HUMAN PERCEPTION OF ENVIRONMENTAL SOUNDS

EDITED BY: Francesco Aletta, Bert De Coensel and PerMagnus Lindborg
PUBLISHED IN: Frontiers in Psychology and Frontiers in Sustainable Cities

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-257-1
DOI 10.3389/978-2-88971-257-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

HUMAN PERCEPTION OF ENVIRONMENTAL SOUNDS

Topic Editors:
Francesco Aletta, University College London, United Kingdom
Bert De Coensel, Ghent University, Belgium
PerMagnus Lindborg, City University of Hong Kong, SAR China

Citation: Aletta, F., De Coensel, B., Lindborg, P., eds. (2021). Human Perception of Environmental Sounds. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88971-257-1

Table of Contents

frontiers in Psychology

EDITORIAL
published: 28 June 2021
doi: 10.3389/fpsyg.2021.714591

Editorial: Human Perception of Environmental Sounds

Francesco Aletta[1]*, *Bert De Coensel*[2] *and PerMagnus Lindborg*[3]

[1] *Institute for Environmental Design and Engineering, The Bartlett, University College London, London, United Kingdom,* [2] *WAVES Research Group, Department of Information Technology, Ghent University, Ghent, Belgium,* [3] *School of Creative Media, City University of Hong Kong, Hong Kong SAR, China*

Keywords: soundscape, perception, environment, acoustics, urban planning, health, design, noise reduction

Editorial on the Research Topic

Human Perception of Environmental Sounds

1. INTRODUCTION

Environmental sounds are a key component of the human experience of a place as they carry meanings and contextual information, together with providing situational awareness. They have the potential to either support or disrupt specific activities as well as to trigger, to inhibit, or simply to change human behaviors in context. The experience of acoustic environments can result in either positive or negative perceptual outcomes, which are in turn related to well-being and Quality of Life. In spite of its relevance to the holistic experience of a place, the auditory domain is often not given enough prominence in environmental psychology studies. Environmental sounds are typically considered in their negative perspective of "noise" and treated as a by-product of society. However, the research (and practice) focus is gradually shifting toward using environmental sounds as mediators to promote and enrich communities' everyday life. Designers explore how natural sounds can be mixed into urban life.

While there is a lot of research happening in this area, the underlying mechanisms connecting environmental sounds, the physical and social context where they occur, and their perceptual effects on users, are still not fully understood (Axelsson et al., 2019). Furthermore, when exploring the aforementioned relationships, more challenges arise in terms of psychometrics and ecological validity of the methodologies involved. All such issues need to be addressed by researchers and practitioners of the built environment. For this reason, a broad spectrum of submissions was invited for this Research Topic. Article types ranged from conceptual analyses, to reviews and research papers. The studies presented here dealt with the characterization and perception of single environmental sounds or complex acoustic environments, as well as their management and design implications for the urban realm. The focus could either be on theoretical aspects (e.g., relationships between sounds and psychological and physiological aspects) or methodological aspects (e.g., protocols and procedures to gather objective and subjective data).

Edited and reviewed by:
Giuseppe Carrus,
Roma Tre University, Italy

***Correspondence:**
Francesco Aletta
f.aletta@ucl.ac.uk

Specialty section:
This article was submitted to Environmental Psychology, a section of the journal Frontiers in Psychology

Received: *25 May 2021*
Accepted: *04 June 2021*
Published: *28 June 2021*

Citation:
Aletta F, De Coensel B and Lindborg P (2021) Editorial: Human Perception of Environmental Sounds. Front. Psychol. 12:714591. doi: 10.3389/fpsyg.2021.714591

2. RESEARCH THEMES

Considering the broad scope of the call for papers, the topics and research questions addressed by the submissions we received were very diverse. Looking retrospectively at them, we tried to identify common themes and eventually clustered them under four main categories. These were: Soundscape theory; Soundscape for health and well-being; Sound perception in urban environments; and Soundscape design. As we see, soundscape is a recurring concept; this is unsurprising considering how "perception"—which was the core aspect of this Research Topic—

is an intrinsic aspect in soundscape studies (Kang et al., 2016). The term soundscape itself was standardized and defined as an "acoustic environment as *perceived* and/or understood by a person, or people, in context" (ISO, 2014). The standardization process for other methodological aspects is still in progress and is expected to be informed by the intense activity currently taking place both in soundscape research and practice.

2.1. Soundscape Theory

While the first harmonization efforts in soundscape studies started more than a decade ago, soundscape theory *per se* could still be considered at an early stage of development for many aspects. If consensus has been found on some basic definitions and frameworks, there is still a lot of debate around methodological approaches, as well as theoretical models underpinning the soundscape concept itself, and how it relates to human psychology and physiology. Thus, contributions to this particular research strand were particularly welcome to advance the scientific conversation on these issues. Chen and Ma synthesized data from semi-structured interviews with 75 participants using a Grounded Theory approach and proposed a conceptual model to define and characterize healthy acoustic environments. Fiebig et al. proposed a conceptual paper about the application of emotion theory to soundscape studies. Their analysis revolves around three main themes, namely: the effect that the acting of collecting soundscape itself can have on people's emotional response to an acoustic environment; whether the affective qualities of a soundscape are actually consciously accessible to people in the first place; and whether it is possible for people to separate the emotion related to a sound environment from affective predisposition. Lionello et al. worked on a large-scale soundscape survey dataset collected in accordance with the ISO/TS 12913-2:2018 and explored how people interpret the Likert scale metrics associated with that soundscape data collection instrument in psychometric terms.

2.2. Soundscape for Health and Well-Being

On the one hand, many aspects of the negative influence of noise on people's health and well-being are well-established through research over the past several decades. For example, sustained exposure to noise near airports or high-density road traffic, even at levels well within legal regulations, has been linked to higher incidence of a range of cardiac illnesses, and overly reverberant classrooms blur phoneme perception and slow down children's learning of language. This knowledge and concerted efforts to enforce regulations have gradually come to influence urban planning (European Parliament and Council, 2002) and construction (Department for Education, 2003). On the other hand, positive effects of sound and soundscape have only more recently moved into focus (Aletta et al., 2018). This is investigated in four articles of the Research Topic. Zhou et al. conducted an experiment with 70 hospital inpatients. Participants listened to soundscape and music recordings through a virtual reality headset, and the researchers measured physiological markers on psychological stress recovery. Using a similar experimental setup, Benfield et al. presented visual images of natural parks while manipulating the soundscape by adding different types of extraneous noise. Responses from 229 volunteers were analyzed with a time series approach. Eqlimi et al. measured the effect of different kinds of ecologically valid noise, such as highway traffic or multi-talker babble, on a learning task. They linked noise type to specific neural correlates that have a negative impact on overall attentive state and capacity to decode spoken language. Gasco et al. adopted a big-data approach: they used geo-referenced social media images from Flickr to characterize the city soundscape of London and built a model with socioeconomic variables, official noise exposure levels, and the soundscape estimated from social media as indicators to predict health outcomes for the population. Ratcliffe offered a comprehensive narrative literature review on the growing research area investigating the relationships between soundscapes, the experience of nature, and restorative environments, which is of theoretical interest for health-related studies.

2.3. Sound Perception in Urban Environments

While the previous experimental studies employed relatively controlled audio stimuli, the next three articles chart complex acoustic environments in built-up, urban spaces, with a bearing on architecture and urban planning. Taghipour et al. set up a laboratory listening experiment and explored how conventional room acoustics parameters would perform in predicting the perceived acoustic comfort in outdoor proximity spaces (e.g., courtyards, balconies) of residential buildings. In Versümer et al. authors surveyed a large number of participants who were asked to recall and describe low-level sounds in everyday situations. The researchers identified a range of sound source types, and explored their impact of on annoyance, valence, arousal, and mental fade-out ability, along with individual and demographic predictors. Lenzi et al., report an observational study during several months of the initial pandemic lockdown in the Basque Country. Pleasantness, eventfulness, and sound source type were analyzed to yield a picture of how people and animals reacted to the extraordinary circumstances in terms of their acoustic communication, as well as people's use of different modes of transport and outdoor-indoor behavior.

2.4. Soundscape Design

The fourth group is exemplary of applied sound perception research. The articles described how the sonic material was deliberately varied and perceptions measured while at the same time the authors had an aesthetic design goal with their work. Rajguru et al. prepared a mini-review about challenges and opportunities in spatial sound perception and soundscape studies, with a focus on augmented and virtual reality methodologies. Cuadrado and collaborators conducted an experimental study on 253 primary school children where some of the sonic elements of an audio story were marked through spatial design. In a mixed-methodology approach, emotional responses were measured with electrodermal resistance, and the children self-reported immersion and mental imagery. Finally, Trudeau et al. reported a case study at a public plaza in Montreal.

Visitors evaluated the perceived quality of the sonic environment, where three different designs of a water feature alternated.

3. CONCLUDING REMARKS

While the four themes discussed above certainly do not cover the full range of questions being debated in the context of soundscape studies, they do detect a few "hot topics" and areas of interest for researchers and practitioners in the field. The psychological theory underpinning the environmental sounds perception processes could still be considered (at least) as evolving. There is a clear interest in making a connection with health and well-being frameworks and also an outlook toward design and co-creation of open public spaces. Virtual reality techniques are now commonly used in perceptual experiments, together with onsite surveys and the analysis of "big data" from public sources. Going forward, it will be essential to include all possible stakeholders in the debate: the public, researchers, practitioners, artists, and professionals with different skills and expertise. This will help testing new hypothesis and triangulate methodologies and results.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

The Editors are grateful to all authors, reviewers, and technical staff at Frontiers for their invaluable contributions to this Research Topic.

REFERENCES

Aletta, F., Oberman, T., and Kang, J. (2018). Associations between positive health-related effects and soundscapes perceptual constructs: a systematic review. *Int. J. Environ. Res. Public Health* 15:2392. doi: 10.3390/ijerph15112392

Axelsson, Ö., Guastavino, C., and Payne, S. R. (2019). Editorial: soundscape assessment. *Front. Psychol.* 10:2514. doi: 10.3389/fpsyg.2019.02514

Department for Education (2003). *Acoustic Design of Schools: Performance Standards - Building Bulletin 93*.

European Parliament and Council (2002). *Directive 2002/49/EC Relating to the Assessment and Management of Environmental Noise*.

ISO (2014). *ISO 12913-1:2014 Acoustics - Soundscape - Part 1: Definition and Conceptual Framework*.

Kang, J., Aletta, F., Gjestland, T. T., Brown, L. A., Botteldooren, D., Schulte-Fortkamp, B., et al. (2016). Ten questions on the soundscapes of the built environment. *Build. Environ.* 108, 284–294. doi: 10.1016/j.buildenv.2016.08.011

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 March 2020
doi: 10.3389/fpsyg.2020.00344

Room Acoustical Parameters as Predictors of Acoustic Comfort in Outdoor Spaces of Housing Complexes

Armin Taghipour[1]*, *Sahand Athari*[1], *Arnthrudur Gisladottir*[1,2], *Tessa Sievers*[1,3] and *Kurt Eggenschwiler*[1]

[1] *Laboratory for Acoustics/Noise Control, Empa, Swiss Federal Laboratories for Materials Science and Technology, Dübendorf, Switzerland,* [2] *Department of Engineering, Aarhus University, Aarhus, Denmark,* [3] *The Carl von Ossietzky University of Oldenburg, Oldenburg, Germany*

OPEN ACCESS

Edited by:
Francesco Aletta, University College London, United Kingdom

Reviewed by:
Louena Shtrepi, Politecnico di Torino, Italy
Marko Horvat, University of Zagreb, Croatia
Laura Estevez Mauriz, Chalmers University of Technology, Sweden

***Correspondence:**
Armin Taghipour
armin.taghipour@empa.ch

Specialty section:
This article was submitted to Environmental Psychology, a section of the journal Frontiers in Psychology

Received: *14 October 2019*
Accepted: *13 February 2020*
Published: *04 March 2020*

Citation:
Taghipour A, Athari S, Gisladottir A, Sievers T and Eggenschwiler K (2020) Room Acoustical Parameters as Predictors of Acoustic Comfort in Outdoor Spaces of Housing Complexes. Front. Psychol. 11:344. doi: 10.3389/fpsyg.2020.00344

Room acoustical parameters have frequently been used to evaluate or predict the acoustical performance in rooms. For housing complexes in urban areas with high population density, it is important to improve acoustic performance not solely indoors, but outdoors as well; for example on the balconies or in the yards. This paper investigates to what extent classic room acoustical parameters would be able to predict the perceived acoustic comfort in outdoor spaces (i.e., courtyards) of virtual housing complexes. Individual and combined effects of a series of independent variables (such as facade absorption, sound source, and observer position) on short-term acoustic comfort were investigated in three laboratory experiments. ODEON software was used for virtual inner yard simulation, whereby 2D spatialization was carried out for a playback over five loudspeakers. Moderate facade absorption was found to increase acoustic comfort. Relatively pleasant and relatively unpleasant sounds were associated with comfort and discomfort, respectively. Lower acoustic comfort ratings were observed at receiver positions with high sound pressure levels and/or strong flutter echoes. A further analysis of the results is carried out here with respect to the room acoustical parameters and their ability to predict the acoustic comfort ratings. Speech transmission index (STI), definition (D50), clarity of speech (C50) and music (C80), early decay time (EDT), and lateral energy fraction (LF80) were found to be significantly correlated with acoustic comfort. They were found to be significant predictors of acoustic comfort in a series of linear mixed-effect models. Furthermore, linear mixed-effect models were established with the A-weighted equivalent continuous sound level, LAeq, as a significant predictor of acoustic comfort.

Keywords: acoustic comfort, inner yard, room acoustical parameters, psychoacoustic experiment, virtual acoustics

1. INTRODUCTION

Development and densification of urban areas has led to an alteration of the urban sound environment and many inhabitants are exposed to high noise levels in their everyday life. Noise emitted from classic noise sources (aircraft, railway etc.) has been related to several health implications and disturbances (WHO, 2018) and thus, the reduction of noise emission in urban areas has been the main objective of conventional and construction acoustics.

One approach to the mitigation of noise in urban living areas is the construction of housing complexes with courtyards or inner yards, where the buildings perform as shields, lowering sound levels from road traffic on one side of the buildings. This allows for several rooms of dwellings to face a "quiet" side of the building complex (Öhrström et al., 2006). Another advantage is that inner yards give access to a recreational outdoor space with lower sound levels from road traffic, which has the opportunity to support various needs of the residents for relaxation, sports or other activities (Gidlöf-Gunnarsson and Öhrström, 2010). Thus, inner yards of housing complexes are under investigation in Switzerland as building typologies with capacities for improvement of the sound environment and acoustic comfort (Sturm and Bürgin, 2016; Sievers et al., 2018; Sturm et al., 2019).

Gidlöf-Gunnarsson and Öhrström (2010) highlighted physical environmental aspects presence of which is found to be highly valuable within inner yards, one of them being the protection from disturbing noise. Although the housing complex benefits from a shielding effect from the street, the inhabitants are confronted with daily life sounds from within the yard itself (Taghipour et al., 2019b,c). Depending on material properties and the building structure among others, the housing complex can induce complicating acoustic effects within the yard, such as multiple reflections, diffraction, and diffusion (Yang et al., 2013). A sound pressure level (SPL) increase of up to 8 dB has been reported due to multiple reflections outside of an apartment complex in comparison to a semi-free field (Yang et al., 2013). Thus, an improvement of acoustic comfort in the building design could benefit the residents. As an example, the use of material with absorptive properties on surfaces outside housing complexes could reduce the SPL, which could then result in increased acoustic comfort (Calleri et al., 2017).

Although the use of sound absorbing materials is well known for the improvement of acoustic comfort in closed rooms/buildings (Battaglia, 2014; Xiao and Aletta, 2016; Thomas et al., 2018), less is known about the use of such materials for the improvement of acoustic comfort for residents of housing complexes with shared inner yards (Taghipour et al., 2019c). Within other exterior spaces of the urban layout, facade absorption has been found to affect the acoustic performance. In public squares, facade absorption has proved to be influential in the subjective assessment of space wideness (Calleri et al., 2017). Alongside streets, building facade and balcony absorption has been found to reduce levels from traffic noise (Lee et al., 2007; Yeung, 2016) and leisure noise (Badino et al., 2019) along facades. Hornikx and Forssén (2009) have found that the use of absorptive facade materials in a shielded canyon could lead to SPL reduction for various observer positions within the canyon. By combining the use of facade absorption and geometrical modification, such as in balcony design, building facades seem to be potentially effective mitigators of noise (Lee et al., 2007). In a study of the effect of facade shape and acoustic cladding on the reduction of leisure noise levels in a street canyon, Badino et al. (2019) have stated that by adding sound absorbing materials on a geometrically optimized facade, a reduction of up to 10 dB in the A-weighted SPL can be achieved. This includes optimized design of balconies, which can greatly influence the facade noise levels (Echevarria Sanchez et al., 2016; Badino et al., 2019). Generally, balconies on building facades have been found to provide significant protection from a noise source on the ground or on the roadway (Hossam El Dien and Woloszyn, 2005; Tang, 2017), although the protective effect can be weakened by reflective balcony ceilings (Hossam El Dien and Woloszyn, 2004; Wang et al., 2015). The shape and placement of balconies also have to be considered. Hossam El Dien and Woloszyn (2005) found that inclined parapet can provide equivalent reduction in SPL as insulation treatments while multiple rectangular balconies were found to be problematic reflectors (Tang, 2005). Another facade property that has been found to shape the perceived acoustic characteristics of urban spaces is the scattering coefficient of the applied facade materials (Calleri et al., 2018). However, the scattering properties were reported not to have a significant influence on reduction in SPL (Onaga and Rindel, 2002; Badino et al., 2019).

From the literature above, it is obvious that acoustic performance of outdoor urban areas—including inner yards of housing complexes—is affected by architectural design and configuration. Every sound is modified and articulated by the materiality and shape of surrounding surfaces (Maag et al., 2019) and thus, architectural design has a great potential to enhance acoustic comfort in cities (Badino et al., 2019). The challenge is that decisions made regarding the design of an outdoor acoustic space is in the hands of various professionals of the built environment, such as planners, architects, engineers and urbanists, and in some cases, acousticians and sound quality experts. The professionals from various backgrounds have different understandings of the acoustic phenomena and partially different objectives (Brown et al., 2016; Coelho, 2016; Sturm and Bürgin, 2016; Sturm et al., 2019). Often, sound has been seen as an unresolved problem, rather than a planned and designed quality (Maag et al., 2019).

Over the last few decades, consideration of acoustic comfort and soundscape quality within the urban living environment has become more eminent (Schafer, 1993; ISO 12913-1, 2014; Brown et al., 2016; Kang, 2017; ISO 12913-2, 2018), with a focus on the design of a relatively pleasant sound environment instead of focusing on noise emission alone. While the definition of soundscape has been standardized (ISO 12913-1, 2014), the term "acoustic comfort" has a broader and more colorful definition in the literature (Taghipour et al., 2019b). In many studies, the improvement of acoustic comfort has been presented as the general improvement of acoustics, measured in objective acoustical and/or room acoustical parameters (such as lower SPL) (Xiao and Aletta, 2016; Thomas et al., 2018). Other studies have used a subjective evaluation of the acoustic comfort (Yang and Kang, 2005; Kang and Zhang, 2010; Battaglia, 2014; Taghipour et al., 2019b), where acoustic comfort was found to be related to the SPL (Yang and Kang, 2005).

With this background, the present paper investigates whether room acoustical parameters would be proper indicators of acoustic comfort in outdoor areas (i.e., inner yards). Room acoustical parameters were actually developed for performances of music and speech in rooms

(ISO 3382-1, 2009; IEC 60268-16, 2011; ODEON, 2018), but have been also used for partially-bounded spaces with open ceilings, such as ancient theaters (ODEON, 2018), historical courtyards for musical performances (Iannace, 2018), and urban spaces (Calleri et al., 2018; Taghipour et al., 2019c). These parameters are not too complex, e.g., a number of them are simple energy ratios, which are available in many simulation software and measurement tools. It is therefore compelling to investigate these parameters for the acoustic quality in partially-bounded outdoor spaces, such as inner yards. This would be particularly useful for architects, acousticians, urban soundscape designers, etc.—who typically have access to simulating software—for the design and development of housing complexes.

The underlying experimental data for the present paper originated from three psychoacoustic laboratory experiments on acoustic comfort in virtual inner yards (Taghipour et al., 2019b). Portions of this study have been published before by Sievers et al. (2018), Taghipour et al. (2019b), and Taghipour et al. (2019c). While Sievers et al. (2018) briefly presented Experiment 1, Taghipour et al. (2019b) reported all three experiments with an analysis of the results with respect to the experimental design variables. Furthermore, Taghipour et al. (2019c) presented a first and brief analysis of the data with respect to the room acoustical parameters, which will now be reported in an expanded length in the present paper. In order to offer the reader a complete picture and to serve as a standalone manuscript, this paper reports the original experiments (Taghipour et al., 2019b), including additional information (e.g., level-time histories and spectral contents of the sound signals, statistical analysis regarding the rating time, etc.), before reporting the analysis with respect to the room acoustical parameters and their association with acoustic comfort.

2. METHODS

Comfort and discomfort reactions to sounds in virtual (acoustic) outdoor spaces of housing complexes were investigated by means of three psychoacoustic laboratory experiments. The observed "short-term" comfort or discomfort ratings related to acute comfort in response to each stimulus, rather than long-term comfort or well-being which is relevant in *post-hoc* field surveys. Specifically, the term "short-term" refers to the time period during and after an acoustic stimulus' playback and before the next stimulus is presented (Taghipour et al., 2019a,b). Furthermore, the term "acoustic comfort" is subjective (and perceptual) and refers to how comfortable a subject was in the presence of each stimulus in the virtual inner yard.

To investigate possible differences in short-term comfort in inner yards with different building facades, sound propagation was simulated in virtual outdoor spaces. Thereby, single-channel recordings were auralized for a multi-channel playback system (Taghipour et al., 2019b).

Note: This study was approved by Empa's Ethics Committee (Approval Nr. CMI 2018-194).

2.1. Experimental Questions

All three experiments presented in this paper investigated which sound sources were associated with short-term acoustic comfort or discomfort. Furthermore and more importantly, the aim of the three experiments was to investigate the effect of the facade's cladding (absorbing vs. reflecting materials) on the perceived acoustic comfort.

- Experiment 1 dealt with the question whether there was a difference between acoustic comfort from sounds in virtual inner yards with reflecting or absorbing facade setups. Furthermore, it was investigated whether receiver (i.e., observer) positions in the yard or on the balcony might be distinctively influenced by the facade covering.
- Experiment 2 investigated the influence of the degree of facade absorption on acoustic comfort. Furthermore, it was investigated whether perceptual differences existed on the balconies of different floors.
- Experiment 3 dealt with the usage of additional facade absorbing materials on the balcony ceilings and their possible influence on the perceived acoustic comfort.

Taghipour et al. (2019b) stated a series of experimental hypotheses resulting from these questions. More details about the design of the experiments and the independent variables in each experiment will be provided in Section 3.

2.2. Listening Test Facility

The three experiments presented in this paper were conducted in the listening test facility of Empa, named AuraLab, which has a separate listening and control room allowing for audio-visual supervision to comply with ethical requirements (Taghipour et al., 2019a,b).

AuraLab satisfies room acoustical requirements for high-quality audio reproduction in terms of its background noise and reverberation time (Taghipour et al., 2019a). A 3D immersive sound system with 16 separate audio channels is installed in AuraLab. Fifteen loudspeakers "KH 120 A" (Georg Neumann GmbH, Berlin, Germany) are located in a hemispherical arrangement on 3 levels (0, 30, and 60° vertically) in a distance of 2 m from the central listening spot. Bass management is performed by two subwoofers "KH 805" (Georg Neumann GmbH, Berlin, Germany) and a digital signal processor (Taghipour et al., 2019a). Stimuli of the experiments presented here were played back by a 2D setup over the five loudspeakers at the vertical level of 0° (subject's ear level) and both subwoofers (see **Figure 1**). The reason for this is that ODEON delivers a 2D surround sound—i.e., a five-channel signal—for playback; more details in Section 2.3. Furthermore, the carpeted floor was covered with additional absorbers on the floor (Taghipour et al., 2019b). **Figure 1** shows the setup in Auralab, where the subject was seated in the central listening spot.

2.3. Recording, Simulation, and Auralization

Figure 2 shows a block diagram of the signal processing from recording to playback. To collect sound sources, single-channel recordings were carried out in a semi-anechoic chamber by

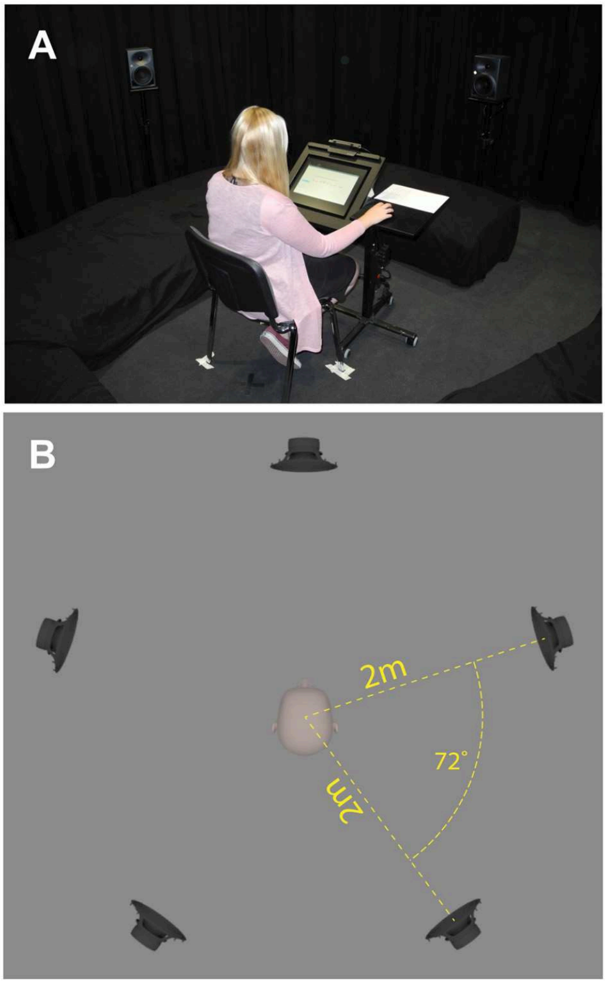


FIGURE 1 | The experimental setup in AuraLab, where the subject was seated in the central listening spot **(A,B)**. For this experiment, additional porous absorbers were put on the floor **(A)**. The stimuli were played back through five satellite loudspeakers and two subwoofers **(B)**.

means of a B&K 4006 microphone (Brüel & Kjaer, Nærum, Denmark), positioned on the reflecting floor. After suitable 8-s extracts were cut from the recordings, they were normalized to the A-weighted level (i.e., A-weighted equivalent continuous sound level, L_{Aeq}) of the signal with the largest maximum absolute value of the amplitude (Taghipour et al., 2019b,c). **Figure 3** shows level-time histories (L_{AF} curves) and one-third octave spectra of the 8-s extracts for a normalized L_{Aeq} of 50 dB(A). As shown in **Figure 3**, several sounds—typical for outdoor living environment—were used in the course of this study. Although, generally, all sounds are neutral in value, they can be determined as pleasant or unpleasant in a particular context and setup by human listeners (ISO 12913-1, 2014; Taghipour et al., 2019b). Prior to the experiments presented in this paper and in a relative approach (i.e., amongst each other), the sounds used in this study were judged by acousticians as to be relatively more or less pleasant. The aim was to facilitate the subjects (of the main experiments) with a variety of sounds that are associated with comfort or discomfort in a laboratory setup.

Room acoustic simulations were done with ODEON v. 14.03 (Odeon A/S, Kgs. Lyngby, Denmark), which uses geometrical acoustics with image-sources and a ray tracing algorithm. Geometrical acoustics methods are currently the most widely used methods in modeling room acoustics, auralization, and outdoor acoustics applications (Georgiou, 2018). Although geometrical acoustics methods have their limitations, e.g., not being able to precisely model wave phenomena such as diffraction (Elorza, 2005), they are popular because of their simplicity and computation efficiency and the ability to model up to high frequencies. Element-based numerical wave-based methods such as Finite Difference Time Domain (FDTD) and Pseudo-Spectral Time Domain (PSTD) in time domain and Finite Element Method (FEM) and Boundary Element Method (BEM) in frequency domain are used for precise acoustic simulations (Hornikx, 2016), but since their solvers need discretized domains and this means a large number of voxels or meshes on a 3D geometry, the computation expense grows much heavily. Hybrid methods have also been used to model higher frequencies using geometrical acoustics and lower frequencies with wave-based methods to obtain a compromise. But still, the most efficient and feasible method seems to be the geometrical acoustics method (Georgiou, 2018), especially with regards to reliable auralization.

An omni-directional sound source was placed in the yard 1.2 m above the ground. Impulse responses were calculated by the software for various observer positions[1]. The simulations were carried out with 200,000 rays and impulse responses of 3.5 s of length. The transition order was set to two. Details about the facade materials are provided by Sievers et al. (2018) and Taghipour et al. (2019b,c). A 2D auralization (2D Surround sound based on first-order B-format Ambisonics) was carried out for five loudspeakers based on the setup in AuraLab (i.e., separated from one another by 72° horizontally, at the vertical angle of 0°). Although ODEON input signals (of different sources) had the same L_{Aeq} (see above), the ODEON outputs exhibited diverging L_{Aeq}, as they possessed unequal spectral and temporal characteristics, to which the virtual rooms reacted differently. The stimuli reported in this paper were simulated considering a single source, one source position, and several observer positions (Taghipour et al., 2019b).

The multi-channel ODEON output signal was upsampled from 44.1 to 48 kHz, as this is a requirement of the playback system. Furthermore, it was low pass (f_c = 10 kHz) and high pass (f_c = 20 Hz) filtered. After being gated with squared-cosine ramps, the multi-channel signal was allocated to the corresponding loudspeakers: front, front-left, front-right, back-left, and back-right. By means of crossover filtering, low-frequency components of the signals were played back over the two subwoofers in the room. Beside the room acoustical

[1]The observer positions in the yard, 1.2 m above the ground, represented a person sitting on a bench for the purpose of recreation/relaxing. As—from an experimental design perspective—the authors had decided to have one source position for all the sound sources, they decided this height to represent playing children, as well as conversations, etc., therefore, 1.2 m above the ground.

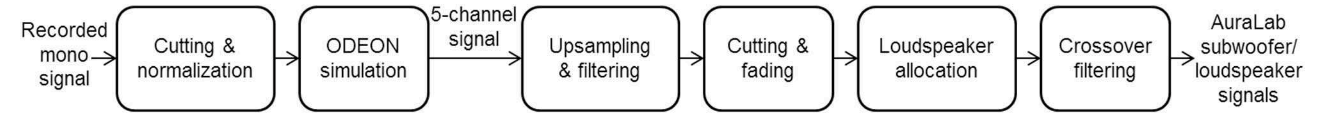


FIGURE 2 | Block diagram of the signal processing steps from recording to playback (Taghipour et al., 2019c).

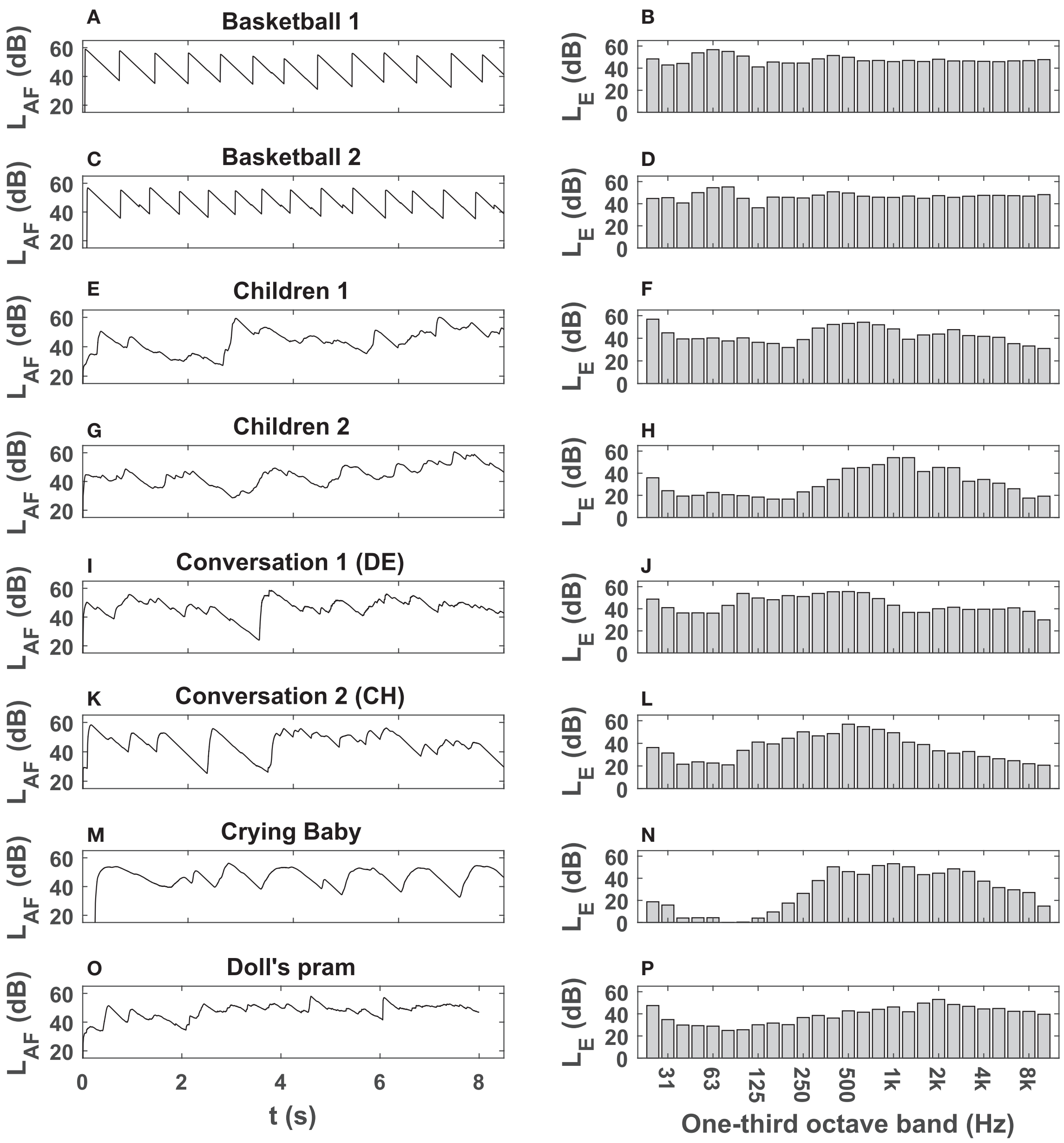


FIGURE 3 | L_{AF} curves, i.e., A-weighted fast level-time histories (left: **A,C,E,G,I,K,M,O**) and one-third octave spectra (right: **B,D,F,H,J,L,N,P**) of the 8 signals used as ODEON inputs in the three experiments. All signals have an L_{Aeq} of 50 dB(A).

simulation in ODEON, all signal processing steps (see **Figure 2**) were done in the MATLAB environment v. R2016b (MathWorks, Natick, MA, USA).

2.4. Reference Inner Yard

The reference inner yard used in this study was a simplified 3D model of an existing housing complex in Dübendorf,

FIGURE 4 | The reference inner yard: **(A)** the building complex and **(B)** its ODEON model.

Switzerland. The geometric model was built in the SketchUp software environment (Trimble Inc., Sunnyvale, CA, USA) and was imported into the ODEON software environment using the plug-in SU2Odeon (Taghipour et al., 2019b). **Figure 4** shows the inner yard and its ODEON model. The walls were of brickwork and concrete with large glass windows. Since ODEON works with bounded/closed room models, the inner yard was modeled as an unceiled room (100 × 20 × 20 m), which—for practical reasons—was inserted in a larger box (10 meters away from each side) with a perfect absorbing inner surface, representing free field (Taghipour et al., 2019b).

2.5. Experimental Sessions

The three psychoacoustic experiments were conducted as focused listening tests in form of a complete block design with repeated measures. Subjects did the tests individually. After reading the study information, they signed a consent form. Thereafter, they answered the first part of the questionnaire about their hearing and well-being (see **Appendix**). The subjects were then introduced to the listening test software which guided them through the test. After the listening test, the subjects filled out the rest of the questionnaire (demographic data) (Taghipour et al., 2019b).

Experiment 1 was conducted as a single listening test with 27 subjects (7 females and 20 males, aged between 19 and 57 years old, median 38 years). Experiments 2 and 3 were conducted as two listening tests in one experimental session with 42 subjects (13 females and 29 males, aged between 18 and 64 years old, median 41 years), whereby the order of the Experiments 2 and 3 was (randomly) counterbalanced between the subjects (Taghipour et al., 2019b). It was reported by Taghipour et al. (2019b) that all subjects declared to have normal hearing (self judgment) and to feel well. Since no audiometric test was performed, the subjects were characterized as self-reporting normal-hearing.

2.6. Listening Test Software, Procedure, and the Comfort Scale

To familiarize with the sounds and the test software, subjects listened to several orienting and training stimuli. They were chosen such that the subjects were familiarized with the range of different sources, facade types, and sound pressure levels, before starting the main experiment.

- Experiment 1: ten orienting and three training stimuli, out of a total of 60 stimuli.
- Experiment 2: six orienting and four training stimuli, out of a total of 40 stimuli.
- Experiment 3: four orienting and two training stimuli, out of a total of 27 stimuli.

The main listening test began thereafter. For each stimulus, subjects completed the following statement: "In this virtual inner yard and in the presence of this sound, I feel ... " (Taghipour et al., 2019b). Their short-term acoustic comfort were recorded during or after stimulus playback on a verbal bipolar 7-point scale: very uncomfortable (−3), uncomfortable (−2), to some extent uncomfortable (−1), neither comfortable nor uncomfortable (0), to some extent comfortable (+1), comfortable (+2), and very comfortable (+3).

To support the neutral category "neither uncomfortable nor comfortable" in its actual purpose as the scale middle category and to avoid its misuse as an avoiding or diverting answer, an additional "don't know" push button was provided to the subjects (Taghipour et al., 2019b). This option was, however, rarely used by the subjects.

The stimuli were played back in a random order after one another, with a 1.2-s break between stimuli after complete playback. By means of a push button, an option was given to the subjects to listen to each stimulus (only) one more time[2], if they wished to. Subjects rarely made use of this option (Taghipour et al., 2019b).

2.7. Statistical Analysis

The statistical analysis was carried out with IBM SPSS Statistics, v. 25 (IBM Corporation, Armonk, USA). Tested

[2]The repetition was offered "only one more time" in order to prevent a too lengthy experiment. Giving subjects the opportunity to listen to the stimuli "as many times as they wish to" would cause two problems. First, the experimental schedule would not be totally under experimenters' control. That is, there would be a need for planning very long sessions, just to avoid any overlaps. This would not be an efficient use of the laboratory setup and staff. Second, a very essential point in laboratory experiments is to provide all subjects with a similar and controlled situation/setup in the lab. This essential point would not be satisfied, if one subject rarely used any repetitions and another subject regularly used several repetitions. They would not be participating in the same listening test scheme.

effects of the independent variables (and their interactions) on the dependent variable "short-term acoustic comfort" were considered significant if the probability, p, of the observed results under the null hypothesis (H0) was less than 0.05 (Taghipour et al., 2019b).

The individual and combined associations of the independent variables (i.e., experimental design variables) on short-term acoustic comfort were investigated as follows (Taghipour et al., 2019b).

- The complete block design of the experiments enabled carrying out repeated-measures multi-factorial analysis of variance (ANOVA) to compare the mean acoustic comfort ratings for different categories of the categorical independent variables. If necessary, failed assumption of sphericity was corrected by the Greenhouse-Geisser method.
- In order to investigate further the directions of the effects, *post-hoc* pairwise comparisons were done by Fisher's protected least significant difference (LSD) test, corrected by the Bonferroni method.
- Furthermore, when helpful, linear mixed-effects models were fitted to the observed data with independent variables of different types; i.e., categorical variables, covariates, and random intercept (comparison of the models by means of Akaike information criterion (AIC) (Akaike, 1998) and Bayesian information criterion (BIC) (Schwarz, 1978); i.e., choosing the model with the lowest AIC/BIC).

Furthermore, it was investigated whether room acoustical parameters and L_{Aeq} are good predictors of acoustic comfort. To this aim, the following analyses were carried out.

- Scatter plots (of acoustic comfort as a function of the individual room acoustical parameters) were visually examined.
- Correlations between acoustic comfort and the individual room acoustical parameters were calculated, reporting Pearson's correlation coefficient, r, and its significance (Taghipour et al., 2019c). Furthermore, correlations were investigated between mean acoustic comfort and the individual room acoustical parameters.
- Linear mixed-effects models were fitted to the observed data to further investigate the combined analysis of sound source, the individual room acoustical parameters (or L_{Aeq}), and random intercept as predictors of acoustic comfort. Sound source was taken into this analysis, because this variable is independent of the room and absorption characteristics.
- Furthermore, the data from all three experiments were put together in order to make a combined overall analysis of the results possible.

3. EXPERIMENTS

3.1. Experiment 1

Three design variables were used in Experiment 1: inner yard (4 levels), source type (5 levels), and observer position (3 levels) (Sievers et al., 2018). Four yards were modeled in ODEON: the (reflecting) reference yard and three further yards with "exaggerated reflecting building facades," with "absorbing facades," and with "exaggerated absorption." Five different sound sources were tested: a bouncing basketball, a doll's pram, a German conversation, and two sounds of happily playing and laughing children. Three observer points were chosen: two observer points in the yard (1.2 m above the ground, representing the position of someone sitting on a bench), 5 and 20 m away from the source, and one observer point on the second floor balcony about 28 m away from the source. Note that, on average, $L_{\mathrm{Aeq,Balcony}} < L_{\mathrm{Aeq,20m}} < L_{\mathrm{Aeq,5m}}$ and that the observer position at the balcony was considerably more affected by echos and flutter echos (Taghipour et al., 2019b,c). In total, 60 stimuli were prepared for this experiment: 4 × 5 × 3 = 60. The A-weighted equivalent continuous sound levels, L_{Aeq}, of the auralized stimuli were between 42 and 59 dB(A) [mean $L_{\mathrm{Aeq}} = 52$ dB(A)]. Each stimulus was 9 s long.

3.2. Experiment 2

Three design variables were used in Experiment 2: the weighted absorption coefficient α_w (ISO 11654, 1997) (5 levels), source type (4 levels), and observer location (2 levels). α_w was varied with an approximately exponential progression. To avoid major frequency-dependent differences in absorption properties of materials, a simple material model was chosen, for which the frequency dependency of α remained approximately constant as α_w was increased (Taghipour et al., 2019b,c). Doing so, the facade was covered with absorbing materials exhibiting α_w values of 0.05, 0.15, 0.30, 0.55, and 0.95. Four different sources were used: a bouncing basketball, a crying baby, a Swiss German conversation, and a sound of playing children. Two observer points were chosen at the ground floor balcony (patio) and the second floor balcony, 12 and 15 m away from the source. The second floor balcony exhibited lower L_{Aeq} than the ground floor balcony (Taghipour et al., 2019b). In total, 40 stimuli were prepared for Experiment 2: 4 × 5 × 2 = 40. The L_{Aeq} of the auralized stimuli was between 49 to 64 dB(A) [mean $L_{\mathrm{Aeq}} = 60$ dB(A)]. Each stimulus was 10 s long.

3.3. Experiment 3

Three design variables were used in Experiment 3: facade α_w (3 levels), source type (3 levels), and balcony ceiling α_w (3 levels). α_w was varied for Experiment 3 between 0.05, 0.30, and 0.95 for the absorption of the facade, as well as the balcony ceiling. Three sound sources were used: a bouncing basketball, a German conversation, and a sound of playing children. The observer was placed at the second floor balcony, 15 m away from the source (Taghipour et al., 2019b,c). In total, 27 stimuli were prepared for Experiment 3: 3 × 3 × 3 = 27. The L_{Aeq} of the auralized stimuli was between 49 to 59 dB(A) (mean $L_{\mathrm{Aeq}} = 56$ dB(A)). Each stimulus was 10 s long.

4. ROOM ACOUSTICAL PARAMETERS

Room acoustical parameters were originally developed to measure and estimate performances of music and speech in rooms (ISO 3382-1, 2009; IEC 60268-16, 2011; ODEON, 2018). However, they are also used in the case of partially-bounded spaces, i.e., spaces with solid floor and walls, but with open ceilings, e.g., from ancient theaters to modern stadiums (Calleri et al., 2018; Iannace, 2018; ODEON, 2018). Since inner yards of building complexes have a similar partially-bounded shape, several classic room acoustical parameters are considered in this paper regarding their ability to predict acoustic comfort. The hypothesis was that they could be used as measures of the quality of the room acoustical experience in the presence of every day life sounds—such as conversations—in outdoor living environments, e.g., inner yards or street canyons. The room acoustical parameters used in this paper will be introduced in the following.

- Speech transmission index, STI, is a quantitative expression of the extent of speech intelligibility (Houtgast and Steeneken, 1973; IEC 60268-16, 2011). STI is derived for an average gender-independent voice spectrum and is expressed as decimal numbers from 0.00 to 1.00. Values in the ranges 0.00–0.30, 0.30–0.45, 0.45–0.60, 0.60–0.75, and 0.75–1.00 correspond to bad, poor, fair, good, and excellent speech intelligibility, respectively.
- Definition (*Deutlichkeit*), D50, is the ratio of the useful energy (the first 50 ms) to the total energy (ISO 3382-1, 2009). It is expressed as percentage values in this paper.
- Clarity (Speech), C50, is the energy ratio before and after 50 ms, expressed in dB, and is stated by ISO 3382-1 (2009) to be appropriate for clarity of speech. Differences between D50 and C50 are that C50 is expressed in dB while D50 is a fraction or percentage and that the integration times for late reverberation are different. C50 is defined as ten times the logarithm of the ratio of the useful energy (of the first 50 ms) to the late energy (after 50 ms).
- Clarity (Music), C80, is an extension of D50 and C50, but is often used for evaluating the space for music (ISO 3382-1, 2009). It is sometimes referred to as "clarity for music." The only difference between C50 and C80 is the 50-ms or 80-ms limit used in their calculation.
- Early Decay Time, EDT, is a measure that indicates how listeners perceive the reverberation of speech or music at specific listening positions (ISO 3382-1, 2009). It is defined as six times the time during which sound level is attenuated by 10 dB, after turning off the sound source (ISO 3382-1, 2009).
- Lateral Energy Fraction, LF80, is an indication of lateral (bi-directional) energy compared to the early energy (the first 80 ms) (ISO 3382-1, 2009). Late reflections that arrive from lateral directions contribute to the perception of spaciousness (Griesinger, 1997; ISO 3382-1, 2009; ODEON, 2018). These reflections lie between 5 ms and 80 ms. LF80 is an indication of the apparent source width (ASW).
- Dietsch's echo criterion predicts if there is a certain peak in the impulse response that indicates an unwanted audible echo (Dietsch and Kraak, 1986; Kuttruff, 2017).

Room acoustical parameters were calculated in the ODEON environment v. 15 (Odeon A/S, Kgs. Lyngby, Denmark). For D50, C50, C80, and EDT, the average of the values for the two octave bands centered at 500 and 1,000 Hz was used here, as recommended by ISO 3382-1 (2009). A similar approach was applied to EchoD. Furthermore, average LF80 values were used which were provided by the ODEON (ODEON, 2018), according to recommendations by ISO 3382-1 (2009). The LF80 average values were calcultaed over 125, 250, 500, and 1,000 Hz octave bands (ISO 3382-1, 2009; ODEON, 2018).

5. RESULTS

This section briefly discusses the most important results delivered by Taghipour et al. (2019b) accompanied by further experimental results, before reporting the analysis with the room acoustical parameters.

5.1. Independent Design Variables

Figure 5 illustrates mean acoustic comfort ratings and their 95% confidence intervals for data from Experiments 1, 2, and 3, originally presented by Taghipour et al. (2019b).

5.1.1. Results of Experiment 1

In total, 1620 (i.e., 27 subjects $\times$ 60 stimuli) acoustic comfort ratings were collected in Experiment 1. Significant main effects on acoustic comfort were found for all three design variables, i.e., inner yard [$F_{(2.2,57.2)} = 49.6$], sound source [$F_{(3.1,79.3)} = 33.1$], and observer position [$F_{(1.2,31.9)} = 25.8$], all $p < 0.001$ (Taghipour et al., 2019b).

Compared to the two reflecting inner yards, acoustic comfort was rated higher for the two absorbing inner yards, $p < 0.001$. No further differences were found between the inner yards, all $p > 0.05$. Acoustic comfort was rated higher for the two children sounds than for the three other sound sources, all $p < 0.001$. No further significant differences were found between the sound sources, all $p > 0.05$. Furthermore, acoustic comfort ratings revealed to be significantly different for the three observer positions, all $p < 0.01$. The observer positions at 20 m distance in the yard and at the second floor balcony were found to be "more" and "less" comfortable than the position at 5 m distance in the yard, respectively (Taghipour et al., 2019b).

Furthermore, a series of significant interactions were reported and discussed extensively by Taghipour et al. (2019b).

5.1.2. Results of Experiment 2

In total, 1678 (i.e., 42 subjects $\times$ 40 stimuli $-$ 2 missing data points) acoustic comfort ratings were collected in Experiment 2. Significant main effects on acoustic comfort were found for all three design variables, i.e., facade's α_w [$F_{(3.0,121.8)} = 21.0$], sound source [$F_{(2.7,112.1)} = 71.6$], and observer position [$F_{(1.0,41.0)} = 29.3$], all $p < 0.001$ (Taghipour et al., 2019b).

A linear mixed-effect model was fitted to the data of Experiment 2 to investigate the effect of facade's α_w on acoustic comfort (also considering sound source, observer position, and subjects' random intercept). A parabolic relationship was found. That is, acoustic comfort was rated the highest for moderate (i.e., middle) α_w values. Basketball 1 and children 2 were rated as least and most comfortable sounds, respectively, all $p < 0.01$. No further significant difference was found with respect to sound sources, $p > 0.05$. Acoustic comfort was rated higher for balcony 2 than for balcony 0, $p < 0.01$ (Taghipour et al., 2019b).

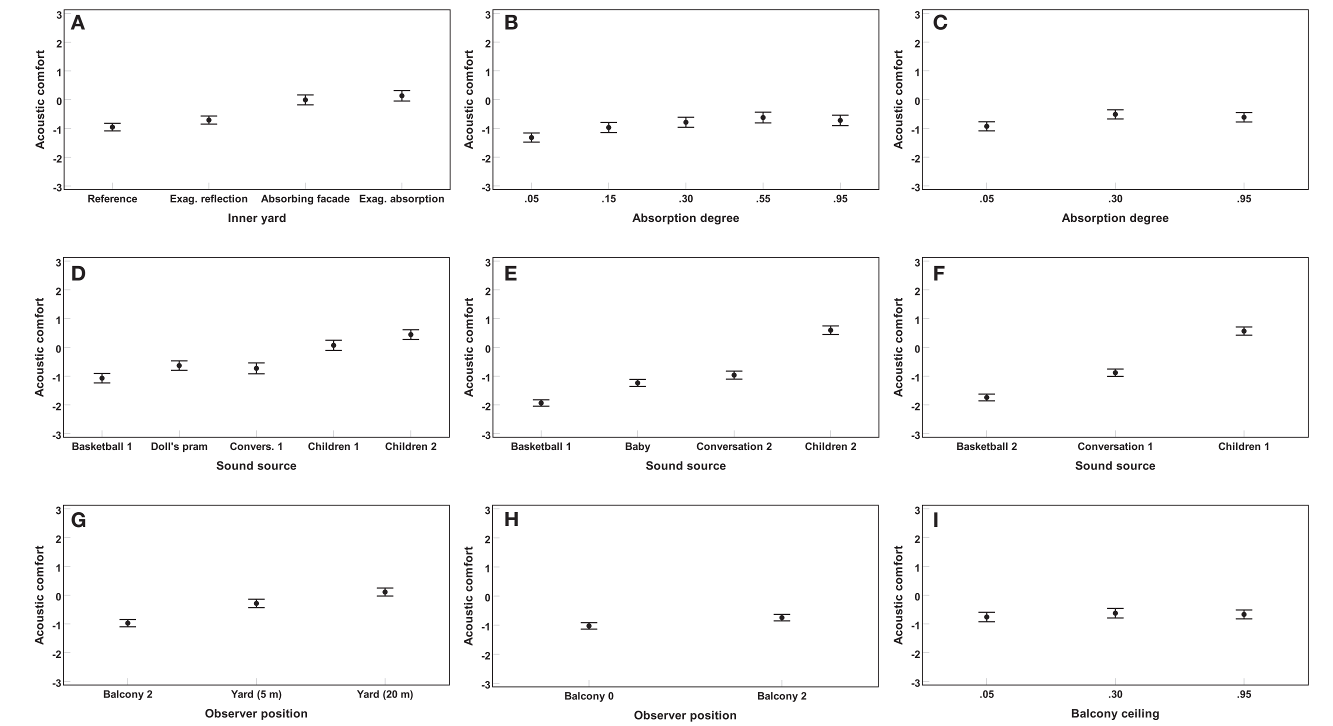


FIGURE 5 | Results of Experiments 1 (left: **A,D,G**), 2 (middle: **B,E,H**), and 3 (right: **C,F,I**): Mean acoustic comfort ratings across subjects and their 95% confidence intervals are shown on the ordinate for different design parameters along the abscissa. In Experiment 1, three variables, i.e., inner yard type (4 levels), sound source (5 levels), and observer position (3 levels), were tested. In Experiment 2, three variables, i.e., facade's weighted absorption coefficient α_w (5 levels), sound source (4 levels), and observer position (2 levels), were tested. In Experiment 3, three variables, i.e., facade's weighted absorption coefficient α_w (3 levels), sound source (3 levels), and balcony ceiling's weighted absorption coefficient α_w (3 levels), were tested (Taghipour et al., 2019b).

Furthermore, a series of significant interactions were reported and discussed extensively by Taghipour et al. (2019b).

5.1.3. Results of Experiment 3

In total, 1134 (i.e., 42 subjects × 27 stimuli) acoustic comfort ratings were collected in Experiment 3. Significant main effects on acoustic comfort were found for facade's α_w [$F_{(1.4,56.6)} = 9.5$] and sound source [$F_{(2.0,80.0)} = 105.7$], all $p < 0.01$. Balcony ceiling's α_w was not found to affect acoustic comfort significantly, although such a non-significant tendency could be observed [$F_{(2.0,80.3)} = 2.8$], $p = 0.07$ (Taghipour et al., 2019b).

Fitted by a linear mixed-effect model, the effect of facade's α_w on acoustic comfort was very similar to that in Experiment 2. Balcony ceiling's α_w, however, was not found to be significantly contributing to the model. Nevertheless, in the absence of any absorbers on the facade, absorbing balcony ceilings tended to improve acoustic comfort. Regarding sound source, basketball 2 and children 1 were rated to be less and more comfortable than conversation 1, respectively, all $p < 0.001$ (Taghipour et al., 2019b).

5.2. Rating Time

Figure 6 shows mean rating time (response time) as a function of acoustic comfort. For all three experiments, a parabolic relationship can be observed. That is, when subjects felt very uncomfortable or very comfortable, they gave their response faster than when they rated their comfort in response to the acoustic stimuli in the middle range of the scale. The absolute fastest mean ratings (in all three experiments) were collected when the stimuli was perceived to be very uncomfortable.

5.3. Playback Sequence

It was investigated whether playback sequence (i.e., the order of stimuli's playback) affected short-term acoustic comfort. **Table 1** shows Pearson's correlation coefficient, r, between acoustic comfort and playback sequence.

It can be observed in **Figure 7** that, overall, acoustic comfort decreased slightly with increasing playback sequence. Nevertheless, as **Table 1** shows, correlation between acoustic comfort and playback sequence was either weak or non-significant. The effect of playback sequence on acoustic comfort was not further analyzed in linear mixed-effect models fitted to the observed data.

5.4. Room Acoustical Parameters

5.4.1. Scatter Plots and Correlations

Figure 8 shows a series of scatter plots of mean acoustic comfort rating as a function of the individual room acoustical parameters. Except for the echo criterion by Dietsch and Kraak (1986), the

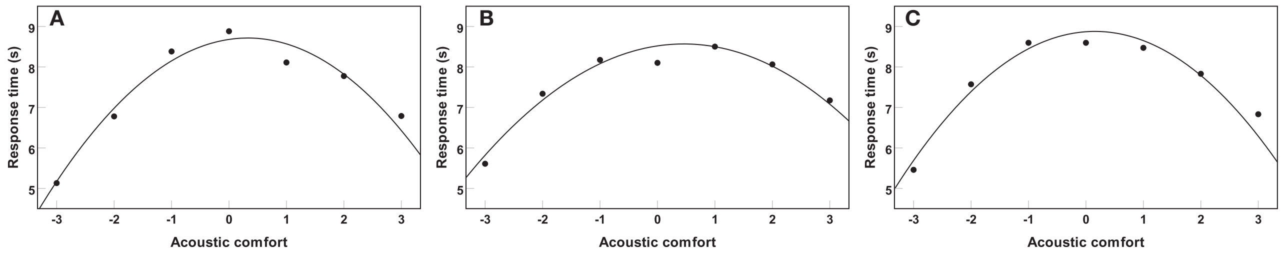


FIGURE 6 | Mean rating time as a function of acoustic comfort rating for **(A)** Experiment 1, **(B)** Experiment 2, and **(C)** Experiment 3.

TABLE 1 | Pearson's *r* for correlations between playback sequence and acoustic comfort.

	Exp. 1	Exp. 2	Exp. 3	Overall
Sequence	−0.02	−0.03	−0.11**	−0.06**

*All significant cases are significant at the level of **p < 0.01.*

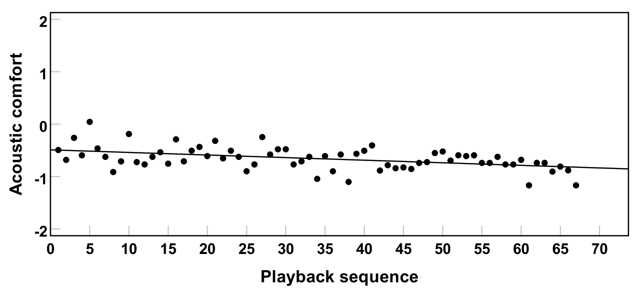


FIGURE 7 | Mean acoustic comfort ratings (for the combined data of the three experiments) across subjects and sound sources as a function of playback sequence. Notice that the second and the third experiments were conducted in one experimental session. Therefore, in the combined analysis, the playback number was registered from 1 to 67 (i.e., 40 and 27 stimuli in Experiments 2 and 3, respectively).

slope sign (i.e., positiveness vs. negativeness) for each room acoustical parameter is consistent for all three experiments. This is further quantified by the significant correlations between acoustic comfort and the individual parameters. Pearson's correlation coefficient, *r*, is reported in **Table 2** for correlations between acoustic comfort ratings and the room acoustical parameters of **Figure 8**. Except for the echo criterion by Dietsch and Kraak (1986) in Experiment 3, all correlations were found to be significant. Note that, whereas, for Experiment 1, rather moderate correlations were observed, for Experiments 2 and 3, correlations were very weak. Furthermore, strong correlations were found between mean acoustic comfort and the individual room acoustical parameters (see **Table 2**).

While short-term acoustic comfort increased with increasing STI, D50, C50, and C80, it decreased with increasing EDT and LF80 (see **Figure 8** and **Table 2**). Since the correlation between acoustic comfort and the EchoD is rather inconsistent for the results of the three experiments reported here—i.e., negative correlation in Experiment 1, positive correlation in Experiment 2, and no significant correlation in Experiment 3—(see **Figure 8** and **Table 2**), this criterion was considered not to be a proper predictor of acoustic comfort. Therefore, the echo criterion by Dietsch and Kraak (1986) was not investigated in the further analysis.

5.4.2. Linear Mixed-Effect Models Including Individual Room Acoustical Parameters

Several linear mixed-effect models were fitted to the observed data to investigate the relationship between the dependent variable acoustic comfort and the individual room acoustical parameters. That is, instead of the physical design parameters (i.e., inner yard, observer position, and α_w of the facade and the balcony ceiling), individual room acoustical parameters were considered in the models as independent variables, accompanied by the categorical independent variable sound source and subjects' random intercept (Taghipour et al., 2019c). Note that, since a majority of the correlations between individual room acoustical parameters and acoustic comfort in **Table 2** are weak, it would be careless to interpret the room acoustical parameters as good predictors of acoustic comfort without considering other independent variables (such as sound source). That is, only if, in the presence of sound source and subject's random intercept, the room acoustical parameters were significant predictors in the fitted linear mixed-effect models, their prediction ability should be taken seriously.

A series of models were fitted to the data which considered sound source, random intercept, and one room acoustical parameter. For the results of all three experiments, including STI, D50, C50, C80, EDT, or LF80 in the models contributed significantly to predict the corresponding acoustic comfort ratings. The direction of their effect was analogous to correlations in **Table 2**. That is, if they were correlated with acoustic comfort positively (or negatively), the corresponding β in the linear mixed-effect model was positive (or negative); see **Table 3**. The linear mixed-effect models were defined by the following equations:

$$y_{ik} = \mu + \tau_{\mathrm{Src},i} + \beta \cdot RAP + u_k + \epsilon_{ik}. \quad (1)$$

In Equation (1), y_{ik} is the dependent variable acoustic comfort, μ is the overall grand mean, $\tau_{\mathrm{Src},i}$ denotes the categorical variable

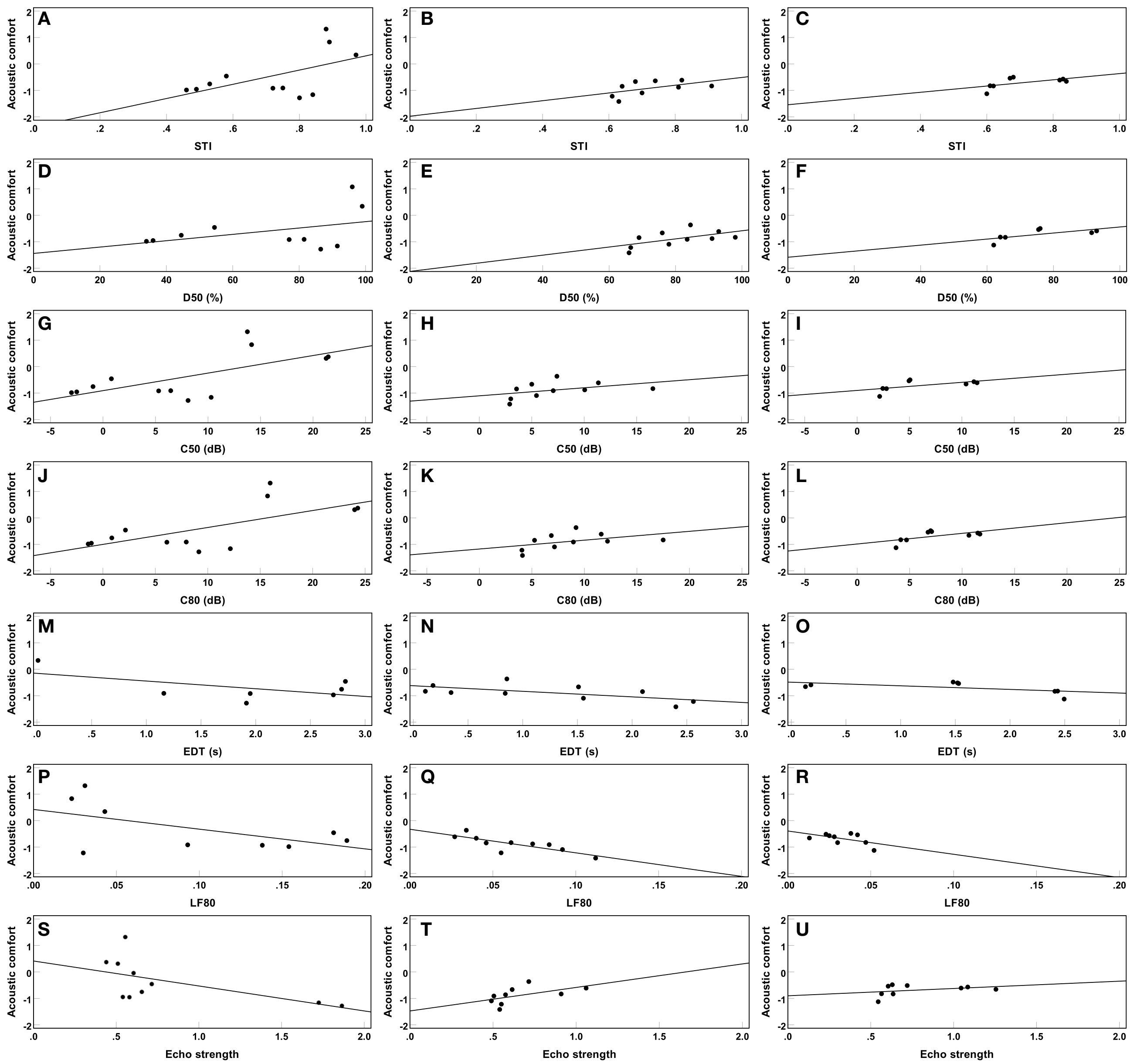


FIGURE 8 | Scatter plots showing mean acoustic comfort ratings across subjects and sound sources as a function of the individual room acoustical parameters. The lines indicate simple linear regressions based on the respective room acoustical parameters. The left panel (i.e., **A,D,G,J,M,P,S**), the middle panel (i.e., **B,E,H,K,N,Q,T**), and the right panel (i.e., **C,F,I,L,O,R,U**) show the results of Experiments 1, 2, and 3, respectively.

source type (five levels in Experiment 1: $i = 1 - 5$, four levels in Experiment 2: $i = 1 - 4$, three levels in Experiment 3: $i = 1 - 3$), *RAP* is the continuous variable (each individual) room acoustical parameter, and β is its regression coefficient. The (unstructured) random effect term u_k is subjects' random intercept (Experiment 1: $k = 1 - 27$, Experiment 2: $k = 1 - 42$, Experiment 3: $k = 1 - 42$). Finally, the error term ϵ_{ik} is the random deviation between observed and expected values of y_{ik}. All parameters contributed significantly to most of the models (all $p < 0.05$). Only for Experiment 1 and in the case of EDT and LF80, the random intercept (i.e., subject) was a non-significant predictor. That is, only in these two cases (out of a total of 18 cases), the significant linear mixed-effect model equations did not include u_k. All models were better than the basic model with sound source and random intercept (without room acoustical parameters). That is, adding individual room acoustical parameters improved the basic models significantly and led to lower AIC and BIC (Schwarz, 1978; Akaike, 1998).

TABLE 2 | Pearson's *r* for correlations between room acoustical parameters and acoustic comfort (and mean acoustic comfort).

		STI	D50	C50	C80	EDT	LF80	EchoD
Experiment 1	Acoustic comfort	0.29**	0.26**	0.32**	0.33**	−0.32**	−0.23**	−0.24**
	(mean acoustic comfort)	(0.55)	(0.41)	(0.66)	(0.67)	(−0.58)	(−0.56)	(−0.59)
Experiment 2	Acoustic comfort	0.08**	0.10**	0.08**	0.08**	−0.11**	−0.14**	0.09**
	(mean acoustic comfort)	(0.53)	(0.57)	(0.43)	(0.46)	(−0.63)	(−0.80)	(0.54)
Experiment 3	Acoustic comfort	0.06*	0.08**	0.06*	0.07*	−0.07*	−0.07*	0.04
	(mean acoustic comfort)	(0.59)	(0.66)	(0.59)	(0.60)	(−0.59)	(−0.53)	(0.35)
Overall	Acoustic comfort	0.18**	0.15**	0.20**	0.21**	−0.18**	−0.11**	−0.08**
	(mean acoustic comfort)	(0.51)	(0.38)	(0.59)	(0.62)	(−0.46)	(−0.31)	(−0.30)

*$p < 0.05$.
**$p < 0.01$.

TABLE 3 | Values of β in Equation (1) for different room acoustical parameters in the three experiments (all $p < 0.01$).

	STI	D50	C50	C80	EDT	LF80
Exp. 1	2.76	1.82	0.07	0.06	−0.44	−6.16
Exp. 2	1.50	1.54	0.03	0.03	−0.21	−8.97
Exp. 3	1.10	1.04	0.03	0.04	−0.12	−8.80
Overall	2.27	1.65	0.06	0.06	−0.31	−6.73

Table 3 shows the values of β in Equation 1 for each room acoustical parameter and experiment. Furthermore, β values are listed for the combined (i.e., overall) analysis of all three experiments; see Section 5.4.3. Note that, in total, there are 24 models in the form of Equation 1 and their complete reporting would not be possible in this paper.

It is important to compare the β values from linear-mixed effect models reported in **Table 3** with the simple linear regressions shown in **Figure 8**. While the linear mixed-effect models additionally consider the strong effect of sound source and random subject intercept, the relationships between acoustic comfort and individual room acoustical parameters resemble those reported in **Figure 8**. The signs of the relationships (i.e., positive or negative correlations) are identical in **Table 3** and **Figure 8**. Furthermore, for each individual room acoustical parameter, the differences between the regression coefficients (i.e., the slopes) in the three experiments show a similar pattern. That is, even considering other predictors, a fairly similar relationship holds between acoustic comfort and the individual room acoustical parameters as to that from **Figure 8**.

5.4.3. Linear Mixed-Effect Models for the Combined Data of the Three Experiments

As mentioned above, linear mixed-effect models were established with the "combined data" from the three experiments (e.g., see **Tables 2**, **3**, "overall"). The same Equation (1) was found to be appropriate for all individual room acoustical parameters; see **Table 3** for β values. In this case as well, all models were better than the basic counterpart with sound source and random intercept (without room acoustical parameters). Based on AIC and BIC, the strongest to the weakest models were with C80, C50, EDT, STI, LF80, and D50. A similar model with L_{Aeq} instead of the individual room acoustical parameters was found to be weaker than all the other models and than the basic model.

5.5. Multiple Room Acoustical Parameters

Since room acoustical parameters are (partially) strongly and significantly correlated with each other (e.g., in many cases: Pearson's $r > 0.90$, $p < 0.01$), special care is needed when multiple room acoustical parameters are being considered in a model. Possible collinearities must be avoided. A series of models were fitted to the observed data considering sound source, random intercept, and multiple room acoustical parameters, with or without L_{Aeq}. Several models were found significant. While these models will not be introduced here in further details, it should be noted that adding more than two room acoustical parameters simultaneously typically did not improve the models any further.

5.6. L_{Aeq}

Generally, short-term acoustic comfort decreased with increasing L_{Aeq}. This confirms findings reported by Yang and Kang (2005). However, the picture was more complicated than this statement.

In the course of the further analyses, for each experiment, a model was fitted to the data considering sound source, random intercept, and L_{Aeq}. For Experiment 1, L_{Aeq} contributed significantly in the model, however, only in interaction with sound source. The following model was found to be appropriate for Experiment 1:

$$y_{ik} = \mu + \tau_{\text{Src},i} + \beta \cdot L_{\text{Aeq}} + \beta_{\text{Src},i} \cdot L_{\text{Aeq}} + \epsilon_{ik}. \quad (2)$$

In Equation (2), L_{Aeq} is the continuous variable L_{Aeq} and β is its regression coefficients. Subjects' random intercept was not found to be significantly contributing to the model. Model coefficients are shown in **Table 4**.

For Experiments 2 and 3, L_{Aeq} contributed significantly in the model, however, only without interaction with sound source.

TABLE 4 | Experiment 1: model coefficients (Coeff.), their 95% CI, and probabilities (*p*) of the linear mixed-effects model for acoustic comfort.

Parameter	Symbol	Coeff.	95% CI	*p*
Intercept	μ	0.400	[−4.018;4.818]	0.859
Sound source	$\tau_{\mathrm{Src},i}$ = Basketball 1	1.546	[−1.667;4.758]	0.345
	$\tau_{\mathrm{Src},i}$ = Doll's pram	3.235	[0.608;5.86]	0.016
	$\tau_{\mathrm{Src},i}$ = Conversation 1	−3.216	[−5.771;−0.661]	0.014
	$\tau_{\mathrm{Src},i}$ = Children 1	−2.747	[−5.245;−0.250]	0.031
	$\tau_{\mathrm{Src},i}$ = Children 2	0[a]		
L_{Aeq}	β	0.001	[−0.034;0.035]	0.961
Source×L_{Aeq}	$\beta_{\mathrm{Src},i}$ = Basketball 1	−0.068	[−0.136;0.001]	0.051
	$\beta_{\mathrm{Src},i}$ = Doll's pram	−0.090	[−0.143;−0.038]	0.001
	$\beta_{\mathrm{Src},i}$ = Conversation 1	0.041	[−0.009;0.091]	0.107
	$\beta_{\mathrm{Src},i}$ = Children 1	0.047	[−0.002;0.095]	0.058
	$\beta_{\mathrm{Src},i}$ = Children 2	0[a]		

The parameters and symbols are explained in Equation (2).
[a]Redundant coefficients are set to zero.

TABLE 5 | Experiment 2: model coefficients (Coeff.), their 95% CI, and probabilities (*p*) of the linear mixed-effects model for acoustic comfort.

Parameter	Symbol	Coeff.	95% CI	*p*
Intercept	μ	10.709	[8.513;12.904]	0.000
Sound source	$\tau_{\mathrm{Src},i}$ = Basketball 1	−4.197	[−4.586;−3.809]	0.000
	$\tau_{\mathrm{Src},i}$ = Baby	−2.100	[−2.260;−1.941]	0.000
	$\tau_{\mathrm{Src},i}$ = Conversation 2	−1.894	[−2.058;−1.729]	0.000
	$\tau_{\mathrm{Src},i}$ = Children 2	0[a]		
L_{Aeq}	β	−0.164	[−0.199;−0.129]	0.000

The parameters and symbols are explained in Equation (3).
[a]Redundant coefficients are set to zero.

TABLE 6 | Experiment 3: model coefficients (Coeff.), their 95% CI, and probabilities (*p*) of the linear mixed-effects model for acoustic comfort.

Parameter	Symbol	Coeff.	95% CI	*p*
Intercept	μ	6.646	[3.468;9.824]	0.000
Sound source	$\tau_{\mathrm{Src},i}$ = Basketball 2	−2.925	[−3.280;−2.570]	0.000
	$\tau_{\mathrm{Src},i}$ = Conversation 1	−1.524	[−1.678;−1.369]	0.000
	$\tau_{\mathrm{Src},i}$ = Children 1	0[a]		
L_{Aeq}	β	−0.107	[−0.162;−0.051]	0.000

The parameters and symbols are explained in Equation (3).
[a]Redundant coefficients are set to zero.

The following model was found to be appropriate for these two experiments:

$$y_{ik} = \mu + \tau_{\mathrm{Src},i} + \beta \cdot L_{\mathrm{Aeq}} + u_k + \epsilon_{ik}. \tag{3}$$

Model coefficients for the analysis of Experiments 2 and 3 corresponding to Equation (3) are shown in **Tables 5**, **6**.

While for the latter two experiments, for each sound source, acoustic comfort decreased with increasing L_{Aeq}, for Experiment 1, the picture was more complex (Taghipour et al., 2019c). Relatively unpleasant sound sources showed a similar pattern with increasing L_{Aeq} as for Experiments 2 and 3. However, acoustic comfort was slightly increased with increasing L_{Aeq} for the two pleasant children sounds. Taghipour et al. (2019b) offered a detailed discussion of the effect of L_{Aeq} on acoustic comfort and the implications of the differences between the mean L_{Aeq} of the three experiments.

6. DISCUSSION

6.1. Discussion and Implications of the Results

It was mentioned in Section 1 that, in the literature, the term "acoustic comfort" has been used with various definitions and been measured with different subjective and objective methods. Taghipour et al. (2019b) offered a discussion on the differences in approaches and measures related to this term. The short-term acoustic comfort here was rated subjectively on a bipolar 7-point scale (see Section 2.6). Compared to any other methods and definitions in the literature, this would be more comparable to the acoustic comfort used by Yang and Kang (2005) which was rated on a bipolar 5-point scale.

Moderate facade absorption was found to increase acoustic comfort. Both reflective and too absorptive facades were associated with low acoustic comfort ratings. While in Experiments 2 and 3, the absorption degree of the whole facade (beside glass windows) was varied systematically (i.e., based on α_w), in Experiment 1, different surfaces were either reflective or absorptive. This suggests that, in the design stage, both approaches could be useful: applying materials with moderate absorption characteristics (here middle-ranged α_w values) for the facade or using highly absorbing materials for only a selected portion of the facade. Furthermore, in the absence of any facade absorption, absorbing materials on the balcony ceilings tended to increase acoustic comfort on the balconies. That would be a simple and cheap solution which can also be applied post construction (Taghipour et al., 2019b). The results of this study seem to be generally in accord with findings in other studies which suggested use of absorbing materials on the facade and/or balconies (Lee et al., 2007; Hornikx and Forssén, 2009; Yeung, 2016; Calleri et al., 2017; Badino et al., 2019).

A dominant factor that influenced the acoustic comfort in the virtual inner yards was the sound source, i.e., the content of the sound present in the yard. While almost all sounds yielded a negative rating concerning the perceived acoustic comfort, relatively pleasant and relatively unpleasant sounds were found to increase and decrease acoustic comfort, respectively. Enabling facilities that invite relatively pleasant sounds, e.g., playing children as well as water features, birds and vegetation (Jeon et al., 2010; De Coensel et al., 2011; Taghipour and Pelizzari, 2019) and avoiding facilities which encourage relatively unpleasant sounds and noisy activities (such as basketball) might improve the overall acoustic comfort in inner yards. This point should be, however, treated with caution, due to inherent differences between short-term responses in a laboratory setup and long-term effects of the sounds in a living environment (Schäffer et al., 2016; Taghipour et al., 2019a).

A number of classic room acoustical parameters were found to be significant predictors of short-term acoustic comfort in linear-mixed effect models fitted to the observed data, including sound source, individual room acoustical parameters, and subject's random intercept. Only the echo criterion proposed by Dietsch and Kraak (1986) was not found to be a significant predictor of acoustic comfort. The room acoustical parameters investigated here are an initial set of acoustic indicators, however, not originally defined for acoustic scenarios in outdoor areas. The main purpose of the analysis in this paper was whether they could serve as indicators of acoustic performance in yards. The results support this hypothesis. In linear mixed-effect models with multiple room acoustical parameters, no more than two room acoustical parameters were found to be needed simultaneously. On the one hand, this suggests that the list presented here could be shortened and optimized. Hereby the results suggest that C50 (or C80), EDT, and STI would be the most important room acoustical parameters related to acoustic comfort. On the other hand, there might be other acoustic indicators which operate similarly or more successfully for this purpose. It might also be possible to define new parameters based on the results here and from similar future studies.

It was reported in Section 4 that—for the majority of room acoustical parameters used here—the statistical analysis was done with averaged room acoustical parameter data. That is, the average for the two octave bands centered at 500 and 1,000 Hz was chosen as representative for each room acoustical parameter, as recommended by ISO 3382-1 (2009). This suggestion, however, holds for performance spaces, not for outdoor living environments. In order to test whether this averaging was suitable in the case of the present study, an alternative averaging system has also been tested, whereby the average for the four octave bands centered at 250, 500, 1,000, and 2,000 Hz was used. The results of all statistical analyses were, however, stronger with the 500–1,000 averaged data. Therefore, only these analyses were reported in this paper.

Another set of established acoustic indicators are the so called psychoacoustic parameters. Rather than being based only on the objective physical characteristics of the acoustic situation, they are derived from subjective perception of sound by humans. The authors suggest to also investigate exploiting psychoacoustic parameters in future investigations to model acoustic comfort.

Acoustic comfort decreased slightly with increasing playback sequence (see Section 5.3). This might be because the sound exposure level (L_{AE}) increased with increasing playback sequence. That is, subjects were gradually exposed to a higher number of sounds, which increased the cumulative L_{AE}. It is reasonable to assume that higher (or cumulative) sound pressure levels are generally associated with lower acoustic comfort ratings (see Section 5.6 for a discussion on the effect of sound pressure level on acoustic comfort). This is consistent with other laboratory experiments, whereby "short-term noise annoyance" was reported either not to be significantly affected by increasing playback sequence or to increase with it (Schäffer et al., 2016, 2019; Taghipour and Pelizzari, 2019; Taghipour et al., 2019a). Hereby, it is noted that increased noise annoyance is typically associated with decreased acoustic comfort (Yang and Kang, 2005). For laboratory psychoacoustic experiments, the effect of the playback sequence indicates that a randomization of the playback list for the subjects—and counterbalancing where needed—is necessary, as carried out in the course of this study.

6.2. Limitations

The room acoustical simulation method used by ODEON is based on a geometrical approach with image sources and ray tracing. This method can have limitations, for example, regarding diffraction. Furthermore, while AuraLab is capable of a 3D playback over up to 15 loudspeakers, the 2D Surround sound ODEON output is limited to a horizontal plane at the ears' level. It should, however, be possible to use the B-Format ODEON output and decode the four-channel B-Format signal as a 3D scenario. This was not done in this study.

The experiments presented here did not include any visual stimuli. This made an investigation of the acoustical perception possible without any confounding effects of additional visual stimuli. Nevertheless, there is evidence for aural-visual interactions which might influence the overall perception (including comfort) in the laboratory (Viollon et al., 2002; Maffei et al., 2013a,b; Ernst and Bülthoff, 2014; Schäffer et al., 2019).

While interpreting the outcome of this study, one should consider the general differences between laboratory experiments and on-site experience and that most of the subjects of the three experiments work for authors' research institute. It should be further differentiated between short-term and long-term comfort.

The current study was carried out with single static sources in each stimuli. However, typically a mixture of (static and moving) sounds in background and foreground are present in reality. This limitation occurred partially because of computational limitations in ODEON and partially in order to reduce complexity. This should be improved in future studies.

The statistical models presented in this paper were fitted on the observed data. They investigated the relationship between the dependent variable acoustic comfort and a series of independent variables or alternatively the mediating variables room acoustical parameters. Therefore, the conclusions of this paper should not be generalized. In order to have a generic predictive model linking the room acoustical parameters to the acoustic comfort a larger amount of data would be needed and the predictive models would need to be further validated.

7. CONCLUSIONS

This paper investigated to what extent classic room acoustical parameters are suitable to predict perceived acoustic comfort in outdoor spaces of housing complexes. Subjective acoustic comfort ratings were collected in the course of three psychoacoustic experiments in the laboratory. The acoustic stimuli consisted of sounds from virtual inner yards of housing complexes. The analysis revealed that, beside the strong effect of the sound source (i.e., relative pleasantness or unpleasantness of the sound source), also L_{Aeq} and a series of room acoustical parameters could be used as predictors/indicators of acoustic comfort. In the design stage of housing complex projects, the estimated values for relevant room acoustical parameters could indicate the degree of subjective acoustic comfort. Thus, design changes which lead to an optimization (maximization/minimization) of estimated room acoustical parameters could be useful in improving acoustic performance. This should be helpful for architects, urban soundscape designers, and acousticians to improve the perceived acoustic comfort for the residents of housing complexes.

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available. Requests to access these datasets should be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Empa's Ethics Committee (Approval Nr. CMI 2018-194). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

This project was conceptualized and supervised by KE and AT and managed by AT. Experiments were designed by AT, KE, and TS and were carried out by TS and AT. Data analysis was done by AT and graphs were prepared by AT and TS. The analysis with room acoustical parameters was carried out by AT and SA. A first draft of this paper was conceptualized and written by AT, AG, and SA. All authors participated in reviewing the paper.

FUNDING

This project was an internally funded project of Empa (project *Innenhof*, assignment number 5213.00138).

ACKNOWLEDGMENTS

The authors are grateful to the participants of the preliminary and main experiments of this study and to people whose recorded voices were used in the experiments. The authors would also like to thank Matthias Blau for the co-supervision of the first experiment (Sievers et al., 2018) and Reto Pieren for preparing the listening test software. Helpful suggestions by Beat Schäffer and English proofreading by Lél Bartha are very much appreciated.

REFERENCES

Akaike, H. (1998). "Information theory and an extension of the maximum likelihood principle," in *Selected Papers of Hirotugu Akaike. Springer Series in Statistics (Perspectives in Statistics)*, eds E. Parzen, K. Tanabe, and G. Kitagawa (New York, NY: Springer).

Badino, E., Manca, R., Shtrepi, L., Calleri, C., and Astolfi, A. (2019). Effect of facade shape and acoustic cladding on reduction of leisure noise levels in a street canyon. *Build. Environ.* 157, 242–256. doi: 10.1016/j.buildenv.2019.04.039

Battaglia, P. (2014). "Achieving acoustical comfort in restaurants," in *Proceedings of Meetings on Acoustics*, Vol. 22 (Indianapolis, IN), 015001. doi: 10.1121/2.0000019

Brown, A. L., Gjestland, T., and Dubois, D. (2016). "Acoustic environments and soundscapes," in *Soundscape and the Built Environment*, eds J. Kang and B. Schulte-Fortkamp (Boca Raton, FL: CRC Press), 1–16. doi: 10.1201/b19145-2

Calleri, C., Shtrepi, L., Armando, A., and Astolfi, A. (2017). "On the influence of different facade materials on the auditory perception of urban spaces," in *INTER-NOISE and NOISE-CON* (Hong Kong), 27–30.

Calleri, C., Shtrepi, L., Armando, A., and Astolfi, A. (2018). Evaluation of the influence of building façade design on the acoustic characteristics and auditory perception of urban spaces. *Build. Acoust.* 25, 77–95. doi: 10.1177/1351010X18757353

Coelho, J. (2016). "Approaches to urban soundscape management, planning, and design," in *Soundscape and the Built Environment*, eds J. Kang and B. Schulte-Fortkamp (Boca Raton, FL: CRC Press), 197–214. doi: 10.1201/b19145-9

De Coensel, B., Vanwetswinkel, S., and Botteldooren, D. (2011). Effects of natural sounds on the perception of road traffic noise. *J. Acoust. Soc. Am.* 129, EL148–EL153. doi: 10.1121/1.3567073

Dietsch, L., and Kraak, W. (1986). Ein objektives kriterium zur erfassung von echostörungen bei musik-und sprachdarbietungen. *Acta Acust. United Acust.* 60, 205–216.

Echevarria Sanchez, G. M., Van Renterghem, T., Thomas, P., and Botteldooren, D. (2016). The effect of street canyon design on traffic noise exposure along roads. *Build. Environ.* 97, 96–110. doi: 10.1016/j.buildenv.2015.11.033

Elorza, D. (2005). *Room acoustics modeling using the raytracing method: implementation and evaluation*. (Ph.D. thesis), University of Turku, Turku, Finland.

Ernst, M., and Bülthoff, H. (2014). Merging the senses into a robust percept. *Trends Cogn. Sci.* 8, 162–169. doi: 10.1016/j.tics.2004.02.002

Georgiou, F. (2018). *Modeling for auralization of urban environments*. (Ph.D. thesis). Eindhoven University of Technology, Eindhoven, The Netherlands.

Gidlöf-Gunnarsson, A., and Öhrström, E. (2010). Attractive "quiet" courtyards: a potential modifier of urban residents' responses to road traffic noise? *Int. J. Environ. Res. Public Health* 7, 3359–3375. doi: 10.3390/ijerph7093359

Griesinger, D. (1997). The psychoacoustics of apparentsource width, spaciousness and envelopment inperformance spaces. *Acta Acust. United Acust.* 83, 721–731.

Hornikx, M. (2016). Ten questions concerning computational urban acoustics. *Build. Environ.* 106, 409–421. doi: 10.1016/j.buildenv.2016.06.028

Hornikx, M., and Forssén, J. (2009). Noise abatement schemes for shielded canyons. *Appl. Acoust.* 70, 267–283. doi: 10.1016/j.apacoust.2008.04.002

Hossam El Dien, H., and Woloszyn, P. (2004). Prediction of the sound field into high-rise building facades due to its balcony ceiling form. *Appl. Acoust.* 65, 431–440. doi: 10.1016/j.apacoust.2003.11.002

Hossam El Dien, H., and Woloszyn, P. (2005). The acoustical influence of balcony depth and parapet form: experiments and simulations. *Appl. Acoust.* 66, 533–551. doi: 10.1016/j.apacoust.2004.09.004

Houtgast, T., and Steeneken, H. J. (1973). The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acta Acust. United Acust.* 28, 66–73. doi: 10.1121/1.1913632

Iannace, G. (2018). The use of historical courtyards for musical performances. *Build. Acoust.* 23, 207–222. doi: 10.1177/1351010X16678219

IEC 60268-16 (2011). *Sound System Equipment – Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index*. Geneva: International Organization for Standardization.

ISO 11654 (1997). *Acoustics – Sound Absorbers for Use in Buildings – Rating of Sound Absorption*. Geneva: International Organization for Standardization.

ISO 12913-1 (2014). *Acoustics – Soundscape – Part 1: Definition and Conceptual Framework*. Geneva: International Organization for Standardization.

ISO 12913-2 (2018). *Acoustics – Soundscape – Part 2: Data Collection and Reporting Requirements*. Geneva: International Organization for Standardization.

ISO 3382-1 (2009). *Acoustics – Measurement of Room Acoustic Parameters – Part 1: Performance Spaces*. Geneva: International Organization for Standardization.

Jeon, J. Y., Lee, P. J., You, J., and Kang, J. (2010). Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds. *J. Acoust. Soc. Am.* 127, 1357–1366. doi: 10.1121/1.3298437

Kang, J. (2017). *Urban Sound Environment, 1st Edn*. Oxfordshire: Taylor & Francis Group.

Kang, J., and Zhang, M. (2010). Semantic differential analysis of the soundscape in urban open public spaces. *Build. Environ.* 45, 150–157. doi: 10.1016/j.buildenv.2009.05.014

Kuttruff, H. (2017). *Room Acoustics*. Boca Raton, FL: CRC Press. doi: 10.1201/9781315372150

Lee, P., Kim, Y., Jeon, J., and Song, K. (2007). Effects of apartment building facade and balcony design on the reduction of exterior noise. *Build. Environ.* 42, 3517–3528. doi: 10.1016/j.buildenv.2006.10.044

Maag, T., Bosshard, A., and Anderson, S. (2019). Developing sound-aware cities: a model for implementing sound quality objectives within urban design and planning processes. *Cities Health* 157, 1–15. doi: 10.1080/23748834.2019.1624332

Maffei, L., Iachini, T., Masullo, M., Aletta, F., Sorrentino, F., Senese, V., et al. (2013a). The effects of vision-related aspects on noise perception of wind turbines in quiet areas. *Int. J. Environ. Res. Public Health* 10, 1681–1697. doi: 10.3390/ijerph10051681

Maffei, L., Masullo, M., Aletta, F., and Di Gabriele M. (2013b). The influence of visual characteristics of barriers on railway noise perception. *Sci. Total Environ.* 445–446, 41–47. doi: 10.1016/j.scitotenv.2012.12.025

ODEON (2018). *ODEON Room Acoustics Software – User's Manual, v. 15*. Lyngby: Odeon A/S.

Öhrström, E., Skånberg, A., Svensson, H., and Gidlöf-Gunnarsson, A. (2006). Effects of road traffic noise and the benefit of access to quietness. *J. Sound Vib.* 295, 40–59. doi: 10.1016/j.jsv.2005.11.034

Onaga, H., and Rindel, J. H. (2002). Acoustic characteristics of urban streets in relation to scattering caused by building facades. *Appl. Acoust.* 68, 310–325. doi: 10.1016/j.apacoust.2006.01.010

Schafer, R. (1993). *The Soundscape: Our Sonic Environment and the Tuning of the World, 1st Edn*. Rochester, VT: Destiny Books.

Schäffer, B., Pieren, R., Wissen Hayek, U., Biver, N., and Grêt-Regamey, A. (2019). Influence of visibility of wind farms on noise annoyance - a laboratory experiment with audio-visual simulations. *Landsc. Urban Plan.* 186, 67–78. doi: 10.1016/j.landurbplan.2019.01.014

Schäffer, B., Schlittmeier, S., Pieren, R., Heutschi, K., Graf, R., and Hellbrück, J. (2016). Short-term annoyance reactions to stationary and time-varying wind turbine and road traffic noise: a laboratory study. *J. Acoust. Soc. Am.* 139, 2949—2963. doi: 10.1121/1.4949566

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136

Sievers, T., Eggenschwiler, K., Taghipour, A., and Blau, M. (2018). "Untersuchungen zur raumakustischen aufenthaltsqualität in innenhöfen von wohnbauten," in *DAGA* (Munich), 1038–1041.

Sturm, U., and Bürgin, M. (2016). *Stadtklang – Wege zu einer Hörenswerten Stadt (Band 1: Perspektiven)*. Zurich: Kompetenzzentrum Typologie und Planung in Architektur (CCTP); vdf Hochschulverlag AG.

Sturm, U., Bürgin, M., and Schubert, A. (2019). *Stadtklang – Wege zu einer Hörenswerten Stadt (Band 2: Klangraumgestaltung von Aussenräumen)*. Zurich: Kompetenzzentrum Typologie und Planung in Architektur (CCTP); vdf Hochschulverlag AG.

Taghipour, A., and Pelizzari, E. (2019). Effects of background sounds on annoyance reaction to foreground sounds in psychoacoustic experiments in the laboratory: limits and consequences. *Appl. Sci.* 9:1872. doi: 10.3390/app9091872

Taghipour, A., Pieren, R., and Schäffer, B. (2019a). Short-term annoyance reactions to helicopter and propeller-driven aircraft noise: a laboratory experiments. *J. Acoust. Soc. Am.* 145, 956–967. doi: 10.1121/1.5090500

Taghipour, A., Sievers, T., and Eggenschwiler, K. (2019b). Acoustic comfort in virtual inner yards with various building facades. *Int. J. Environ. Res. Public Health* 16:249. doi: 10.3390/ijerph16020249

Taghipour, A., Sievers, T., and Eggenschwiler, K. (2019c). "Room acoustical parameters as indicators of short-term acoustic comfort in inner yards," in *ICSV* (Montreal, QC: DEGA).

Tang, S.-K. (2005). Noise screening effects of balconies on a building facade. *J. Acoust. Soc. Am.* 118, 213–221. doi: 10.1121/1.1931887

Tang, S.-K. (2017). A review on natural ventilation-enabling façade noise control devices for congested high-rise cities. *Appl. Sci.* 7:175. doi: 10.3390/app7020175

Thomas, P., Aletta, F., Vander Mynsbrugge, T., Filipan, K., Dijckmans, A., De Geetere, L., et al. (2018). "Evaluation and improvement of the acoustic comfort in nursing homes: a case study in flanders, belgium," in *Euronoise 2018* (Crete), 27–31.

Viollon, S., Lavandier, C., and Drake, C. (2002). Influence of visual setting on sound ratings in an urban environment. *Appl. Acoust.* 63, 493–511. doi: 10.1016/S0003-682X(01)00053-6

Wang, X., Mao, M., Yu, W., and Jiang, Z. (2015). Acoustic performance of balconies having inhomogeneous ceiling surfaces on a roadside building facade. *Build. Environ.* 93, 1–8. doi: 10.1016/j.buildenv.2015.06.027

WHO (2018). *Environmental Noise Guidelines for the European Region*. Copenhagen: WHO Regional Office for Europe.

Xiao, J., and Aletta, F. (2016). A soundscape approach to exploring design strategies for acoustic comfort in modern public libraries: a case study of the library of birmingham. *Noise Mapp.* 3, 264–273. doi: 10.1515/noise-2016-0018

Yang, H.-S., Kim, M.-J., and Kang, J. (2013). Acoustic characteristics of outdoor spaces in an apartment complex. *Noise Control Eng. J.* 61, 1–10. doi: 10.3397/1.3702001

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

Yeung, M. (2016). "Adopting specially designed balconies to achieve substantial noise reduction for residential buildings," in *INTER-NOISE and NOISE-CON* (Hamburg), 21–24.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

The first part of the questionnaire included six questions about the subjects' hearing (4 questions) and well-being (2 questions). The answers to these questions were used as inclusion/exclusion criteria. The data of no subject had to be excluded from the final analysis. The questions were:

- How good is your hearing? (5-point bipolar scale)
- Do you suffer from any hearing loss, hearing illnesses or ear noises (e.g., tinnitus)? (Y/N)
- Do you use a hearing aid? (Y/N)
- Do you have a cold at the moment? (Y/N)
- Are you feeling healthy and well? (Y/N)
- Are you feeling very tired? (Y/N)

frontiers in Psychology

ORIGINAL RESEARCH
published: 06 May 2020
doi: 10.3389/fpsyg.2020.00737

Arousing the Sound: A Field Study on the Emotional Impact on Children of Arousing Sound Design and 3D Audio Spatialization in an Audio Story

Francisco Cuadrado[1], Isabel Lopez-Cobo[1], Tania Mateos-Blanco[2] and Ana Tajadura-Jiménez[3]*

[1] Communication and Education, Universidad Loyola Andalucía, Seville, Spain, [2] Department of Theory and History of Education, and Social Pedagogy, Universidad de Sevilla, Seville, Spain, [3] DEI Interactive Systems Group, Department of Computer Science and Engineering, Universidad Carlos III de Madrid, Madrid, Spain

Sound from media increases the immersion of the audience in the story, adding credibility to the narration but also generating emotions in the spectator. A study on children aged 9–13 years ($N = 253$), using an audio story, investigated the emotional impact of arousal vs. neutral treatment of sound and 3D vs. stereo mix spatialization. The emotional impact was measured combining three different measures: physiological (Electrodermal activity), self-report (pre-post exposition), and richness of mental images elicited by the story (using Think-aloud technique after exposition). Results showed higher emotional impact of the arousal and 3D audio conditions with different patterns according to the age of the participants and distinctive types of interaction when both variables were combined.

Keywords: audio-story, sound-design, 3D-sound, emotion, immersion, imagery

OPEN ACCESS

Edited by:
PerMagnus Lindborg, City University of Hong Kong, Hong Kong

Reviewed by:
Lonce Wyse, National University of Singapore, Singapore
Stefania Serafin, Aalborg University Copenhagen, Denmark

****Correspondence:***
Francisco Cuadrado
fjcuadrado@uloyola.es

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

Received: *10 December 2019*
Accepted: *26 March 2020*
Published: *06 May 2020*

Citation:
Cuadrado F, Lopez-Cobo I, Mateos-Blanco T and Tajadura-Jiménez A (2020) Arousing the Sound: A Field Study on the Emotional Impact on Children of Arousing Sound Design and 3D Audio Spatialization in an Audio Story. Front. Psychol. 11:737. doi: 10.3389/fpsyg.2020.00737

INTRODUCTION

For the majority of us, sounds are present in many aspects of our lives. They accompany many of our actions, such as opening a door or walking down the stairs; they signal the presence of other individuals, animals, objects, and even environmental events, such as a thunderstorm; and, through reflections in walls and other surfaces, sounds provide information of the geometry of the space we are in. Our auditory system, as the rest of our sensory systems, has evolved to monitor the surrounding environment, obtain information, and alert us of significant events so that we can adapt our behavior accordingly and keep safe (Graziano, 2001). In this respect, the auditory system is especially good at detecting changes and quickly orienting our behavior toward them; often, the auditory system acts faster than, for instance, vision does (McDonald et al., 2000). For this reason, the auditory system is often known as "a warning system" (Juslin and Västfjäll, 2008).

Hearing a sound will offer trigger an emotional response in listeners. Indeed, sounds can elicit a full range of emotional responses in listeners (Bradley and Lang, 1999, 2000; Juslin and Västfjäll, 2008). People can be startled by a sudden scream in the middle of the night, annoyed by the traffic noise, pleased by a bird song, or thrilled by hearing football crowds cheering. Emotional responses produce changes in our physiological state, behavior, and feelings, getting our body ready for action (e.g., Levenson, 1994; LeDoux, 1998; Seth, 2013).

Previous research investigating emotional responses to sound has mostly focused on trying to connect physical sound attributes, such as intensity, frequency, or the time structure of the sound signal (Schirmer et al., 2016), with basic emotional responses. For instance, equal pleasantness contours for tones varying in frequency and intensity have been developed (Todd, 2001), and a few studies have suggested a correspondence between sound intensity and emotional arousal since increasing loudness results in an increase in the orienting response (e.g., Sokolov, 1963; Lang et al., 1990). Also, correspondences between sound clarity (a parameter directly connected to the amount of high frequency in a sound) and the emotional valence has been found (Cho et al., 2001). Nevertheless, other studies have evidenced that looking at physical properties alone cannot fully capture emotional responses to sounds. For instance, the study by Landstrom et al. (1995) showed that only around 20% of noise-induced annoyance related to physical characteristics of the noise (see also Bjork, 1999; Bradley and Lang, 2000). In another study, everyday sounds were used (e.g., a cow and a rollercoaster), but the identification of these sounds was impaired by using a neutralization algorithm that preserved the physical properties of the sounds; this was done in order to show that only 20–25% of the emotional responses to these everyday sounds depended of the physical properties of the sound (Asutay et al., 2012).

The studies above have suggested that listeners do not react emotionally just to acoustic waves but also to sound sources and sound events, and that emotional responses to sound depend on the interpretation and meaning (i.e., relevance) the listeners attribute to these particular sound sources and events (Jäncke et al., 1996; Gygi, 2001; Juslin and Västfjäll, 2008; Tajadura-Jiménez, 2008). Furthermore, this interpretation is the result of an interaction between the sound itself, the context of when and where sound is heard, and the listener (Blauert and Jekosch, 1997; Jekosch, 1999). Therefore, when studying emotional responses to sound, it is important to consider other variables apart from the physical properties of the sound. These variables relate to whether sounds can be identified as objects or events (Jäncke et al., 1996; Bradley and Lang, 2000; Asutay et al., 2012); the context, such as the events that preceded the sounds, the presence of other multisensory events, or the space where the sound is heard (Västfjäll et al., 2002; Tajadura-Jiménez et al., 2010; Berger et al., 2018); and the individual differences of listeners. The same sound may be interpreted in a substantially diverse manner by different listeners; listeners may vary in their previous experiences, expectations, personality traits, or individual goals (Grimshaw, 2014); therefore, the same sound may elicit a rather different emotional response in different listeners (social and cultural memory; Tajadura-Jiménez, 2008).

All these changes affect attention, cognitive, and perceptual processes (De Gelder and Vroomen, 2000) and influence our judgments and decisions (Peters et al., 2006). As a result, it is common 4 that sound is used in products or media applications in order to transmit information, grab the attention of users, or influence their attention. Sound and music from different types of media products (film, TV series, documentaries, podcasts, and videogames) contribute to the success of the audience experience, adding credibility to the created story, making the narration more understandable, and also generating emotions in the spectator. Thanks to the veridiction pact (Zunzunegui, 1995) and the semi-conscious perception of sound (Murch, 2001), sound design has the power to increase the immersion and participation of the audience in the story.

Regarding audio-visual media products, different studies have analyzed the emotions elicited by sound as part of a media narration, considering the presence or absence of sound and musical narrative elements. These studies have found a greater response in EDR (electrodermal response), heart rate, and temperature in stimuli with sound effects compared to silence (Shilling et al., 2002); also exhibited were a significant increase in EDA and questionnaires in stimuli with sound compared to silence (Scorgie and Sanders, 2002). Also, a better performance in the accomplishment of a task (driving game) has been achieved when the music is selected by the participant. Taking into consideration the diegetic vs. non-diegetic approach, Grimshaw (2008) found that diegetic sound provides a higher level of immersion, while music increases immersion and reduces tension and negative affect.

Focusing on the audio-only kind of media products, such as radio programs, podcasts, or audio narrations, one of the few studies with children in this field concluded that the use of narration, character's direct voice, and sound effects in an audio story generated more enjoyment, attention, and positive emotional impact in children aged 3–4 years (Ritterfeld et al., 2005). A slightly different approach in the emotional impact of media sound is the consideration of the relationship between the narration (voice over or dialogues), sound effects, and the use of different sound shots (the placement of sound in several distances from the listener perspective). In a study focused on the analysis of mental images and attention level in sound fictional stories, Rodero (2012) compared four versions of the same stimulus: (1) narration, (2) narration with sound effects, (3) narration with different sound shots, and (4) narration with sound effects and different sound shots. Results showed a higher level of creation of mental images and attention in stimulus with sound effects vs. stimulus without sound effects. Furthermore, the use of different sound shots in the narration also derived a higher level of creation of mental images and attention compared to narration without the use of sound shots. Finally, the highest level of creation of mental images and attention was found in stimulus that included narration, sound effects, and the use of different sound shots.

As these last findings suggest, space and spatial localization of sound is one of the key elements that increase the immersion and emotional impact on the listener (Murphy and Pitt, 2001). These findings are consistent with Steele and Chon (2007), who found that the spatial location of a virtual sound object, although currently limited in terms of game implementation, has a significant potential related to the emotions.

A further key element related to the listening experience and, more specifically speaking, to the spatial dimension of sound is that the choice of headphones or speakers could be a significant contextual variable (Cox, 2008; Hong et al., 2017), particularly in terms of location and immersion (Grimshaw, 2007) and emotional impact. In a comparable study, Murphy and Pitt (2001) showed a preference for the use of headphones, arguing that it "... allows the designer to incorporate more complex sound objects whose subtleties will not be lost due to background noise, speaker conversation, etc." Headphones seem to produce a more immersive experience, and the commercial availability of a wide range of headphones (many designed specifically for computer games) suggests that the use of headphones is common in a player's natural environment (LaGrou, 2014). These studies provide evidence of how the spatial dimension of sound, in this case related to the use of headphones, may impact on immersion and emotional impact.

OBJECTIVE AND HYPOTHESES

The aim of the present study was to investigate the potential impact of "emotionally marked" sound effects and of 3D spatialization on emotional responses and quality of mental images elicited in children when listening to an audio story. According to Valkenburg and Beentjes (1997), a story presented in auditory form is expected to stimulate imagination and fantasy in children.

This study is part of the research project "Unconscious listening," which is focused on the analysis of the emotional impact of sound in children and its possibilities to increase and improve learning in the scholar environment. As stated by Ritterfeld et al. (2005), audio stories might support cognitive and emotional development in children. Following the research design of the "Unconscious listening" project, the study focused on Primary and Secondary Education children. Although no references have been found in previous studies about differences in the emotional impact of sound in children from distinctive ages, this has been considered in the present study, according to the various educational level of participants.

According to the previous findings, several hypotheses were formulated:

H1: the use of "emotionally marked" (i.e., arousing vs. neutral) sound in the design and production of a sound story will elicit more intense emotional responses in the listener compared to a sound story without this emotional manipulation in its design and production.

H2: the use of "emotionally marked" sound in the design and production of a sound story will generate richer and more detailed mental images in the listener than a soundtrack without this emotional intention.

H3: a soundtrack mixed in 3D sound format will elicit a more intense emotional response in the listener compared to a soundtrack mixed in stereo.

H4: a soundtrack mixed in 3D sound format will generate a greater number of mental images as a well as richer and more detailed mental images in the listener compared to a soundtrack mixed in stereo.

H5: the emotional impact elicited by both emotional marked sounds and/or 3D sound mix format takes place at an unconscious level and therefore will not be reported by the listener.

H6: the emotional impact and the mental images generated by both emotional marked sounds and/or 3D sound mix format exhibit different effects according to the educational level of the participants.

METHOD

Sample Description

The participant sample consisted of 253 children from two schools in Seville: Ntra. Señora del Águila (SSAA) and San José SSCC (PPBB). The participants were students from two educational levels: 128 participants from 4th Primary Education (9–10 years old) and 125 participants from 1st of Secondary Education (12–13 years old). Once the participation of each school in the project was agreed with schools' administrators, the sample selection was made by voluntary participation of students from the different school classes.

Ethical Implications

Participation in the project did not involve any physical or psychological risk to participants. Participants and their parents were conveniently informed about the whole project, and they signed informed consent forms before taking part in the study. All information collected followed the necessary protocols to safeguard the privacy and confidentiality of participants. The collected data were only used for the purposes of this research; the data were also protected so that only researchers could access it. The experiment was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki, as revised in 2008, and approved by the Ethics Committee of the Universidad Loyola Andalucía.

Stimuli

The stimulus consisted of a sound-only story (similar to a fictional radio story), based on an existing written story suitable for children between 9 and 14 years old: "*Los cohetes tienen forma de flauta*" ("*The rockets are flute-shaped*"). Due to the length of the whole tale, the two first chapters were selected to create the stimulus, resulting in a story length of 1860 words. The original text was adapted to produce a radio story. The adaptation basically consisted of increasing the number of character interventions and dialogues and reducing the amount of voiceover narration. According to Rodero (2012) a dramatized story generates a greater level of imagery and involvement in the listener compared to a narrated story. The narration and dialogues voices were recorded in a studio: a professional voice-over actor performed the role of the narrator, while two children with acting experience (a boy and a girl, aged 11 and 9, respectively) performed the role of the two characters of the story: Salva, a 10-year-old boy, and his 8-year-old sister, Elena.

According to the previous findings in research literature and the study hypothesis, the design of the stimuli centered around two independent variables: arousal level of sound design and sound spatialization. The arousal level of sound design was developed mainly through sound effects and ambiences. Two sound design proposals were elaborated: neutral and arousal marked. The sound treatment applied to every condition was based on very specific and subtle modifications of sound instead of on the presence or absence of certain types of sound (presence or absence of sound effects, dialogue, or music), which has been the approach of previous studies. The neutral condition consisted of sound effects and ambiences that, according to the description of the story, the locations, and the characters' action, movements, and dialogues, could be heard in a real-world situation. Also, a global equalization was applied to all the neutral sound effects and ambiences tracks, rolling off frequencies below 13 Hz and over 5.6 KHz (in both cases with 12 dB/octave slopes), in order to subtly reduce the clarity of the sound (related to a higher emotional elicitation, Cho et al., 2001) and the low-frequency impact.

The arousal-marked condition combined two distinctive approaches—sound parameter modification and sound source modification—which resulted in different procedures. Following the first approach, in certain cases, the same sound effects or ambiences from the neutral version were used but with changes in certain sound parameters. The modifications included equalization, modifying the high or the low frequencies of the sound in each case; changes in pitch to increase the clarity of the sound or to create a sensation of movement within the sound; changes in the loudness of specific sound effects; and added reverberation to increase the spaciousness of a specific sound. Following the second approach, some of the neutral sound effects and ambiences were substituted by others, looking for sound elements that supported a movement or an action, which illustrated a description or enriched a location. Special attention was given to the fact that both versions should include the same amount of sound elements, in the same moments of the story in order to provide to all the listeners a comparable sound story (avoiding the risk of obtaining different results based solely on the presence or absence of elements).

Furthermore, the strength of the variances in sound treatment between conditions was focused on moments of the story where it could be more narratively and dramatically effective, according to the development of the action throughout the story. Regarding arousal treatment, there were specific moments of the story that were identified to allow for a clearer difference between the sound design in the neutral vs. arousal conditions. Sound excerpts of these moments have been attached to this article, in both versions (neutral and arousal marked), as examples of the treatment applied in each case. These moments have been included here.

- "Meteor fall" (**Supplementary Audios 1**, **2**: meteor_fall_neutral.mp3 and meteor_fall_arousal.mp3): the neutral version of the stimulus included a "woosh" sound to illustrate this fall. The arousal version included a denser "woosh" sound, with more high and low frequency components, as well as added reverberation.
- "Tic-tac": While Salva says *"the time is relative,"* the neutral condition used the sound of the footsteps of the boy in the room. The arousal version used the sound of a reverberant tic-tac sound that, at the end of the sentence, becomes progressively slower until it stops.
- "Chronometer" (**Supplementary Audios 3**, **4**: chronometer_neutral.mp3 and chronometer_arousal.mp3): while Elena plays the music on the flute, the sound of a chronometer counts the time she takes to play it. The neutral condition included the normal sound of a chronometer with no modifications. The arousal version modified this same sound, increasing the pitch and speed of the sound to make it higher pitched and to synchronize the tempo of the tic-tac sound with the tempo of the music Elena plays).
- "Down" effect: This effect is introduced to finish the moment in which Salva imagines the speed of his sister playing the flute while she is on a skateboard, and the narrator says *"the experiment was a total failure."* In the neutral condition, the usual sound of children movement was used. In the arousal condition, the whole sound ambience that is listened while Salve speaks is pitched down until all the sounds disappear in a very low frequency register.

The sound treatment of most of these moments is consistent with previous studies, focusing in the modification of the parameters volume (Sokolov, 1963; Lang et al., 1990) and frequency spectrum (Cho et al., 2001), where it has been reported a higher emotional impact of sounds with great amount of high-frequency content.

Figure 1 shows an excerpt of the sound design script used during the stimuli production, specifically from one of the selected moments for arousal intervention: "meteor fall."

Figures 2, **3** refer to the same timeline period (the first minute of the story). These figures display two screen captures of the tracks, sound effects, and ambiences used for each condition: neutral (**Figure 2**) and arousal (**Figure 3**). The number of audio tracks and the amount of sound layers are greater in the arousal version. For instance, a combination of different sound effects was used to recreate the gabble of children when leaving school. Also, two distinctive ambiences were layered to design a richer sound that recreated the environment of the village where the characters live.

Regarding sound spatialization, two versions of the stimulus were prepared, one mixed in stereo format and the other one mixed in surround 3D format. Both versions were produced to be listened through headphones. Nuendo 7.1 (from Steinberg) and Spatial Audio Designer (from New Audio Technology) were used to produce both mixes. The same recorded dialogue tracks were used in both mixes, keeping the volume and the clarity between the stereo and the 3D audio versions. The creation of the soundtrack followed the usual sound design and postproduction processes involved in film and media sound production: Foley effects recording and sound effects libraries were the main sound sources. All the ambiences and sound effects were edited using Nuendo 7.1. Different equalization, dynamic, and modulation plugins were used to process the sound, including several types of reverberation to recreate accurately the singular spaces represented in the story. From all the edited audio material (dialogues, sound effects, and ambiences), two separate mixes were produced. The stereo mix used the left-right panning, as well as volume, equalization, and reverberation to simulate the different spaces, position, and movement of sound sources. The 3D sound mix was produced with a 9.1 surround configuration (compatible with Dolby Auro 3D systems): five main channels, four elevated channels, and one LFE channel. Apart from the use of volume, equalization and reverberation, the spatial recreation was achieved through the movement of the sound objects in the 3D audio space (combining the front–back, left–right, and up–down axis). All the mixes (stereo and 3D) and parameter automation were done using the Spatial Audio Designer plugin, keeping the same peak and RMS levels between the two mixes, controlling in detail the clarity and understandability of dialogue in both versions. The final mixes in both formats were produced using the Headphone Surround 3D technology from the Spatial Audio Designer software. This technology consists of a binaural

TAKE 01.04		
Voice	SFX Neutral	SFX Arousal
Salva: "And of course, as the meteorite was very large, it filled everything with dust, like when Mama shakes the carpets, and you can see nothing. Also, it gave a bit of disgust to the dinosaurs, I think.	• Mama shaking the carpet + footsteps	• Meteorite whoosh with great 3D spatial movement of the sound. Dinosaurs' roaring, with added reverberation

FIGURE 1 | Excerpt of the sound design script.

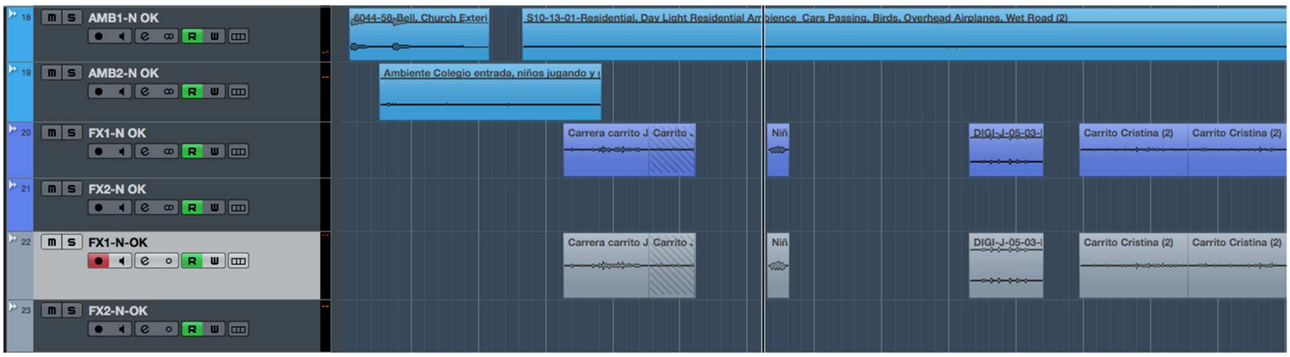


FIGURE 2 | Screen capture showing the sound effects and ambiences used in the Neutral version for the first minute of the story.

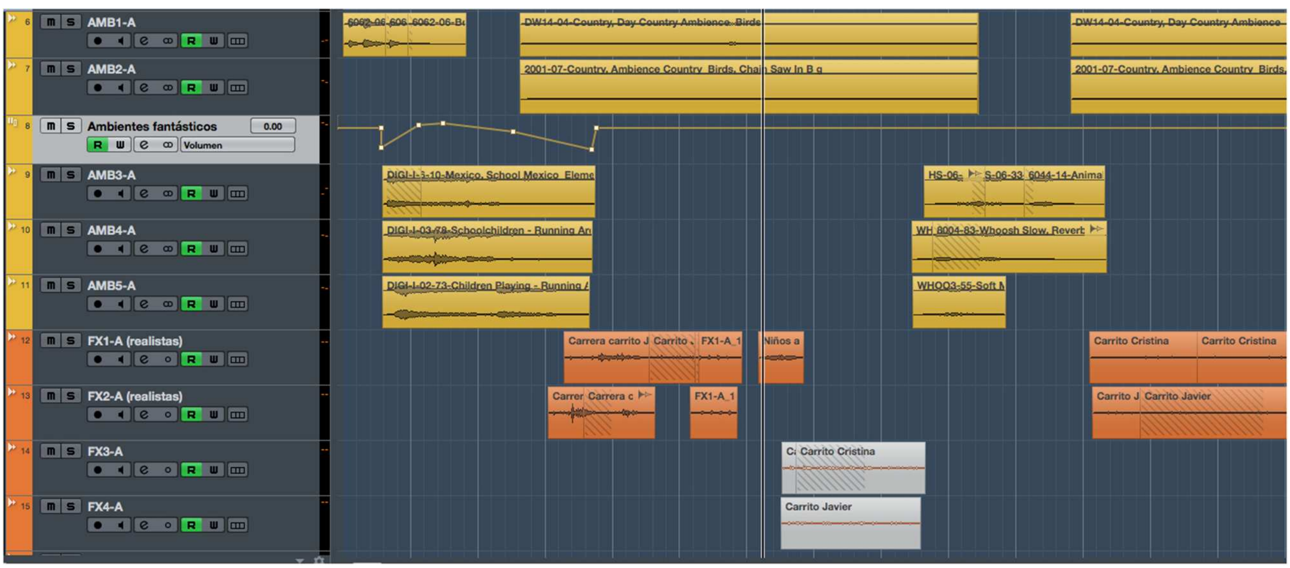


FIGURE 3 | Screen capture showing the sound effects and ambiences used in the Arousal version for the first minute of the story.

simulation of various mixing studios with multiple virtual loudspeaker arrays; this makes it possible to produce 3D mixes using all three dimensions and different locations and to listen to these 3D mixes using a pair of stereo headphones. **Figure 4** shows a screen capture of the 3D sound mix configuration and sound object position in one of the specific moments of the story

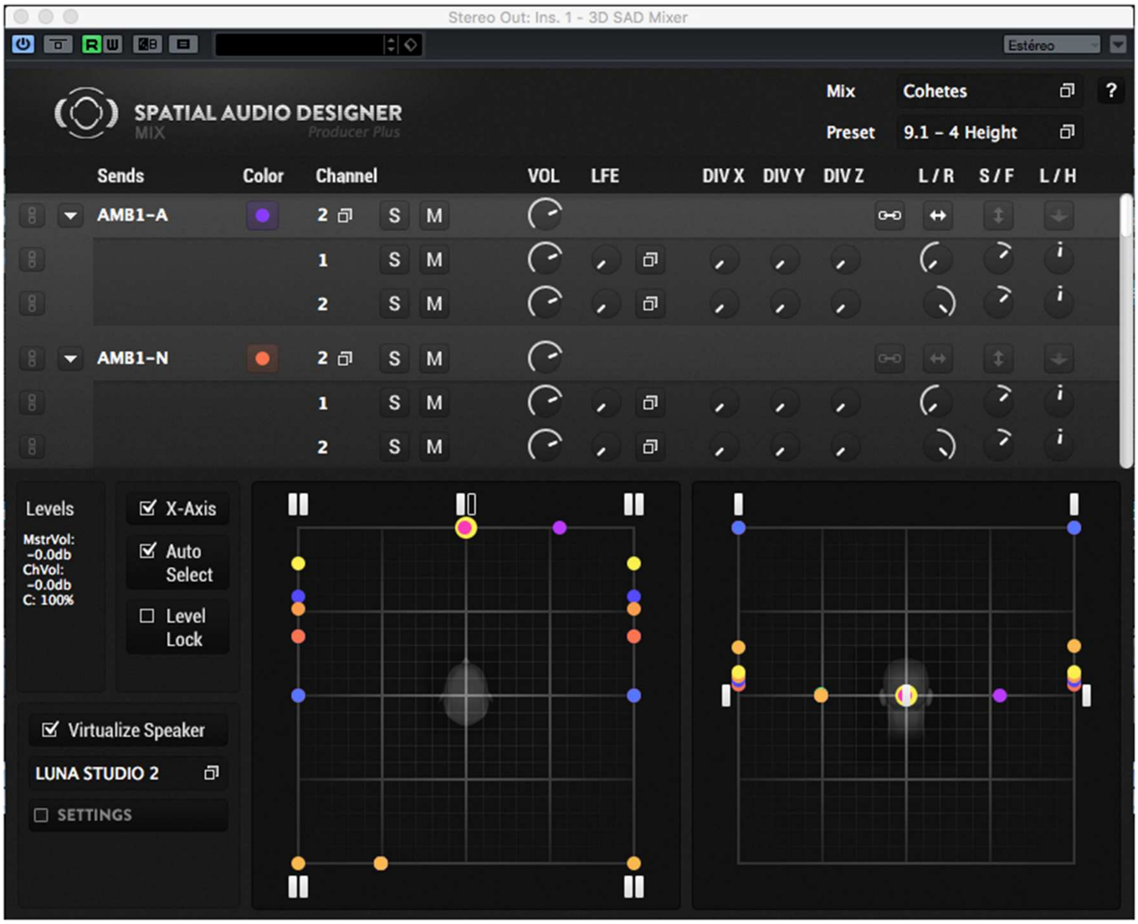


FIGURE 4 | Screen capture of the 3D sound mix state in a specific moment of the story when sound elements are positioned and moved across the 3D field (horizontal and vertical).

when sound elements are positioned and moved across the 3D field (horizontal and vertical). Each dot corresponds to a separate sound object (specific sound effects and ambiences). The master audio files produced with this software are compatible with any .wav or .mp3 file format player, and the 3D spatialization can be listened in any device using a pair of normal stereo headphones.

Although, in the 3D mix conditions, the whole sound was mixed using this spatial conception of space, there were also specific moments in the story that allowed for a clearer movement of sound through the three dimensional space. Sound excerpts of these moments have been attached to this article, in both versions (stereo and 3D), as examples of the treatment applied in each case. It is recommended to listen to the examples with headphones in order to notice the differences in the spatialization:

- "Salva dressing as a scientist" (**Supplementary Audios 5**, **6**: scientist_st.mp3 and scientist_3D.mp3): Salva moves around the room to look for elements to dress himself like Albert Einstein. The 3D treatment combines the use of footsteps and the dialogue of Salva moving around the 3D space (following a specific path from center to left, to the back side of the room, to the right back of the room, and coming back to the front) together with specific sound effects (moving objects, manipulating boxes and household items, etc.) located in several specific places in the 3D space where Salva "stopped" to look for more elements to be use in his dressing.
- "Sad day for light-speed": Salva speaks about how his sister is going to surpass light-speed with her flute performance over the skateboard. The 3D condition included a reverberant three-dimensional sound ambience that included the sound of a tower bell, thunder, the clapping and shouts of a large audience, and a high pitch "woosh" moving from the back left to the front right of the virtual sound space.
- "Skate-fall" (**Supplementary Audios 7**, **8**: skate-fall_st.mp3 and skatefall_3D.mp3): Elena throws on the ramp with the skateboard while trying to play the flute. Different sound elements were located and moved through the 3D space: the sound of Elena screaming from left back to right front, the shouts of Salva from the very back left bottom; an intense wind sound and the rolling of the skateboard, also moving

in the same direction of Elena's voice; and the final crash at the front right.

The sound treatment of most of these moments is also consistent with previous studies that consider spatial localization of sounds to be one of the key elements that increase the immersion and emotional impact on the listener (Murphy and Pitt, 2001; Steele and Chon, 2007). Furthermore, the use of binaural simulation of surround sound through headphones coincides with the conclusions of different authors (Murphy and Pitt, 2001; Grimshaw, 2007; Cox, 2008; Hong et al., 2017) about the increase of the location of sound sources and its relation to the immersion and emotional impact.

The combination of both variables (neutral vs. arousal and stereo vs. 3D) resulted in four distinctive sound conditions:

- Sound condition A: neutral sound design + stereo mix
- Sound condition B: neutral sound design + 3D sound mix
- Sound condition C: arousal sound design + stereo mix
- Sound condition D: arousal sound design + 3D sound mix

All the final stimuli were produced in wav file format with a 44.1 Khz sample rate and 16-bit depth to retain the standard of most consumer audio file formats.

Measures

The emotional impact was measured by combining three different approaches: the physiological response (electrodermal response during exposition to the stimulus), self-reported emotional state (Self-Assessment Mannequin test—SAM, pre- and post-stimulus, and an immersion questionnaire, post stimulus) and mental images elicited by the story (specific questions and verbal expression of participants, using the Think-aloud technique after exposition). All the technical details of the equipment used are presented as an appendix at the end of this article (**Appendix 1**: Apparatus).

Below are the three different approaches:

- Physiological response: the electrodermal response (EDR) linked to emotional arousal or emotional intensity (Boucsein, 2012; Venables and Christie, 1973) was measured, providing the EDR per second for each participant.
- Self-reported emotional state: the Self-Assessment Mannequin test [SAM (Bradley and Lang, 1994)] was used to register the self-perception of each participant's emotional state. The SAM test is a picture-based instrument that measures three dimensions of a perceived emotion: valence (positive–negative), arousal (passive–active), and dominance (dominated–dominant). Only the two first dimensions (valence and arousal) were used in the study. The test was administrated to each participant before and after the exposition to the stimulus. The instrument was specially designed for this study to make possible that participants fulfilled the test using the same touchscreen of the tablet in which all the other activities were programmed (see the "Procedure" sub-section).
- To measure the mental images elicited by the story and the verbal expression of these emotions, a series of questions were designed to guide the participant through the exploration and verbalization of these mental images. Some questions were designed to be answered choosing an option (closed questions), while other questions were designed to register the verbal expressions of participants, using the Think-aloud technique (voice recording of participant's speech, in this case, as answers to specific questions).

The questions proposed to participants are shown in **Figure 5**.

First, we tested whether the distributions of the obtained data were normal using the Shapiro-Wilk test. None of the variables passed the normality test. Nevertheless, Q-Q plots showed moderate deviations from normality. Given that parametric statistical tests (ANOVAs) are quite robust to moderate deviations from normality (e.g., McDonald, 2014), in our analyses we opted for the use of both non-parametric Kruskal-Wallis tests for the four sound conditions, and the Analyses of Variance (ANOVA) tests allowed for the testing of the interactions between the factors sound emotion condition (Neutral, Emotional) and sound spatiality condition (stereo, 3D). The analyses were carried out using the SPSS software, version 24. Significant effects were followed by Mann-Whitney analyses (non-parametric tests of independent samples).

A qualitative analysis of the data obtained through Think aloud was carried out by three independent analysts. A total of 1,152 recordings were collected from the students who answered the questions they were asked. An analysis of the content of all the collected information was made, and an emergent categorical system was drawn up by the research team in which several dimensions of analysis were identified: (1) representation of the story, (2) elicited emotions, (3) mental images, and (4) immersion.

To develop the descriptive phase of the analysis, a total of eight categories and 25 subcategories were defined, identified from the described events and emotional states expressed. For the coding and subsequent qualitative analysis of the collected data, Nvivo 11PRO software was used, providing the open coding (Flick, 2012) of the information. Subsequently, an analysis of the frequency of references made in the responses to each of these categories and subcategories was carried out, and, based on these data, the interpretative phase of the analysis was carried out.

The variables contemplated in the analysis of the different measures were the sound condition (A, B, C, or D) and educational level (to consider age differences).

Procedure

The field study was conducted over 2 weeks during school hours. A special classroom was prepared to give all the participants with enough space so as not to disturb each other while listening to the story, answering the questionnaires, or recording their voices during the Think-aloud tasks. The four versions of the stimulus (corresponding to the four sound conditions) were randomly assigned to each participant, and so none of them were aware of the listened version of the story. The version, gender, educational level (age), and school distribution of the stimuli is detailed in **Table 1**.

Question	Answer options
1. What moment in the story did you find most exciting? Why?	Think-aloud
2. How have you felt at that moment?	Think-aloud
3. Try not to think too much, and speak as things come to your mind: what have you imagined when Elena was going down the school hill on the scooter?	Think-aloud
4. While you were listening to the story, have you imagined what Salva looks like?"	Yes/No
5. Describe how you have imagined Salva	Think-aloud
6. The images that have come to your mind while you were listening were…	Semantic differential o Reals – Fictitious o Clear – Unclear o Exciting – Boring o Known – Unknown
7. After listening to the story … images came to my mind	Multiple option question: o A lot of o Some o A few

FIGURE 5 | Questions and type of answer options for the mental images register.

RESULTS

Effects on Physiological Arousal (EDR)

The measurement tool analyzed phasic activity related to emotion, i.e., the electrodermal response (EDR). The measurement unit was the electrodermal resistance in Kiloohms (KΩ) of each participant. All participants were exposed to a conditioning stimulus before the exposition to the studio stimulus with the purpose of accommodating them to the listening conditions and also to establish an individual baseline in the EDR response for each participant. Al collected data were preprocessed, subtracting the individual baseline level to all the measures for each participant. For the analysis of EDR, an initial ANOVA on the mean EDR values during the stimulus duration with sound condition (A—Neutral stereo, B—Neutral 3D, C—Emotional stereo, and D—Emotional 3D) as between-subjects variable was conducted. This analysis did not yield any significant results ($p > 0.05$). Furthermore, there were no significant differences in EDR between groups, as confirmed by a Mann-Whitney test ($p > 0.05$).

A subsequent analysis was conducted that looking only at the moments with special sound manipulations according to what has been exposed in the stimulus subsection:

- Meteor fall
- Dressing as a scientist
- Tic-tac
- Chronometer
- Sad day for light-speed
- "Down" effect
- Skate-fall

For each of the stimuli in **Table 2**, the maximum EDR value during the stimulus duration was calculated (Martin and Venables, 1980; Boucsein, 2012). Stimuli were all longer than or exactly 5 s long (note that, according to Edelberg, 1967, the EDR may be extended up to 5 s after the onset of the stimuli).

The peak EDR values for each course and sound condition (A—Neutral stereo, B—Neutral 3D, C—Emotional stereo, and D—Emotional 3D) are displayed in **Table 2**. Peak EDR values were used as dependent variables for a Multivariate Analyses of Variance (MANOVAs) with between-subject factors course (4th EP and 1st ESO), sound emotion condition (Neutral and Emotional), and sound spatiality condition (stereo and 3D). Wilks' Lambda was used as the multivariate criterion. The results of the multivariate test revealed that there was a non-significant tendency, indicating an interaction between sound emotion condition and sound spatiality condition [$F_{(7, 238)} = 1.90$, $p = 0.07$, Wilks' Lambda $= 0.947$]. This interaction is explained by the results showing that, while for the neutral conditions there was an increase in EDR peak value from stereo to 3D (conditions A and B), for the emotional conditions, the difference between stereo and 3D conditions was smaller with a slight decrease in

TABLE 1 | Sample description and stimuli assignation.

Stimuli version	Condition A (Neutral-St)	Condition B (Neutral-3D)	Condition C (Arousal-St)	Condition D (Arousal-3D)
N° Participants	60	62	63	68
N° Male	33	27	35	36
N° Female	27	35	28	32
Educational level: 4° EP	29	32	32	35
Educational level: 1° ESO	31	30	31	33
N° participants SSAA	19	21	22	22
N° participants PPBB	41	41	41	46
N° Male SSAA 4° EP	5	7	6	4
N° Male SSAA 1° ESO	5	4	5	6
N° Male PPBB 4° EP	12	8	16	12
N° Male PPBB 1° ESO	11	8	8	14
N° Female SSAA 4° EP	3	4	5	6
N° Female SSAA 1° ESO	6	6	6	6
N° Female PPBB 4° EP	9	13	5	13
N° Female PPBB 1° ESO	9	12	12	7
4° EP SSAA	8	11	11	10
4° EP PPBB	21	21	21	25
1° ESO SSAA	11	10	11	12
1° ESO PPBB	20	20	20	21

EDR from condition C to D (see mean values for conditions in **Table 2**).

Univariate tests for each of the stimuli did not reveal a significant effect for any of the stimuli, but we observed tendencies toward an effect of the sound spatiality condition for the stimulus "skate down" [$F_{(1, 244)} = 2.98$, $p = 0.085$] with an overall larger peak EDR for the 3D version vs. the stereo version; for this stimulus, there was also a tendency in the effect [$F_{(1, 244)} = 3.38$, $p = 0.067$], and 1st ESO students displayed an overall larger peak EDR for this stimulus (see **Figure 6**). There was also a non-significant tendency toward an interaction between sound emotion condition and sound spatiality condition for the stimulus "tic tac" [$F_{(1, 244)} = 2.77$, $p = 0.097$], see **Table 2**.

Effects on Self-Reported Emotional State (SAM)

The mean self-reported valence and arousal ratings for each educational level, test time (pre- and post-experience), and sound condition (A—Neutral stereo, B—Neutral 3D, C—Emotional stereo, and D—Emotional 3D) are displayed in **Figures 7**, **8**. Note that higher ratings of valence and arousal represent more pleasant and arousing experiences, respectively.

First, both for self-reported valence and arousal ratings, the difference from pre-test to post-test were entered into Kruskal-Wallis tests, one for each educational level, to investigate potential variations in the pre-post change between the four sound conditions (A—Neutral stereo, B—Neutral 3D, C—Emotional stereo, and D—Emotional 3D). These analyses showed no significant differences between conditions (all ps > 0.05).

Then, in order to test the potential influence of the factor test time, sound emotion condition, and sound spatiality condition in the self-reported emotional state, according to the educational level, self-reported valence and arousal ratings were used as dependent variables for a Multivariate Analysis of Variance (MANOVA). The within-subject factor was test time (Pre vs. Post) and the between-subject factors were educational level (4th EP, 1st ESO), sound emotion condition (Neutral, Emotional), and sound spatiality condition (stereo, 3D). Wilks' Lambda was used as the multivariate criterion. The results of the multivariate test revealed that there was a significant main effect of educational level [$F_{(2, 231)} = 46.36$, $p < 0.001$,Wilks' Lambda = 0.71], a significant main effect of test time [$F_{(2, 231)} = 14.39$, $p < 0.001$,Wilks' Lambda = 0.89], and a significant interaction between educational level and test time [$F_{(2, 231)} = 3.42$, $p = 0.034$,Wilks' Lambda = 0.71]. As it can be seen in **Figure 8**, the 4th EP gave higher ratings, and there were also higher ratings in post-test than in pre-post.

In order to explore the significant interaction between educational level and test time, separate MANOVAs for each educational level were conducted with within-subject factor test time and with between-subject factors sound emotion and sound spatiality conditions. The results of the multivariate test revealed that there was a significant main effect of test time for the 4th EP [$F_{(2, 115)} = 12.53$, $p < 0.001$,Wilks' Lambda = 0.82] and for the 1st ESO [$F_{(2, 115)} = 4.84$, $p = 0.010$,Wilks' Lambda = 0.92]. Univariate tests revealed a significant pre-post effect on valence for both educational level, the 4th EP [$F_{(1, 116)} = 15.22$, $p < 0.001$], and the 1st ESO [$F_{(1, 116)} = 9.76$, $p = 0.002$] as well as a significant pre-post effect on arousal only for the 4th EP [$F_{(1, 116)} = 23.27$, $p < 0.001$]. The rest of univariate tests were not significant. Overall, participants from both educational levels reported more pleasant emotional state in the post-test than in the pre-test, and participants in the 4th EP group reported being more aroused in the post-test than in the pre-test. For all analyses, there was no significant effect of the sound emotion condition or sound spatiality condition, and neither was there an interaction between the sound condition and test time (all ps > 0.05).

Self-Report of Immersion Level

The hypothesis proposes that the immersion level will be higher in the sound conditions where a 3D sound mix has been used. A specific question was presented to the participants—"Have you felt that you were inside the story?"—using a Likert scale with five answer options (None/A little bit/Some/Quite/A lot). The

TABLE 2 | Peak EDR ± SE values [kΩ] for each educational level (4th EP and 1st ESO) and sound condition (Neutral stereo, Neutral 3D, Emotional stereo, and Emotional 3D) for the different events (stimuli).

	Stimuli version	Condition A	Condition B	Condition C	Condition D
4° EP	Meteor fall	0.912 (0.92)	1.027 (0.88)	4.115 (0.88)	0.78 (0.85)
	Dressing as scientist	3.176 (1.03)	3.394 (0.98)	1.794 (0.98)	4.19 (0.95)
	Tic-Tac	0.826 (0.90)	2.23 (0.86)	3.4 (0.86)	0.793 (0.83)
	Chronometer	1.778 (0.97)	1.628 (0.92)	3.543 (0.92)	1.726 (0.89)
	Sad day for light-speed	1.939 (0.99)	3.008 (0.94)	3.356 (0.94)	3.195 (0.91)
	Down effect	0.84 (0.83)	0.607 (0.79)	2.796 (0.79)	0.731 (0.77)
	Skate-down	1.707 (1.1)	1.859 (1.04)	2.293 (1.04)	2.846 (1.01)
	MEAN 4° EP (SD)	1.638 (0.79)	1.836 (0.99)	2.762 (1.08)	2.001 (1.29)
1° ESO	Meteor fall	1.332 (0.90)	1.54 (0.91)	0.859 (0.89)	0.956 (0.87)
	Dressing as scientist	2.952 (0.99)	3.531 (1.01)	3.046 (0.99)	2.498 (0.96)
	Tic Tac	1.317 (0.87)	2.389 (0.89)	0.878 (0.87)	1.887 (0.84)
	Chronometer	2.602 (0.93)	1.816 (0.95)	2.578 (0.93)	2.348 (0.91)
	Sad day for light-speed	2.107 (0.96)	3.317 (0.97)	3.158 (0.96)	3.196 (0.93)
	Down effect	0.843 (0.80)	1.186 (0.82)	1.752 (0.80)	0.896 (0.78)
	Skate-down	2.231 (1.06)	4.196 (1.08)	2.639 (1.06)	5.121 (1.03)
	MEAN 1°ESO (SD)	1.801 (0.78)	2.378 (1.18)	2.025 (0.95)	2.224 (1.45)
	MEAN GLOBAL (SD)	**1.72 (0.76)**	**2.11 (1.09)**	**2.39 (1.06)**	**2.11 (1.33)**

Mean values (SD) for 4th EP and 1st ESO are marked in bold font.

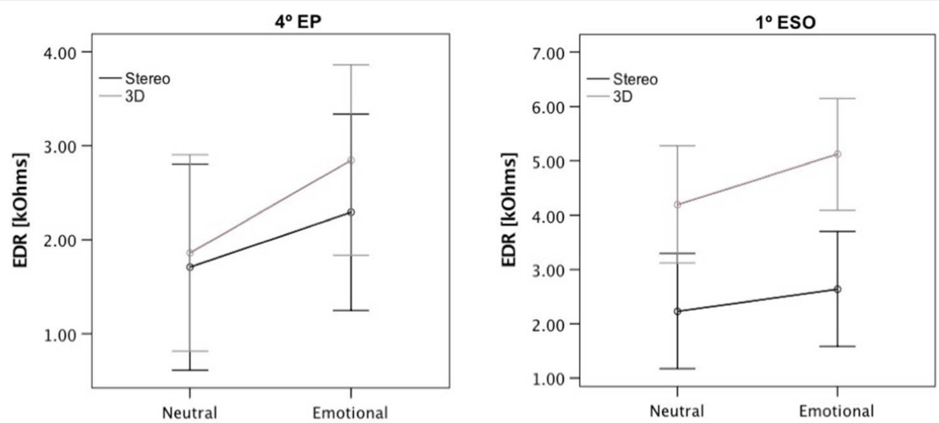


FIGURE 6 | Peak EDR ± SE values [kΩ] for stimuli Skate-Down for each educational level (4th EP and 1st ESO) and sound condition (Neutral stereo, Neutral 3D, Emotional stereo, and Emotional 3D).

mean immersion ratings for each course and sound condition (A—Neutral stereo, B—Neutral 3D, C—Emotional stereo, and D—Emotional 3D) are displayed in **Figure 9**.

First, the immersion ratings were entered into a Kruskal-Wallis test, one for each educational level, to investigate potential modifications between the four sound conditions. This analysis showed no significant differences between conditions ($p > 0.05$). Then, in order to test the potential interaction between the factors sound emotion and sound spatiality, the immersion ratings were used as dependent variables for two ANOVAs, one for each educational level, with between-subject factors sound emotion condition (Neutral and Emotional) and sound spatiality condition (stereo, 3D). These analyses did not yield significant results ($p > 0.05$).

A positive correlation between SAM test and immersion level has been found, using Spearman' rho statistics: higher valence and arousal correspond to higher immersion level (see **Table 3**).

Effects on Perceived Emotions

The hypotheses propose that the different sound conditions will generate distinctive intensity levels of perceived emotions in the listener. Two specific questions were asked to the participants after listening to the story:

- What moment in the story did you find most exciting? Why? (Q1)
- How did you feel at that moment? (Q2)

The answers to both questions were registered using the Think-aloud technique. The qualitative analysis of the recordings

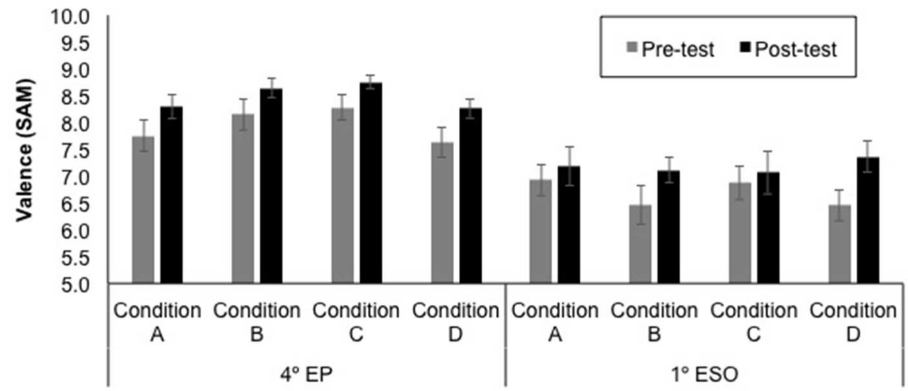


FIGURE 7 | Mean valence ratings ± SE (on a nine-point scale) for each educational level (4th EP and 1st ESO), test time (pre- and post-experience) and sound condition (A—Neutral stereo, B—Neutral 3D, C—Emotional stereo, and D—Emotional 3D).

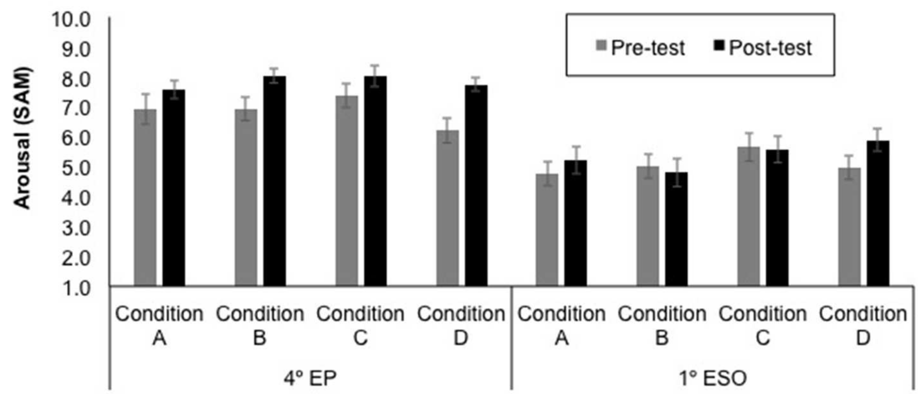


FIGURE 8 | Mean arousal ratings ± SE (on a nine-point scale) for each educational level (4th EP and 1st ESO), test time (pre- and post-experience) and sound condition (A—Neutral stereo, B—Neutral 3D, C—Emotional stereo, and D—Emotional 3D).

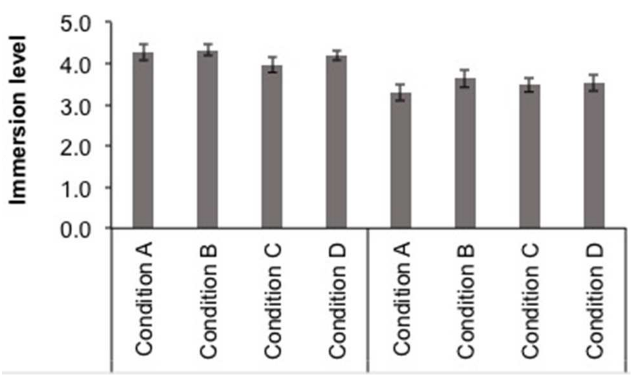


FIGURE 9 | Mean immersion ratings ± SE (on a five-point scale) for each course (4th EP, on the left, and 1st ESO, on the right) and sound condition (A-Neutral stereo, B-Neutral 3D, C-Emotional stereo, and D-Emotional 3D).

TABLE 3 | Correlations (Spearman' rho values) between Immersion, Valence, and Arousal for the two pre-test and post-test measures.

Variables	1	2	3	4	5	6	7
Valence—Pretest 1	1						
Arousal—Pretest 1	0.522**	1					
Valence—Pretest 2	0.529**	0.444**	1				
Arousal—Pretest 2	0.389**	0.683**	0.551**	1			
Valence—Post	0.520**	0.402**	0.619**	0.401**	1		
Arousal—Post	0.411**	0.578**	0.453**	0.620**	0.648**	1	
Immersion	0.307**	0.227**	0.313**	0.288**	0.449**	0.413**	1

***Correlation is significant at the 0.01 level (2-tailed).*

showed that the different sound conditions generated distinctive reported emotional responses in the participants. Several categories were established: moment of history that seemed most exciting; the type of emotion that this moment generated in participants; and the intensity of that emotional response.

In relation to the moment of history that they found most exciting, four subcategories have been identified for this dimension of analysis: history in general; action situation; communication–help situation; and communication situation.

The one that stood out with a high percentage of references was the one related to an action situation, followed by the communication–help situations, as shown in **Figure 10**.

The moments of action narrated mainly focus on the episode in which Elena throws herself on the ramp with the skateboard, showing in most cases the emotion through expressions of intensity, onomatopoeia, or narrating unexpected and catastrophic outcomes.

> "*Elena throws herself with the skateboard from the slope of the school and began to do with her mouth buaaah.*" (Condition A\\263_PPBB-4EP-B-05)

> "*Well, I felt that she was going to throw herself for sure—that she was going to crash, and that in the end she had time to play a fragment of the flute—because...* (it is not understood) *if she enters by the flute and if she plays the flute is impossible to play, it can play*." (Condition A\\302_PPBB-4EP-D-05)

> "*Elena fell with the skateboard because it was a very steep slope and she had to be very careful and I thought she was going to be in a coma and she was going to breathe running and that...*" (Condition A\\625_SSAA-1ESO-A-03)

> "*When she threw himself down the steep and so big slope and when she put on the cow's face, because that part was so funny, I loved it and it was very cool, I want to do it, I love it, I love it. Well the truth is that I did not like it, well I liked everything..." (*Condition C \\ 332_PPBB-4EP-A-14)

Focusing on those situations of action of the story that has elicited more intense emotion in the participants, no significant differences have been found between the two educational levels (4th EP or 1st ESO). As shown in **Figure 11**, in both educational levels, the conditions that collect more fragments in this subcategory are those that correspond to an arousal sound treatment: condition C and condition D. In the 1st ESO participants, condition C was the one that generated a greater amount of emotions in action situations, while in 4th EP participants it was the condition D condition. In those passages in history that relate moments of action, e.g., when Elena throws herself on the ramp with the skateboard or when Salva disguises himself, the arousal treatment of the sound contributes to enrich the description and enhance the action.

In relation to the type of emotions that participants felt while listening to the story, four subcategories of analysis have been identified and defined:

- Positive: expressions that produce participant well-being.
- Negative: expressions that generate participant discomfort.
- Neutral: expressions that do not show reactions either pleasant or unpleasant.
- Contradictory: those expressions where the manifestation of the same feeling by the participant reflects conflicting emotions (positive or negative).

Participants state a high percentage of expressions of emotions, feelings, or positive moods throughout their responses, as shown in **Figure 12**. Negative and contradictory emotions also have a significant presence in this category of analysis.

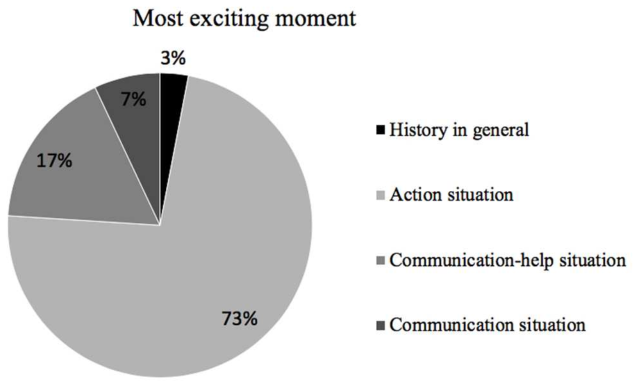


FIGURE 10 | Distribution of the most exciting moment between the identified subcategories.

Depending on the type of condition to which the participants have been exposed while listening to the story, the "Positive" emotions exhibited a higher presence in all four types of conditions, as **Figure 13** shows. The results in this category do not show a significant difference between participant groups either. The condition A stands out, especially with 74%, with a minimum presence of the rest emotions. However, the condition D is the condition that generates in the participants a greater variety of emotions, highlighting the positive, negative, and contradictory types of emotions. According to this result, it is concluded that the condition that combines the two sound treatments (arousal sound design + 3D sound mix) is the one that generates a greater diversity of emotional responses (positive, negative, contradictory, or neutral) in the listener.

The emotional responses felt by the participants, apart from being classified according to a type of emotion, can have greater or lesser intensity (**Supplementary Figure 1**). Most of the emotions raised by the listening of the story in the participants are classified as "High intensity." No significant differences have been found between the four conditions, just minimal variations between them.

This "High intensity of emotions" is expressed by participants through adverbs of quantity, profusion in the manifestation of various emotions, or biological responses to stimuli (e.g., laughter):

> *"I felt super cool."* (Condition A\\263_PPBB-4EP-B-05)

> *"I have felt happy, imaginative. I have felt excited, passionate, deep within the story ehh thoughtful."* (Condition A\\277_PPBB-4EP-B-03)

> *"I had a great time." "It was very cool." "It made me very funny." "It made me laugh a lot." "Very funny, and I laughed a lot because, because I liked it a lot."* (Condition A\\297_PPBB- 4EP-D-02)

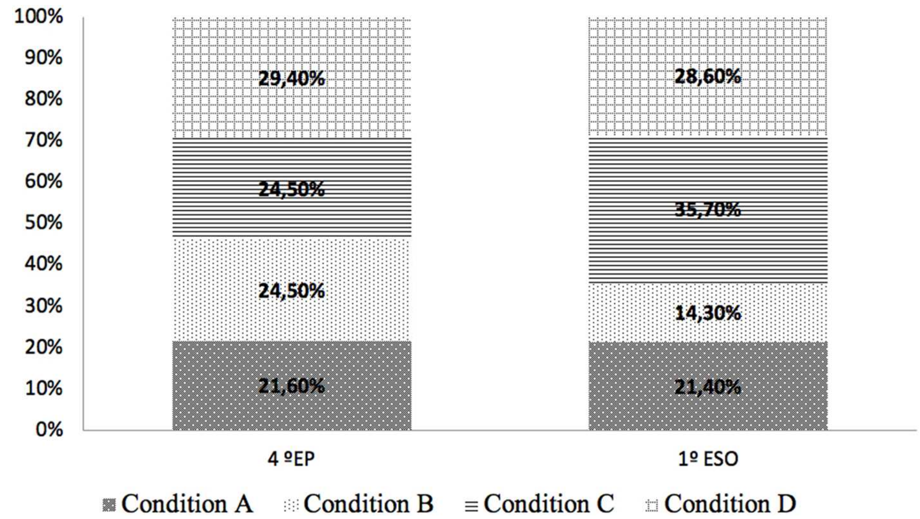


FIGURE 11 | Perceived emotion in action situations comparison between 4th EP and 1st ESO participants.

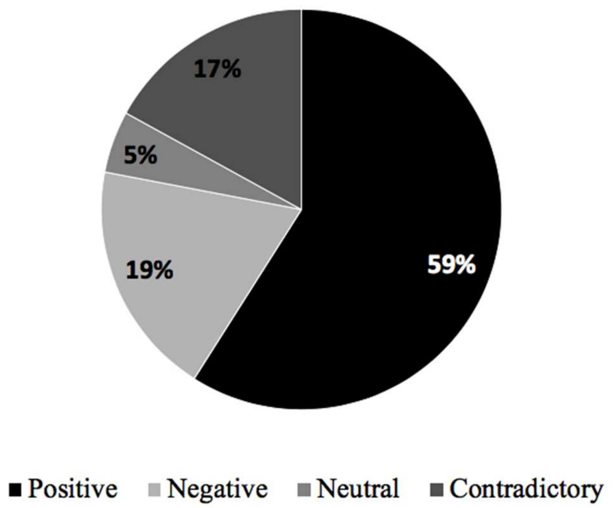


FIGURE 12 | Type of emotions felt by participants.

"I felt very laugh. I felt super, super lively as if it were a humorous story, super cool story, very exciting story." (Condition B\\269_PPBB-4EP-B-06)

"I laughed a lot with them because I liked it a lot, and I loved this story because it is so much fun." (Condition B\\339_PPBB-4EP-A-12)

"I felt happy, funny, entertaining, imaginative." (Condition C\\320_PPBB-4EP-C-13)

"I felt very happy." (Condition D\\633_SSAA-1ESO-A-19)

"I have felt super excited, and I really liked the story." (Condition D\\672_SSAA-4EP-A-02)

Mental Images Elicited by the Story

The proposed hypothesis (hypothesis H4) that both quantity and richness/detail of images will be higher in the CD sound condition than in the other conditions. According to this hypothesis, two dimensions were measured. On one side, the number of mental images elicited by the story and, on the other hand, the richness and detail of such images.

The collected data from the think-aloud technique, as answers to two of the proposed questions to participants, were used to measure both dimensions:

- Try not to think too much, and speak as things come to your mind: what have you imagined when Elena was going down the school hill on the scooter? (Q3)
- Describe how you have imagined Salva (Q5)

The qualitative analysis of the information collected from Think-aloud was carried out according to two main categories: description of the represented situations and the number of elicited images when they narrate an episode of the story.

In the analysis carried out on the description made by the participants of the situations that tell about the story three subcategories have been inferred depending on the content they narrated: as it is told in the story; invented but same plot and characters; or invented with a changed argument but with elements present in other passages. According to the condition of the stimulus to which the participants have been exposed, condition D is the one that has generated the recreation of mental images more faithful to what is told in the story (**Supplementary Figure 2**).

This finding confirms the hypothesis that an "emotionally marked" sound design in the context of a sound story will

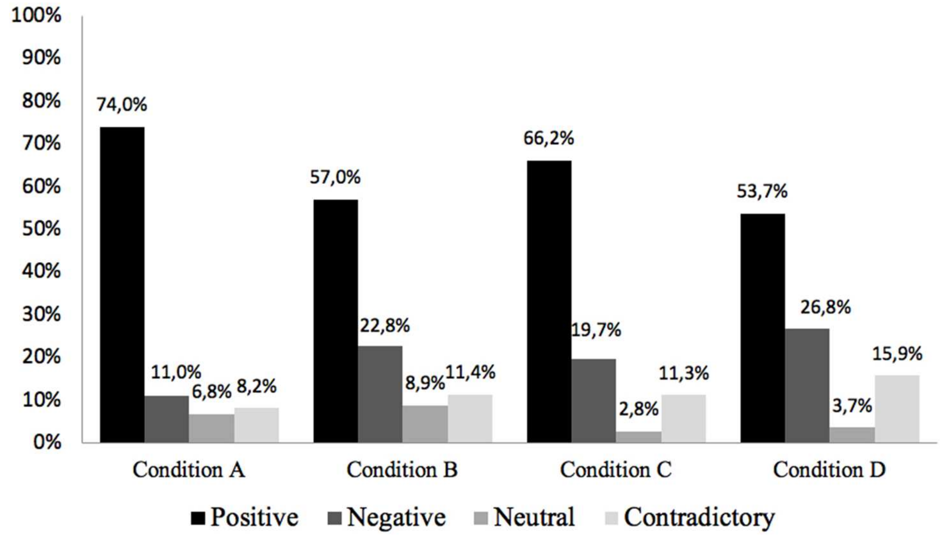


FIGURE 13 | Type of emotion per sound condition (%).

generate richer and more detailed mental images in the listener than a soundtrack without this emotional purpose.

Per educational level, 4th EP participants exposed to Conditions with arousal treatment (C and D) have elaborated richer and more detailed mental images compared to the ones who listened the neutral conditions. On the other hand, in the case of 1st ESO participants, it was the Condition B the one that generated more detailed and richer mental images.

In general, the mental images that the participants represent are present in the story—specifically in two key moments of the story, the scene of Elena going down the street over a skateboard and the scene in which Salva dresses himself as a scientist—as are passages invented by the participants themselves.

> *"The part where she was thrown by, was pulling crazy on the ramp has been very cool, but the image of Salva's dressed as a scientist has come to mind. It has also come to mind and that I found it very funny because he looked like a madman, a mad scientist."* (Condition A\\274_PPBB-4EP-B-04).

> *"Elena—well, Salva's sister whose name was Elena—doesn't know how to play the flute, and Salva the brother went to comfort her, invents words to make her laugh, told her to play a piece of . . . (can't be heard) with the flute."* (Condition B\\654_SSAA-1ESO-B).

> "*Well, at first there were the two brothers in the town. They went uphill, and I was imagining, I was imagining that it was a town, and I was imagining the car races, and also the stops in a pile of a hill, and I also imagined the moment when the sister was sad, and then it seemed funny when the brother took the diapers, the little children diapers, and now takes a costume, well the costume from the grandmother, and then he puts it on, and then he does experiments, does experiments until he success, come on, that the brother is very crooked. The truth is that he is very cool. I would love to hear it again.*" (Condition D\\323_PPBB-4EP-C-21).

Finally, the number of mental images that the story provokes in the participant has been measured depending on the sound condition. As **Supplementary Figure 3** shows, Condition B and Condition C are the conditions that generate the greatest number of mental images in the participants when they tell and recreate the story. As results show, in the 4th EP educational level, the arousal condition shows a higher amount of mental images but only in the stereo condition. Otherwise, in the 1st ESO educational level, the 3D sound condition shows higher number of mental images than stereo condition, though only in the neutral condition.

Finally, to complement the qualitative analysis, two further closed questions were included to quantitatively measure the kind and number of mental images:

- "The images that have come to your mind while you were listening were:"
  - Real—Fictitious (four-point Likert scale)
  - Clear—Unclear (four-point Likert scale)
  - Exciting—Boring (four-point Likert scale)
  - Known—Unknown (Q6) (four-point Likert scale)
- "After listening to the story ... images came to my mind." Multiple option question:
  - A lot of
  - Some
  - A few (Q7)

Responses to each of these questions were used as dependent variables for an ANOVA with between-subject factors educational level (4th EP and 1st ESO), sound emotion condition (Neutral and Emotional), and sound spatiality condition (stereo and 3D). Results showed only a main effect for the questions related to clear/unclear images [$F_{(1,231)} = 5.08$, $p = 0.025$] and exciting/boring images [$F_{(1,231)} = 33.28$, $p <$

0.001]. Non-parametric Kruskal-Wallis were run to confirm the significant effect in relation to clear/unclear images [$H_{(1)} = 5.72$, $p = 0.017$] and exciting/boring images [$H_{(1)} = 31.97$, $p < 0.001$]. In a scale ranging from 1 (clear) to 4 (unclear), 4th EP students found the images more unclear (M = 3.32, SE = 0.08) than 1st ESO students (M = 3.05, SE = 0.08). Furthermore, in a scale ranging from 1 (exciting) to 4 (boring), 4th EP students also found the images more boring (M = 3.75, SE = 0.06) than 1st ESO students (M = 3.25, SE = 0.06).

DISCUSSION

As exposed in the results section [Effects on Physiological Arousal (EDR)], the means of the physiological analyses do not show significant differences between conditions.

When considering the moments that have a specific sound treatment, a non-significant tendency indicated an interaction between sound emotion condition and sound spatiality condition. Overall, there was higher EDR for the "emotional" than for the "neutral" conditions, and only for the "neutral" conditions was the expected increase in EDR from the "stereo" to "3D" condition observed. When looking separately at the two educational levels, 4th EP and 1st ESO, we found differences between them. In all these moments, a constant pattern is found in the 4th EP participants: the arousal condition shows a higher impact of the EDR, confirming the initial hypothesis (H1), but only in the stereo condition. In 3D conditions, the variance in the impact of the arousal condition over the neutral is not so remarkable. Otherwise, in the 1st ESO participants, the 3D sound condition obtains higher EDR levels than stereo condition, but the evolution between neutral and arousal is not so consistent.

On the other hand, upon analyzing hypothesis H3—"a soundtrack mixed in 3D sound format will elicit more intense emotional response in the listener compared to a soundtrack mixed in stereo"—it was found that, as results show, 1st ESO participants are more affected by the 3D sound than by the arousal treatment of sound even in the moments in which there is no special 3D sound treatment. In the moments where the three-dimensionality of the sound is more focused, the higher impact of the 3D sound condition is more evident in both educational levels, which reinforces previous findings (Murphy and Pitt, 2001; Steele and Chon, 2007) and reinforces the initial hypothesis (H3). However, in 4th EP participants, the EDR level of the 3D condition falls when combined with arousal treatment. We hypothesized that this disparity between educational levels may be explained by the fact that it is a "sound only" stimulus, and a greater level of cognitive maturity may be necessary to decodify both processes (arousing and spatialization), which indicates that the impact in younger children is greater in the arousal condition where the sound treatment becomes more evident than the 3D sound mix. This hypothesis needs further investigation in future research, as there are no previous studies on sound emotional impact on different ages that can strength or refute this argument. Furthermore, while older people may be more habituated to the arousal treatment of sound, as it is a very common process used in film sound production, 3D sound is quite a novel narrative technique that may have a greater impact, which may clarify and justify the greater level of emotional impact of 3D condition over stereo and over neutral or arousal treatment.

Results from the Think-aloud analysis (subsection Effects on Perceived Emotions) confirm most of the findings obtained from EDR and SAM measurements. On the one hand, those moments in the story in which both arousal sound treatment and 3D sound mix have been applied are the ones that students stand out as the most exciting ones: action situations (Elena on the skate down the street) and communication–help situations (Salva dressing as a scientist). Although no significant differences have been found between conditions, in the case of action situations, the arousal predominant condition (condition B) has had a greater emotional impact on the 4th EP participants, while in the 1st ESO students, the 3D sound condition is the one that obtains the greatest emotional impact when identifying action situations throughout the story. These results are consistent with the EDR response patterns in both educational levels: younger participants are more affected by the arousal treatment, while 3D mix has a greater impact over older participants.

On the other hand, regarding the valence and intensity of the reported emotions, a high percentage of positive emotions, feelings, and moods have been identified in the responses without finding significant differences between the participants of both educational levels or between the four conditions. In accordance with these results, there is consistency between the data obtained in the quantitative and qualitative analysis, confirming hypotheses H1 and H3.

Another of the elements proposed for analysis was related to H2—"the use of emotional marked sound in the design and production of a sound story will generate richer and more detailed mental images in the listener than a soundtrack without this emotional intention"—and H4—"a soundtrack mixed in 3D sound format will generate richer and more detailed mental images in the listener compared to a soundtrack mixed in stereo."

In relation to the mental images reported through the Think-aloud results, in those cases in which the students have been exposed to 3D sound mix conditions, the description they made regarding the scene of Elena's descent down the street on the skateboard was that it been more real and detailed in accordance with the story. This finding confirms the H4 hypothesis. It is also noted that sound condition B has generated a greater impact on the number of reported mental images in 4th EP students. However, for the 1st ESO students, it is the treatment of 3D sound that caused a greater number of mental images in students. The pattern identified in the EDR response between the different groups is also maintained in the analysis of the mental images.

Finally, according to H5—"the emotional impact elicited by both emotional marked sounds and/or 3D sound mix format takes place at an unconscious level and is not self-perceived by the listener"—and regarding self-reported emotions, we found that, overall, participants from both educational levels reported more pleasant emotional state in the post-test than in the pre-test, showing that they liked the experience. Furthermore, participants from 4th EP reported being more aroused in the post-test than in the pre-test, showing that they also found the story exciting.

The fact that the 1st ESO students did not show this change may be explained by the story being more suitable to the younger students. There were no differences due to the sound treatment in self-reported emotions or immersion level, that is, in the answer of the participants related to their own perception of the emotion level and the immersion level. Otherwise, there were differences in the other measures of the emotion and immersion levels (not self-perception but physiological, amount, richness, and detail of mental images). These results confirm hypothesis H5: the emotional impact elicited by both emotional marked sounds and/or 3D sound mix format takes place at an unconscious level and will therefore not be reported by the listener.

CONCLUSIONS

As a main conclusion, the hypotheses have been partially confirmed. Although no significant correlations have been found between the conditions and the variables considered, different patterns depending of educational levels (and, subsequently, age of participants) have been identified. This opens the topic to further studies in which age (and diverse aspects related to this parameter, much like cognitive development or consumer habits) must be considered in the definition of variables.

Apart from the results obtained analyzing the different variables independently, a relevant finding has emerged from the combination of the two variables (arousal sound treatment and 3D audio mix). The interaction between these variables in the four conditions generates a different response in the participants, particularly in EDR measurement, than the response obtained when only one of the variables is considered.

Some limitations must be considered from the present study:

- First of all, the subtlety and diversity of the sound treatment: with the purpose of giving a step forward in the research field, the differences between conditions have been based on specific changes in the sound that are quite subtle and not consciously perceived by the listener rather than being based on the intervention on more noticeable changes, such as the presence or absence of certain elements. Otherwise, this approach has also made possible to carry out a field study with listening conditions as close to real situations as possible.
- In relation to the aforementioned, the combination of variables in the same stimulus (emotional-marked or neutral with stereo or 3D sound) may have limited the clarity of the results because of the interaction of these variables, as has been detailed in relation to the different cognitive development of the diverse educational level participants, but also considering other contextual factors that refer to the social and cultural memory, such as previous experiences or expectations, which is consistent with previous findings (Tajadura-Jiménez, 2008; Grimshaw, 2014).

Finally, further developments and applications of the present study are proposed:

- Replication of the present study with wider age range and extending the sample to adults.
- Studies derived from the present one, but focused on specific sound treatment, in order to enrich and complete the knowledge base about the emotional impact of different sound with real-world stimulus and listening situations.
- Application of the results to the production of sound stories, but also to video games, films, or advertising. As Dafonte-Gomez (2014, p. 206) concludes in another study on viral advertising, "the obtained results show the outstanding presence of surprise and joy as dominant emotions in the most successful viral videos." According to the results of the present study, this kind of positive response can be achieved through the use of arousal sound treatment.
- Application in the educational environment: as Mora (2005) states, an emotionally marked experience is best remembered, especially because of the connection between the hippocampus and the amygdala, where our emotions are represented. A sound-based educational resource focused in the arousal treatment of the sound may improve the learning experience. With this purpose, a first development from the present study has been carried out: as part of the "Unconscious listening" project "Gale's journey," an educational project based on the use of arousal sound treatment and TUI object interface, has been designed to foster the teaching of different contents from the Primary Education curriculum in Spain: Natural Sciences, Social Sciences, EFL, and Music. A first exploratory study has also been carried out, and its results are reporting a high level of usability (easy to use, clear, and appealing) as well as positive student feedback in terms of motivation, attention level, and learning improvement.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comité de Ética de la Universidad Loyola Andalucía. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

FC and AT-J described the theoretical framework and were in charge of the literature search. FC wrote the Methods section. IL-C and AT-J were in charge of the quantitative analysis and results, while TM-B was in charge of the qualitative analysis and results. All authors participated in the discussion and elaborated the Hypothesis section.

FUNDING

This project has been funded by the Grants for Researchers and Cultural Creators-—2016 Call, from BBVA Foundation

through the project Unconscious listening (http://www.laescuchainconsciente.com).

ACKNOWLEDGMENTS

FC acknowledges the participation of the students and teachers from the schools Ntra. Señora del Águila and San José SSCC. AT-J was supported by grant ryc-2014-15421 from Ministerio de Economía, Industria y Competitividad of Spain.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00737/full#supplementary-material

Supplementary Figure 1 | Intensity of perceived emotions.

Supplementary Figure 2 | Subcategories of mental images elicited by the four sound conditions.

Supplementary Figure 3 | Number of mental images elicited by the four sound conditions.

Supplementary Audio 1 | Meteor_fall_neutral.mp3.

Supplementary Audio 2 | Meteor_fall_arousal.mp3.

Supplementary Audio 3 | Chronometer_neutral.mp3.

Supplementary Audio 4 | Chronometer_arousal.mp3.

Supplementary Audio 5 | Scientist_st.mp3.

Supplementary Audio 6 | Scientist_3D.mp3.

Supplementary Audio 7 | Skate-fall_st.mp3.

Supplementary Audio 8 | Skatefall_3D.mp3.

REFERENCES

Aiger, M., Palacín, M., Cornejo, J. M. (2013). La señal electrodérmica mediante Sociograph: metodología para medir la actividad grupal. *Rev. Int. Psicol. Soc.* 28, 333–347. doi: 10.1174/021347413807719102

Asutay, E., Västfjäll, D., Tajadura-Jiménez, A., Genell, A., Bergman, P., and Kleiner, M. (2012). Emoacoustics: a study of the psychoacoustical and psychological dimensions of emotional sound design. *AES J. Audio Eng. Soc.* 60.

Berger, C. C., Gonzalez-Franco, M., Tajadura-Jiménez, A., Florencio, D., and Zhang, Z. (2018). Generic HRTFs may be good enough in virtual reality. Improving source localization through cross-modal plasticity. *Front. Neurosci.* 12:21. doi: 10.3389/fnins.2018.00021

Bjork, E. A. (1999). Startle, annoyance and psychophysiological responses to repeated sound bursts. *Acustica* 85, 575–578.

Blauert, J., and Jekosch, U. (1997). Sound-quality evaluation: a multi-layered problem. *Acustica* 83, 747–753.

Boucsein, W. (Ed.). (2012). *Electrodermal Activity*, 2nd Edn. New York: NY: Springer.

Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9

Bradley, M. M., and Lang, P. J. (1999). *International Affective Digitized Sounds (IADS): Stimuli, Instruction Manual and Affective Ratings* (Tech. Rep. No. B-2). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Bradley, M. M., and Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology* 37, 204–215. doi: 10.1111/1469-8986.3720204

Cho, J., Yi, E., and Cho, G. (2001). Physiological responses evoked by fabric sounds and related mechanical and acoustical properties. *Textile Res. J.* 71, 1068–1073. doi: 10.1177/004051750107101206

Cox, T. (2008). Scraping sounds and disgusting noises. *Appl. Acoustics* 69, 1195–1204. doi: 10.1016/j.apacoust.2007.11.004

Dafonte-Gomez, A. (2014). Claves de la publicidad viral: de la motivación a la emoción en los vídeos más compartidos. *Comunicar* 43, 199–207. doi: 10.3916/C43-2014-20

De Gelder, B., and Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cogn. Emot.* 14, 289–311. doi: 10.1080/026999300378824

Edelberg, R. (1967). "Electrical properties of the skin," in Methods in Psychophysiology, ed C. C. Brown (Baltimore, MD: Williams and Wilkins), 1–53.

Flick, U. (2012). *Introducción a la Investigación Cualitativa*. Madrid: Ediciones Morata.

Graziano, M. S. A. (2001). A System of multimodal areas in the primate brain. *Neuron* 29, 4–6. doi: 10.1016/S0896-6273(01)00174-X

Grimshaw, M. (2007). "The resonating spaces of first-person shooter games," in *Proceedings of the 5th International Conference on Game Design and Technology* (Liverpool).

Grimshaw, M. (2008). Sound and immersion in the first-person shooter. *Int. J. Intell. Games Technol*. 119–124. Available online at: http://digitalcommons.bolton.ac.uk/gcct_journalspr/3

Grimshaw, M. (2014). Sounding off on sound. *Comp. Games J.* 3, 194–195. doi: 10.1007/BF03395954

Gygi, B. (2001). *Factors in the identification of environmental sounds* (PhD thesis). Indiana University, Bloomington, IN, USA.

Hong, J., He, J., Lam, B., Gupta, R., and Gan, W.-S. (2017). Spatial audio for soundscape design: recording and reproduction. *Appl. Sci.* 7:627. doi: 10.3390/app7060627

Jäncke, L., Vogt, J., Musial, F., Lutz, K., and Kalveram, K. T. (1996). Facial EMG responses to auditory stimuli. *Int. J. Psychophysiol.* 22, 85–96. doi: 10.1016/0167-8760(96)00013-X

Jekosch, U. (1999). Meaning in the context of sound quality assessment. *Acustica* 85, 681–684.

Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 559–575. doi: 10.1017/S0140525X08005293

LaGrou, J. (2014). The future of audio engineering. *Tape Op*. 100, 21–28.

Landstrom, U., Akerlund, E., Kjellberg, A., and Tesarz, M. (1995). Exposure levels, tonal components, and noise annoyance in working environments. *Environ. Int.* 21, 265–275. doi: 10.1016/0160-4120(95)00017-F

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychol. Rev.* 97, 377–395. doi: 10.1037/0033-295X.97.3.377

LeDoux, J. E. (1998). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York, NY: Simon and Schuster.

Levenson, R. W. (1994). "Human emotions: a functional view," in *The Nature of Emotion: Fundamental Questions*, eds P. Ekman and R. J. Davidson (New York, NY: Oxford University Press), 123–126.

Martin, I., and Venables, P. H. (Eds.). (1980). *Techniques in Psychophysiology*. New York, NY: Wiley.

Martínez Herrador, J. L., Garrido Martín, E., Valdunquillo Carlón, M. I. Macaya Sánchez, J. et al. (2008). Análisis de la Atención y la Emoción en el Discurso Político a Partir de un Nuevo Sistema de Registro Psicofisiológico y su Aplicación a Las Ciencias Políticas. DPSA, Documentos de trabajo del Departamento de Psicología Social y Antropología.

Martínez Herrador, J. L., Monge Benito, S., Valdunquillo Carlón, M. I. (2012). Medición de las respuestas psicofisiológicas grupales paraapoyar el análisis de discursos políticos. Trípodos. Núm. 29, 53–72.

McDonald, J. H. (2014). *Handbook of Biological Statistics,* 3rd ed. Baltimore, MA: Sparky House Publishing.

McDonald, J. J., Teder-Sa?leja?rvi, W. A., and Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature* 407, 906–908. doi: 10.1038/35038085

Mora, F. (2005). *Cmo Funciona el Cerebro*. Madrid: Alianza.

Murch, W. (2001). *In the Blink of an Eye*. Los Angeles, CA: Silman-James Press.

Murphy, D., and Pitt, I. (2001). Spatial sound enhancing virtual story telling. *Lecture Notes Comput. Sci*. 2197, 20–29. doi: 10.1007/3-540-45420-9_3

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., and Dickert, S. (2006). Numeracy and decision making. *Psychol. Sci*. 17, 407–413. doi: 10.1111/j.1467-9280.2006.01720.x

Ritterfeld, E., Klimmt, C., Vorderer, P., and Steinhilper, L. K. (2005). The effects of a narrative audiotape on preschoolers' entertainment experience and attention. *Media Psychol*. 7, 47–72. doi: 10.1207/S1532785XMEP0701_3

Rodero, E. (2012). See it on a radio story: sound effects and shots to evoked imagery and attention on audio fiction. *Commun. Res*. 39, 458–479. doi: 10.1177/0093650210386947

Schirmer, A., Escoffier, N., Cheng, X., Feng, Y., and Penney, T. B. (2016). Detecting temporal change in dynamic sounds: on the role of stimulus duration, speed, and emotion. *Front. Psychol*. 6:2055. doi: 10.3389/fpsyg.2015.02055

Scorgie, M. A., and Sanders, R. D. (2002). *The Effect of Sound Delivery Methods on a User's Sense of Presence in a Virtual Environment*. Naval Postgraduate School. Available online at: http://calhoun.nps.edu/bitstream/handle/10945/6025/02Mar_Sanders.pdf?sequence~=1

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci*. 17, 565–573. doi: 10.1016/j.tics.2013.09.007

Shilling, R., Zyda, M. Y., and Wardynsky, E. (2002). "Introducing emotion into military simulation and videogame design: America's army: operations and VIRTE," in *Communication at GameOn Conference* (London).

Sokolov, E. N. (Ed.). (1963). *Perception and the Conditioned Reflex*. Oxford: Pergamon Press.

Steele, D. L., and Chon, S. H. (2007). A perceptual study of sound annoyance. *Audio Mostly* 2007, 9–24. Available online at: https://ccrma.stanford.edu/~shchon/pubs/AudioMostly2007.pdf

Tajadura-Jiménez, A. (2008). *Embodied psychoacoustics: spatial and multisensory determinants of auditory-induced emotion* (Ph.D. thesis). Chalmers University of Technology, Gothenburg, Sweden.

Tajadura-Jiménez, A., Larsson, P., Väljamäe, A., Västfjäll, D., and Kleiner, M. (2010). When room size matters: acoustic influences on emotional responses to sounds. *Emotion* 10:416. doi: 10.1037/a0018423

Tapia, A., and Martín, E. (2015). Neurociencia aplicada al cine: medición de la atención y la emoción en la audiencia de "Memoria de un cine de provincias". *Egitania Sci*. 17, 41–54. Available online at: http://www.egitaniasciencia.ipg.pt/files/egitania_sciencia_17.pdf

Tapia, A., and Martín, E. (2016). Neurociencia aplicada a la elevision: medición de la atención y la emoción de la serie "FOREVER". *Vivat Acad*. 134, 69–82. doi: 10.15178/va.2016.134.69-82

Todd, N. (2001). Evidence for a behavioral significance of saccular acoustic sensitivity in humans. *J. Acoust. Soc. Am*. 110, 380–390. doi: 10.1121/1.1373662

Valkenburg, P. M., and Beentjes, J. W. J. (1997). Children's creative imagination in response to radio and TV stories. *J. Commun*. 47, 21–38.

Västfjäll, D., Larsson, P., and Kleiner, M. (2002). Emotion and auditory virtual environments: affect-based judgments of music reproduced with virtual reverberation times. *Cyberpsychol. Behav*. 5, 19–32. doi: 10.1089/109493102753685854

Venables, P. H., and Christie, M. J. (1973). "Mechanisms, instrumentation, recording techniques, and quantification of responses," in *Electrodermal Activity in Psychological Research*, eds W. F. Prokasy and D. C. Raskin (New York, NY: Academic Press), 1–124.

Zunzunegui, S. (1995). *Pensar la Imagen*. Madrid: Ediciones Cátedra.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX 1. APPARATUS

The physiological measurement of electrodermal activity was made using the Sociograph measuring instrument, patent n° 9902767. It has been used in previous studies, such as those of Martínez Herrador et al. (2008, 2012), Aiger et al. (2013), Tapia and Martín (2015, 2016).

For sound story playblack and data collection (answer to questionnaires and think-aloud recording), an individual 8" Android tablet was used for each participant. It had the following technical specifications

- Manufacturer and model: BQ, Aquaris B8.
- Screen: 8," Resolution: 800 × 1200, 189 ppi, Aspect ratio: 16:10
- Processor:
  - CPU: MediaTek Quad Core MT8163B, 1,3 GHz
  - GPU: MediaTek Mali-T720 MP2, 520 MHz
- RAM: 2 GB.

The headphones used were overhead models with microphone, and they had the following specifications

- Manufacturer and model: Mars Gaming, MH2
- Speaker diameter: 40 mm
- Speaker Impedance: 32 Ohm
- Frequency response: 20 Hz–20 KHz
- Maximum input power: 20 mW
- Sensitivity (SPL): 105 dB +/– 3 dB
- Microphone Dimensions: Dia 6.0 × 5.0 mm
- Sensitivity: –54 db +/– 3 dB
- Directivity: omnidirectional
- Impedance: <2.2 KOhm

frontiers in Psychology

ORIGINAL RESEARCH
published: 21 July 2020
doi: 10.3389/fpsyg.2020.01600

Influence of the Acoustic Environment in Hospital Wards on Patient Physiological and Psychological Indices

Tianfu Zhou[1], Yue Wu[2], Qi Meng[2]* and Jian Kang[3]*

[1] *Department of Architecture, Shanghai Academy of Fine Arts, Shanghai University, Shanghai, China,* [2] *Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, School of Architecture, Harbin Institute of Technology, Harbin, China,* [3] *UCL Institute for Environmental Design and Engineering, The Bartlett, University College London, London, United Kingdom*

OPEN ACCESS

Edited by:
PerMagnus Lindborg, City University of Hong Kong, Hong Kong

Reviewed by:
Jooyoung Hong, Singapore University of Technology and Design, Singapore
Felipe Reinoso Carvalho, KU Leuven, Belgium

***Correspondence:**
Yue Wu
wuyuehit@hit.edu.cn
Qi Meng
mengq@hit.edu.cn

Specialty section:
This article was submitted to Environmental Psychology, a section of the journal Frontiers in Psychology

Received: *06 April 2020*
Accepted: *15 June 2020*
Published: *21 July 2020*

Citation:
Zhou T, Wu Y, Meng Q and Kang J (2020) Influence of the Acoustic Environment in Hospital Wards on Patient Physiological and Psychological Indices. Front. Psychol. 11:1600. doi: 10.3389/fpsyg.2020.01600

Patients in general wards are often exposed to excessive levels of noise and activity, and high levels of noise have been associated with depression and anxiety. Previous studies have found that an appropriate acoustic environment is beneficial to the patient's therapeutic and treatment process; however, the soundscape is rarely intentionally designed or operated to improve patient recovery, especially for psychological rehabilitation. To gain the most accurate, and least variable, estimate of acoustic environmental stimuli/properties, virtual reality (VR) technology should be used to ensure that other environmental factors are stable and uniform in order to reduce the stimulation of other environmental factors. Therefore, this study aims to discuss the influence of the acoustic environment on patient physiological/psychological indicators and the mechanism of the effect on recovery using VR technology. A digital three-dimensional (3D) model of a hospital room was constructed, and experimental subjects wore VR glasses to visualize a real ward scene. Four typical sound categories were selected to analyze the effect of the acoustic environment on recovery; physiological indicators were monitored, and psychological factors were subjectively evaluated. The results show that music plays an important role in reducing stress as it can aid in a patient's physiological (skin conduction levels) and psychological stress recovery. Furthermore, mechanical and anthropogenic sounds exert negative effects on a patient's stress recovery. However, the effect is only limited to psychological stress indicators. The interaction effects of demographic characteristics and the acoustic environment are not significant, and future studies could consider the social–economic characteristics of patients. Based on these findings, we provide evidence that indicates that a hospital's acoustic environment is an important influencing factor on the stress recovery of patients and can serve as a reference for healthcare architects and policy makers.

Keywords: perceptual attributes, acoustic environment, patient feelings, heating region, hospital wards

INTRODUCTION

The indoor environment as a service carrier is most directly influenced by mental feelings, which are linked to patient comfort and mood. Patients stay in the ward almost all day, highlighting the great importance of providing comfortable conditions in hospital buildings. The comfort environment is considered the most important factor influencing patient feelings (Buckles, 1990). Researchers have measured the noise levels or studied the sound source of various healthcare environments, such as critical care wards (Xie et al., 2013), intensive care units (Xie et al., 2009), and entrance halls (Qin et al., 2011). As shown in previous studies, the noise levels measured in the wards frequently exceed the World Health Organization guideline values (45 dBA) by more than 20 dBA (Berglund et al., 1999; MacKenzie and Galbrun, 2007). Noise is a major public health issue, and noise annoyance is the most common and direct response among people exposed to environmental noise. Noise has been identified as a major stressor in hospitals (Farrehi et al., 2016) and will influence an individual's physical and mental health.

Literature Review

The documented association with several diseases and the growing number of exposed persons worldwide (Recio et al., 2016) indicate negative emotional and attitudinal reactions to noise (Okokon et al., 2015). Exposure to noise may interfere with daily activities, feelings, thoughts, rest, or sleep and may be accompanied by negative emotional reactions, such as irritability, distress, exhaustion, and other stress-related symptoms (Beutel et al., 2016). The impacts of stressors on health depend on the complex interactions between stressors and individual coping strategies, which are developed through previous experience, psychology, biology, social factors, competitive stressors, and personality (Jensen et al., 2018). Noise-related health problems are growing, and more severe effects related to cardiovascular morbidity and mortality have been proposed (Belojevic et al., 2011). Studies have found that an increase in daily noise levels of 1 dB(1) resulted in a 6.6% increase in the risk of death in the elderly (Tobías et al., 2015) and have observed a significant increase in blood pressure of 2–4 mmHg after 10 min of high-level exposure (Paunović et al., 2014). Studies have also reported that noise negatively impacts mental health (Hammersen et al., 2016; Jensen et al., 2018), which interacts with a wide range of complex elements, including biological, psychological, social, economic, and environmental factors (Barry and Friedli, 2010; Aletta et al., 2017). These factors include not only objectively measured environmental conditions but also subjective evaluation. When the noise level can no longer be reduced, people can still be annoyed by the noise because their subjective feelings can be affected by other psychoacoustic attributes, such as sharpness and roughness (Zwicker and Fastl, 1999). Noise and noise annoyance have non-standard effects on individuals that might depend on previous experiences or biological susceptibility. When individuals do not have control over the noise, as experienced with noise annoyance, they might suffer from learned helplessness and biological signatures of chronic stress, including overproduction of cortisol (Recio et al., 2016).

Since the formulation of eco-effective design (EED) and evidence-based design, the restorative effects of the environment have attracted wide attention (Ulrich et al., 2008; Shepley et al., 2009). As the primary facility for helping people to recover from illness, hospitals have also begun to focus on developing a healthy spatial environment utilizing natural forces. Through studying soundscapes, sounds in the environment have been regarded as a useful resource, and a favorable and healthy spatial environment can be created through discussing human perception and experience (Kang et al., 2016). As a result of people's perception of the acoustic environment, soundscapes can be positive (such as happy, calm) or negative (such as worry, pressure). The research on the effect of soundscape restoration is based on the development of attention restoration theory (ART) proposed by Kaplan and Kaplan (1989) and stress restoration theory (SRT) proposed by Ulrich et al. (1991). Weakening negative soundscapes is significantly related to health status, and increasing positive soundscapes is significantly related to environmental pressure recovery (Aletta et al., 2018a,b). A series of previous studies revealed that design and occupant choices can have positive health impacts by controlled reduction of noise levels (Evans, 2003; Von Lindern et al., 2016; Aletta et al., 2018c). It was also found that the natural environment had a positive effect on restoration processes (Hartig and Staats, 2003). However, the restorative effects of soundscapes should be correlated not only with subjective evaluation data but also with physiological parameters, including the emotions caused by sound stimulation (Hume and Ahtamad, 2013; Aletta et al., 2016a). Moreover, the soundscape is related to other spatial environmental factors. When people hear a sound, the perceived auditory space around them may modulate their emotional response to it. Small rooms are considered to be more pleasant, calmer, and safer than large rooms, and sounds originating behind listeners tend to be more arousing and elicit larger physiological changes than sources in front of the listeners (Tajadura-Jiménez et al., 2010). In their work on soundscapes in hospitals, researchers have revealed the relationship between the acoustic environment, typical sound sources, and geometry form (Xie and Kang, 2012b). An acoustic environment evaluation system has also been established (Xie and Kang, 2012a), and it has been found that the acoustic environment plays a leading role in the overall environmental evaluation (Wu et al., 2019). However, the impacts of hospital acoustic soundscapes on the physiological and psychological indices of patients require further study.

Perceptual experiences in one modality often depend on activity from other sensory modalities. The renewed interest in the topic of cross-modal correspondences that have emerged in recent years has motivated research that demonstrated that cross-modal matchings and mappings exist between most sensory dimensions (Deroy and Spence, 2016). Individuals reliably match different tastes/flavors

(KnFerle and Spence, 2012), colors (Hamilton-Fletcher et al., 2016), and shapes (Ozturk et al., 2013) to auditory stimuli. For example, individuals consistently match high-pitched sounds to small, bright objects located high up in space (Spence, 2011). In each experimental module, participants were experiencing different hospital indoor environments as the different experimental scenario conditions. Experimental scenarios can be classified as real or artificial. Due to site restrictions, it is difficult to effectively control a large number of irrelevant environmental factors, and it is also difficult to "add" a new environmental factor to the original indoor environment. Therefore, the experimental conditions of real scenarios are limited by their controllability (Stamps, 2007). To gain the most accurate and least variable estimate of acoustic environmental stimuli/properties, the stimulation of other environmental factors should be minimized. With the increasing maturity of virtual reality (VR) technology in recent years, VR environments can provide users with a more realistic and immersive environment (Chamilothori et al., 2019a). Multiple empirical studies show that the physiological, psychological, and behavioral feedback of participants in VR scenarios is similar to those in real scenarios (Heydarian et al., 2015). Yin et al. (2018) found that, in VR scenarios, the user's heart rate, blood pressure, skin conductivity, cognitive ability, and emotional level were very similar to those in real scenarios. Therefore, environmental psychologists began to use VR scenarios for environmental psychology experiments, rather than real scenarios.

Study Framework

This study aims to determine the following: (1) whether the acoustic environment can promote recovery in terms of physiological indicators—we hypothesize that physiological recovery will increase with music and decrease with artificial sounds and mechanical sounds; (2) whether sounds can decrease or increase the psychological function of patients in hospital wards—we hypothesize that music will be helpful for the psychological restoration of patients, as artificial and mechanical sounds will lead to the opposite trend; and (3) whether demographic factors and other environmental factors will cause different degrees of impact—we hypothesize that differences in demographic and environmental factors will lead to differences in the degree of the effect of soundscape recovery, as some previous studies indicated that there are differences between population and other environmental factors in the subjective evaluation of the acoustic environment. A digital three-dimensional (3D) model of a room was constructed, and experimental patients wore VR glasses to visualize the same ward scene and eliminate other visual and landscape distractions. Several different approaches were explored to meet the goals. First, the effect of sound stimuli on the physiological indices of the patients was examined. Second, the effect of sound stimuli on an individual's mental health was examined. Third, differences in the effects of sound on different populations and multiple environmental interactions were observed.

METHODOLOGY

In this study, a combination of physiological measurements and psychological evaluation was utilized. Four typical sound types were presented to experimental patients, and their physiological indicators were monitored by attached detectors. The patients wore VR glasses to observe the same virtual ward space and eliminate interference from other environments. The participants were asked to complete a subjective questionnaire. The obtained data were analyzed to evaluate the restorative effect of sounds in hospitals on individuals utilizing statistical methods.

Participants

The participants were all inpatients of the internal medicine department of the First Hospital of Harbin and the Second and Fourth Affiliated Hospitals of Harbin Medical University. Inpatients from internal medicine tend to have more time to participate in experiments than outpatients, and internal medicine is mainly related to chronic diseases; this provides an ideal experimental object that can exclude the psychological and physiological effects of diseases.

The participants were 70 patients with an average age of 48.2 (SD = 3.42; min = 18; max = 72), including 36 men and 34 women. The proportions of participants younger than 45, 45–60, and over 60 years old were similar to remove the effects of differences between participant groups on the experimental results (Zhang et al., 2016). The number of participants selected in this study was based on relevant experiments conducted in similar fields (Alvarsson et al., 2010; Annerstedt et al., 2013).

All participants were required to have clear cognitive consciousness and sufficient visual, auditory, and behavioral abilities to ensure that they could complete the physiological index measurement and questionnaire survey, and wore comfortable clothing. Additionally, patients with hyperthyroidism and supraventricular tachycardia were excluded from this experiment, as autonomic nervous dysfunction would decrease the accuracy of the measurement and evaluation of physiological stress indicators. The diet and sleep status of the participants needed to be stable. Six hours before the test, the patients did not drink, smoke, or have coffee or other drinks that would stimulate the sympathetic nervous system (Li and Kang, 2019).

The study was approved by the professors' associates in the School of Architecture at Harbin Institute of Technology. Written informed consent was obtained from all participants before the test began. Participants were informed about the goals and contents of the study, privacy, and data protection and that their participation in the study was voluntary. Biological samples were not collected.

Visual Scene

Many studies have been conducted on the influence of audio-visual factors on noise perception (Collignon et al., 2008; Yost and Zhong, 2015; Aletta et al., 2016b; Liu and Kang, 2018), and it has been demonstrated that vision and hearing can influence one other. Therefore, to prevent other factors influencing the patients' psychological and physiological indices, participants wore VR

A B

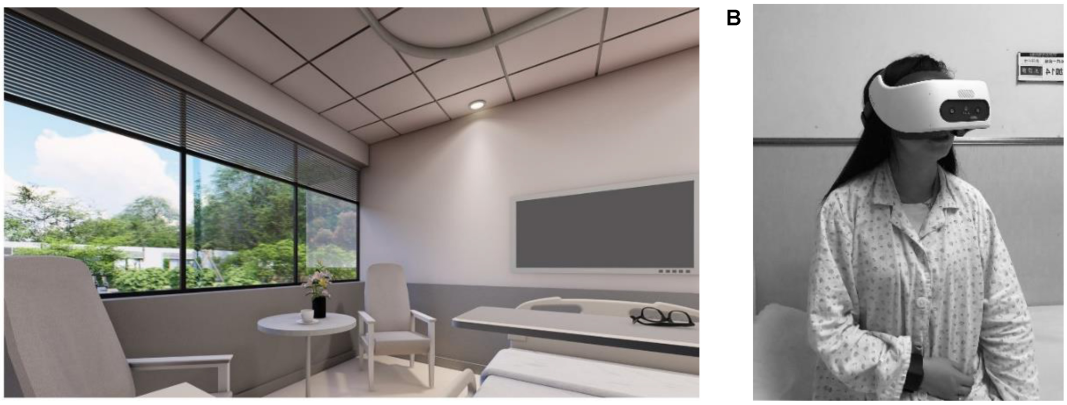


FIGURE 1 | Virtual reality (VR) device HTC Vive Focus Plus and presenting virtual scene. **(A)** Virtual scene of the ward. **(B)** Participant wearing VR.

glasses and observed the same ward scene. A standard single ward was selected as the experimental scene, as shown in **Figure 1A**. The simulated ward was 6.6 × 3.6 × 2.8 m, and the bed size was 2.1 × 0.9 m. A U-shaped rail curtain and 0.45 × 0.45 m bedside cabinet were set around the bed. Participants experienced the scene from the perspective of sitting on the bed in the ward and wore VR headsets, as shown in **Figure 1B**. The transformation of the virtual environment was based on the experimental condition transformation path of basic model construction–experimental parameter adjustment–virtual scene generation. First, a digital 3D model of the indoor space was established by 3DMAX. Then, according to the specific experimental goal, some of the design parameters of the model scenario were adjusted. Finally, the adjusted model was imported into the HTC Vive VR device.

Selection of Sound Stimuli

Common hospital sound sources can be summarized into four typical categories: mechanical, artificial, background, and music (Rashid and Zimring, 2008). Mechanical sounds are produced by hospital equipment, such as wheelbarrows, ventilators, and electrocardiograph monitors. Artificial sounds include patient conversations, children's crying, phones being answered, and other behaviors. Background sounds have no clear dominant source and include mechanical sounds produced by new air systems, elevator operations, and other equipment, as well as artificial sounds produced by the conversations and movements of doctors and patients. Mechanical, artificial, and natural sounds were recorded in the participants' hospital. According to Baker's classification standard of hospital background noise, the background should be stable, with its amplitude changing less than once per minute (<5 dB) (Baker, 1992). "Day and Night" was selected for the music stimulus, which is a particular piece of music that has been widely used in sound masking systems in hospitals (Ferguson et al., 1997; Mlinek and Pierce, 1997), and studies have demonstrated that it is popular among patients and played a positive role in their well-being, making them feel less tense, more relaxed, and safe (Thorgaard et al., 2004). "Day and Night" was favored by most inpatients, with 82% of the patients being very pleased/pleased with the song and 91% of participants defining the sound environment as very pleasant/pleasant (Bitten et al., 2017). Light music without lyrics was selected for this experiment, which avoids experimental deviation caused by the influence of lyrics on patients and their mental health (Baker and Bor, 2008).

The samples used in the experiments were recorded by SQuadriga II with BHS I, and the type of all four sound source samples can be clearly identified. A 5 min sample without dominant sound sources and only ambient noise was recorded for the control group. Five minutes of representative footage from each recording was used as the stimulation material for the experiment, as prolonged use of a VR headset would cause the subjects to become uncomfortable and interfere with the experimental results (Li and Kang, 2019). The 5 min equivalent sound pressure level (SPL) was adjusted to 50 dB(A) (Liu and Kang, 2018) for each audio frequency by AuditionCS6 to remove differences in volume during the stimulation of the four sounds. To ensure that the participants listened to the four auditory stimulus sounds under similar playback SPL conditions, the LAeq of the audio stimuli had been normalized by an artificial head to 50 dB(A) before the experiments to exclude the effect on arousal due to loudness. The background noise was below 45 dB(A) during the acoustic stimulation experiment, and nobody spoke in the room. Subjective loudness evaluation was carried out simultaneously; the results show that the loudness levels of different groups are significantly different due to the difference of their dominant source frequencies, but the loudness levels of different participants in the same group can be ignored.

Measurements

By using VR glasses to observe the 3D virtual hospital ward environment created by virtual simulation and headphones to listen to the four types of sound, participants can experience a more realistic hospital environment. The physiological recovery indices include heart rate and skin conductance, which were measured using the Empatica E4 physiological information monitoring equipment. Information regarding psychological recovery was obtained using a questionnaire. The scale is composed of two parts. The first is psychological feedback, including the anxiety state and perceived environmental restorativeness of the subjects. The anxiety states of participants were measured by using STAI-Y6 with eight questions to indicate their anxiety level (Zijlstra et al., 2017). Perceived restorativeness score (PRS) was adopted to evaluate the subjects' perceived restorativeness (Hartig et al., 1997). The second part refers to environmental appraisal. Perceived environmental quality index (PEQI) was used to describe the participants' perceived environmental quality (Fisher, 1974). The scale consists of a set of bipolar adjectives rated on a seven-point Likert scale, ranging from 1 (extreme negative perception of the environment) to 7 (extreme positive perception of the environment).

Procedure

The participants were asked to sit comfortably on a bed. The investigator explained the entire experimental procedure and asked the participants about their physical and mental states. After the subjects understood and agreed to all the terms, the investigator connected the HTC Vive Focus Plus VR device and Empatica E4. The experiment was started after the completion of the connection process and calibration of the physiological signal. The experimental process is shown in **Figure 2**. First, stress was induced in the subjects using the PASAT (packed audit serial addition task) program. The subjects then received a sound clip (one of the four types of sound sources), and the indoor scene was displayed by the VR equipment. After receiving one experimental condition, the subjects temporarily removed the VR equipment and completed the psychological recovery questionnaire, which took approximately 2 min. The subjects then rested for 2 min, accepted the next sound clip, and followed the process until the four experimental conditions were completed. The sequence of the four experimental conditions in the experiment followed a Latin square design (Morsbach et al., 1986). The Empatica E4 equipment continuously recorded the physiological recovery index data of the participants.

Data Analysis

Regarding physiological data transformation, the heart rate and skin conductance, which are the basic physiological indices of the human body, are easily affected by the physical differences between patients. For example, some individuals may have a relatively high basal heart rate or exhibit more intense physiological responses under the PASAT. Therefore, the physiological stress recovery level of participants cannot be compared by the mean values of physiological indices. In this study, we used the standardized physiological stress recovery rate (R) to estimate the influence of the acoustic environment on the individuals' physiological indices to reduce the potential experimental error caused by physical difference (Payne, 2013; Medvedev et al., 2015; Watts et al., 2016). As shown in formula (1), the R-value can be obtained by dividing the stress recovery level (the difference between the mean values of $F1$ and $F2$) by the stress arousal level (the difference between the mean values of $F1$ and $F0$). RHR and RSCL represent the stress recovery of the heart rate and skin conductance level, respectively. A higher R-value indicates that, under the experimental conditions, the subjects recover from physiological stress faster.

$$R = \frac{\overline{F_1} - \overline{F_2}}{\overline{F_1} - \overline{F_0}} \qquad (1)$$

For statistical analysis, IBM SPSS 25.0 was used to construct a database containing the final results (Meng et al., 2018; Ba and Kang, 2019). The data were analyzed by the following methods: (1) The differences between the physiological and psychological indicators measured at different times and for different sound source types were determined by repeated analysis of variance measurements, and the level of significance was set at $p < 0.05$. (2) Least significant difference (LSD) *post hoc* tests were conducted for pairwise comparisons. The effect sizes (partial η^2) were regarded as minimum, intermediate, and high at thresholds of 0.01, 0.06, and 0.14, respectively.

RESULTS

Effects of the Acoustic Environment on Physiological Stress Recovery

The results showed that the patient's mean heart rate recovery rates (R_{HR}) under the ambient noise, mechanical sound, artificial sound, and music conditions were 0.66 (SD = 0.11), 0.64 (SD = 0.11), 0.63 (SD = 0.12), and 0.68 (SD = 0.12), respectively. As shown in **Figure 3**, the patients' heart rates tended to recover faster under the music soundscape than the others. However, the repeated measures ANOVA results indicated that the main effect of the soundscape on heart rate recovery was not significant ($F = 1.35$, $p = 0.26$, partial $\eta^2 = 0.04$).

As assumed, the highest skin conductance level recovery rate (R_{SCL}; $M = 0.73$, $SD = 0.14$) was observed when the patients were exposed to the music soundscape condition. In contrast, the R_{SCL} of patients decreased by 2.5 and 4.6% under the mechanical and artificial noise conditions, respectively, when compared to the control group (ambient noise). The main effect of the soundscape on R_{SCL} was statistically significant ($F = 3.37$, $p = 0.02$), indicating that there was a significant difference in R_{SCL} during exposure to various experimental conditions. Additionally, according to the effect size exhibited by partial η^2(0.10), the soundscape could exert a substantial impact on R_{SCL}. In the multiple comparisons, a Bonferroni correction is applied. Given there are six sub-groups, 0.0083 is used as a corrected significance threshold. As shown in **Table 1**, the results of the LSD *post hoc* test confirmed that the recovery of the skin conductance level was faster under the music soundscape than under the mechanical (MD = 0.07) and artificial (MD = 0.08) noise conditions. Finally, although exposure to mechanical and artificial noise may lead to slower skin conductance level recovery than ambient noise, the difference was not significant. Combining the results of R_{HR} and R_{SCL} partly confirms our first hypothesis; i.e., the healthcare soundscape could impact the physiological stress

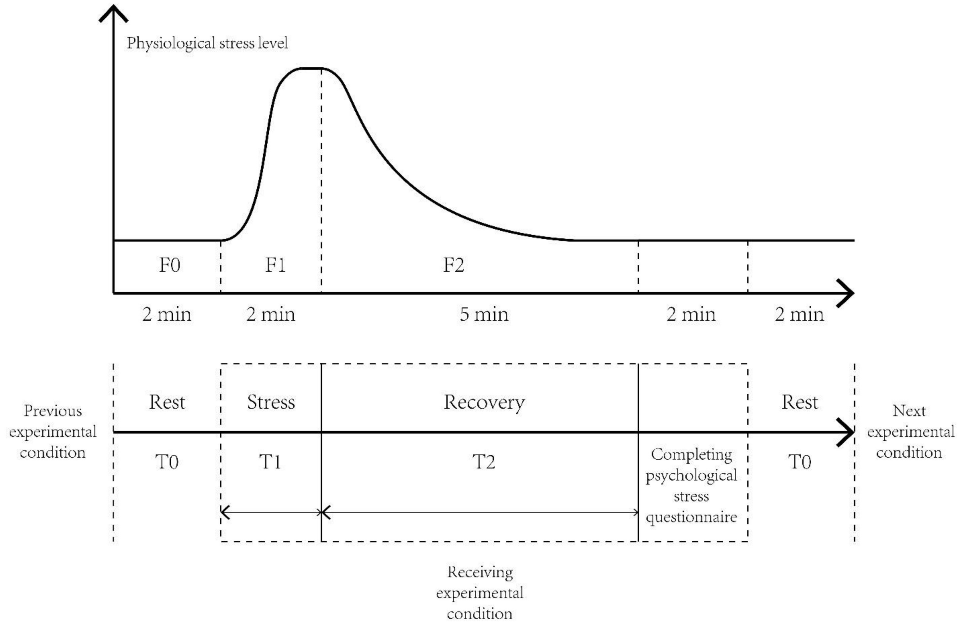


FIGURE 2 | Experimental process.

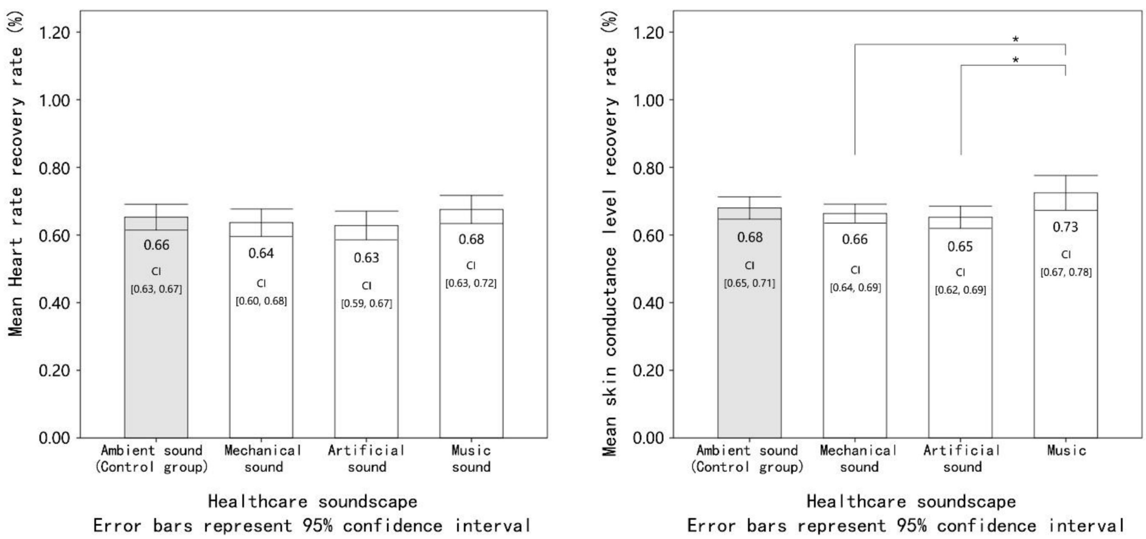


FIGURE 3 | The mean value of patients' physiological stress recovery indicators under exposure to different healthcare soundscapes. Error bars depict 95% confidence interval. *Significant at the 0.05 level. **Significant at the 0.01 level.

recovery response of patients but only in the case of skin conductance level.

Effects of the Acoustic Environment on Psychological Stress Recovery

The repeated measures ANOVA results suggested that the main effect of healthcare soundscapes on the anxiety state of patients was statistically significant ($F = 10.95$, $p = 0.00$, partial $\eta^2 = 0.26$). Music soundscapes could effectively reduce the patients' anxiety state. After experiencing the music soundscape, the anxiety states of the participants were 16.7, 14.4, and 24.5% lower than those under the ambient, mechanical, and artificial noise conditions, respectively. The LSD *post hoc* test indicated that these differences all reached statistical significance ($p < 0.008$). Additionally, artificial noise could cause patients to feel more anxious than the other three soundscapes. As shown in **Table 2**, the LSD *post hoc* test revealed that the difference in the anxiety score between mechanical noise and artificial noise was significant ($p = 0.007$), but that between the mechanical and control groups was not ($p = 0.0027$).

TABLE 1 | Pairwise comparison of physiological stress recovery levels in patients.

	Heart rate recovery rate				Skin conductance level recovery rate			
	MD	95% CI for difference		Sig	MD	95% CI for difference		Sig
		Lower bound	Upper bound			Lower bound	Upper bound	
Ambient sounds–mechanical sounds	0.02	−0.03	0.06	0.48	0.02	−0.02	0.06	0.40
Ambient sounds–artificial sounds	0.03	−0.02	0.07	0.27	0.03	−0.02	0.07	0.24
Ambient sounds–music sound	−0.02	−0.08	0.03	0.42	−0.05	−0.09	0.01	0.07
Mechanical sounds–artificial sounds	0.01	−0.04	0.05	0.70	0.01	−0.03	0.05	0.59
Mechanical sounds–music sound	−0.04	−0.10	0.02	0.19	−0.07	−0.12	−0.01	0.03(*)
Artificial sounds–music sound	−0.05	−0.10	0.01	0.09	−0.08	−0.14	−0.01	0.03(*)

**The mean difference is significant at the 0.05 level.*

TABLE 2 | Pairwise comparison of psychological stress recovery levels in patients.

	Self-reported anxiety state				Perceived restorativeness score			
	MD	95% CI for difference		Sig	MD	95% CI for difference		Sig
		Lower bound	Upper bound			Lower bound	Upper bound	
Ambient sounds–mechanical sounds	0.25	−0.55	1.05	0.530	2.09	0.28	3.91	0.04(*)
Ambient sounds–artificial sounds	−0.97	−1.82	−0.12	0.027(*)	0.78	−1.04	2.60	0.35
Ambient sounds–music sound	1.56	0.52	2.61	**0.01(**)**	−2.69	−4.51	−0.87	**0.00(**)**
Mechanical sounds–artificial sounds	−1.22	−2.08	−0.36	**0.01(**)**	−1.32	−3.13	0.51	0.20
Mechanical sounds–music sound	1.31	0.46	2.17	**0.00(**)**	−4.78	−6.60	−2.96	**0.00(**)**
Artificial sounds–music sound	2.53	1.52	3.54	**0.00(**)**	−3.46	−5.29	−1.65	**0.00(**)**

**The mean difference is significant at the 0.05 level. **The mean difference is significant at the 0.01 level. Values in bold represent the family-wise error rate corrected significant.*

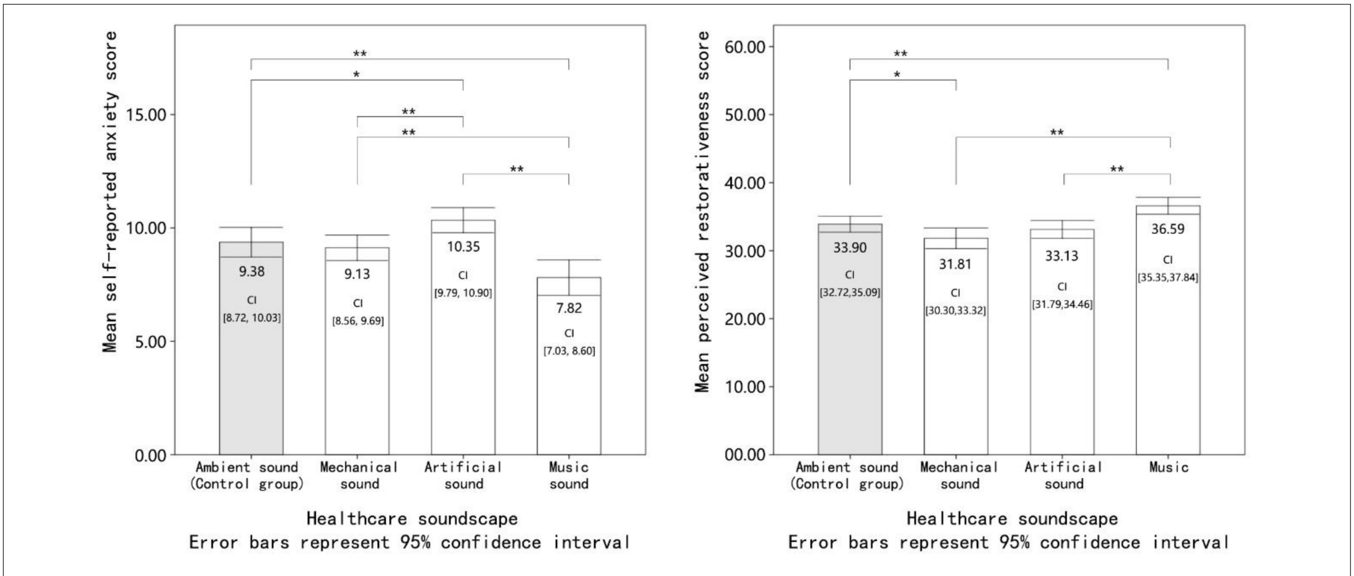


FIGURE 4 | The mean value of patients' psychological stress recovery indicators under exposure to different healthcare soundscapes. Error bars depict 95% confidence interval. *Significant at the 0.05 level. **Significant at the 0.01 level.

The environmental restorativeness scores given by patients significantly differed (F = 9.39, Sig = 0.00) under the ambient noise, mechanical noise, artificial, and music soundscape conditions. The effect size (partial η^2 = 0.23) suggested the substantial effect of the healthcare acoustic environment on the perceived environmental restorativeness scores. As shown in **Figure 4**, consistent with the anxiety result, when the music soundscape was broadcast, patients tended to perceive

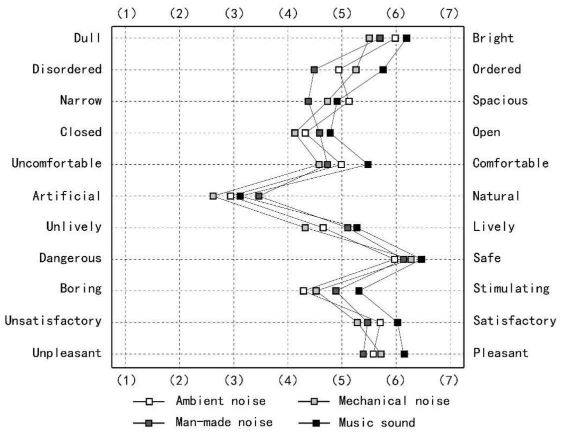


FIGURE 5 | Patients' bipolar adjective environmental appraisal scores in different healthcare acoustic environments.

the surrounding environment as "restorative." The percentage of improvement in the restorativeness scores, *i*, was used to estimate the benefit that certain soundscapes could bring to the participants, which can be calculated as follows: *i*music = (PRSmusic − PRSnoise)/PRSnoise, where *i*music is the percentage of improvement in the restorativeness scores from the music to the noise soundscape conditions, PRSmusic is the mean PRS under music soundscape conditions, and PRSnoise is the mean PRS under noise soundscape conditions. The restorativeness scores given to the music condition were 7.9, 15.0, and 10.5% higher than those given to the ambient, mechanical, and artificial noise conditions, respectively. The LSD *post hoc* test indicated that the restorativeness differences between music and the other three conditions were all significant ($p < 0.008$). In contrast, participants regarded mechanical noise as the least restorative soundscape ($M = 31.81$, $SD = 4.20$), but the difference between mechanical noise and the control group was insignificant.

The repeated measures ANOVA results showed that the main effects of healthcare soundscapes on the three environmental appraisal indices, i.e., sense of order, comfort, and stimulation, were significant. Under the music experimental condition, patients perceived the virtual environment as more orderly, comfortable, and stimulating (**Figure 5**). In contrast, participants experiencing the mechanical noise condition were more likely to use negative adjectives to describe the sound than under the control conditions. Specifically, when patients were exposed to mechanical noise, they evaluated the surrounding environment as narrower, closed, uncomfortable, artificial, and unlively. However, the difference was not significant in any of the environmental appraisal indices. Additionally, healthcare soundscapes may influence some visual environmental appraisal parameters. For example, the results indicated that patients considered the space to be more dull and narrow under artificial and mechanical noise, respectively.

Interaction Effects of the Acoustic Environment and Demographic Factors on Stress Recovery

As shown in **Figure 6**, mechanical noise appeared to exert a more negative impact on physiological stress recovery in female patients. Under the mechanical noise condition, the mean R_{HR} and R_{SCL} values of female patients were 1.68 and 3.09% lower

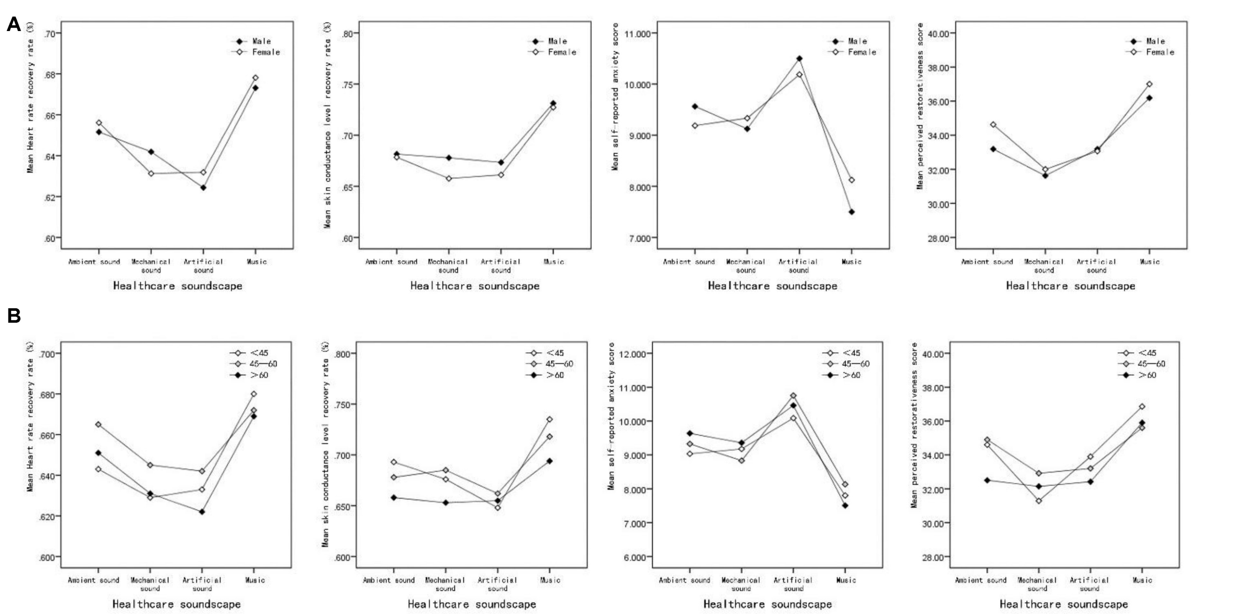


FIGURE 6 | (A) The interaction effects of gender and acoustic environment on patients' stress recovery. **(B)** The interaction effects of age and acoustic environment on patients' stress recovery.

TABLE 3 | MANOVA results for main and interaction effects on physiological and psychological recovery.

	Physiological stress recovery				Psychological stress recovery			
	R_{HR}		R_{SCL}		STAI		PRS	
	F	Sig	*F*	Sig	*F*	Sig	*F*	Sig
Gender	0.01	0.94	0.30	0.59	0.01	0.91	0.91	0.34
Gender*soundtype	0.04	0.99	0.05	0.99	0.57	0.64	0.26	0.86
Age	0.23	0.79	2.00	0.19	0.28	0.73	0.82	0.45
Age*soundtype	0.36	0.91	1.07	0.36	1.78	0.10	2.14	0.05

than those of male patients. However, the ANOVA results showed that the main effects of gender on R_{HR}, R_{SCL}, anxiety state, and perceived restorativeness state were not significant ($p < 0.05$). In addition, there was no significant interaction effect between gender and healthcare soundscape on physiological and psychological stress recovery, indicating that the street recovery outcomes of male and female patients in response to various healthcare soundscapes were similar.

As suggested by **Figure 6**, patients of different age groups tended to respond to different healthcare soundscapes similarly, although patients aged between 45 and 60 appeared to physiologically recover slightly faster than those in the other age groups. A two-way repeated measures ANOVA was conducted to examine the influence of age on the participants' physiological and psychological stress recovery parameters, and the results indicated that neither the main effect of the patients' age nor the interaction effect between age and soundscape condition was significant ($p < 0.05$), as shown in **Table 3**. The interaction effect between age and soundscape on the perceived restorativeness was almost significant ($p = 0.05$). Senior patients were less sensitive to the three types of healthcare noise, and young people (less than 45 years old) perceived the environment as less restorative under the mechanical noise condition.

The results showed that there was no significant interaction effect between soundscape and demographic characteristics on the participants' restoration; therefore, hypothesis (3) is rejected. Although we failed to identify any significant interaction effects, certain groups of participants exhibited some particular environmental feedback tendencies. For example, the restorative outcome of elderly people appeared to be less sensitive to acoustical conditions. Participants aged between 45 and 60 years tended to withstand negative sounds better than those in the other age groups. In future studies, we could consider the patients' socioeconomic characteristics in analysis to explore the potential interaction effects.

DISCUSSION

The results obtained from the skin conductance level partly support the theory that the healthcare soundscape could affect physiological stress recovery. The music soundscape has a restorative effect on healthcare. The results showed that the recovery rate of participants' SCL under the music condition was faster than that under the ambient, mechanical, and artificial sound conditions. However, none of these differences reached significance, indicating that the restorative effects of music are limited in the aspect of physiological stress.

Some previous studies observed a stronger restorative effect of music than this study (Medvedev et al., 2015). The experimental setting we adopted may cause different results, as VR conditions could draw the patients' attention to visual stimuli and weaken the restorative effect of soundscapes. Another possible explanation may be that the music used in this study was not self-selected. Researchers have confirmed that a participant's sense of control could improve restorative effects (Heitz et al., 1992). Additionally, the different musical tastes of patients could affect the restorative impact (Watts et al., 2016).

No significant impact of soundscape on the patients' R_{HR} was observed here, which is consistent with the results of some previous studies (Aletta et al., 2018a; Yu et al., 2018). This may be attributed to the special characteristics of heart rate, which is highly sensitive to the mode of information processing (Ulrich et al., 1991; Meehan et al., 2005). A person's heart rate accelerates dramatically if experimental conditions involve information processing, such as mental counting (Lacey and Lacey, 1974). This study involved no information storage, retrieval, or manipulation under any of the four soundscapes, which may have resulted in an insignificant effect on heart rate. Additionally, although the skin conductance level and heart rate are both indicators of sympathetic nervous system activity, their sensitivities when testing various built environmental stimuli may differ (Cacioppo et al., 2007). Studies have found that skin conductance is sensitive to changes in the luminous environment (Izso et al., 2009) and that heart rate is more responsive to the facade pattern (Chamilothori et al., 2019a).

The second hypothesis was confirmed by the significant effects of soundscape on the participants' anxiety level and perceived restorativeness state. Artificial sound induced higher anxiety than the mechanical experimental conditions in the study. A potential explanation for this result is that artificial sound contains more transient noise (Allaouchiche et al., 2002), which may negatively impact psychological recovery. However, the anxiety level of the participants experiencing the mechanical sound was not significantly higher than that of the participants in the control group. Although this is the first study in this area, the influence of the participants' acoustic expectations may explain this result. The potential function of space (such as socializing and working) could influence a user's environmental expectations and evaluations (Chamilothori et al., 2019b). In this study, healthcare was chosen as the research context, and participants may anticipate certain kinds of mechanical noise before being subjected to the experimental conditions. Thus, its negative impact on recovery could be relieved.

The perceived restorativeness state was also significantly influenced by healthcare soundscapes. Mechanical noise was perceived as the least restorative condition in the study, which was inconsistent with the anxiety state data. This might be due to the different assessment weights between the two psychological recovery indicators. The anxiety state assesses the participants' mental state. For example, the scale we used included items such as "I feel upset" (Marteau and Bekker, 1992). However, the perceived restorativeness state reflects the external environment appraisal. In this study, both parameters may partly reflect the impact of the soundscape on the patients' psychological states, but more work should be conducted to determine the mechanisms and pathways of the effects of healthcare soundscapes on psychological stress recovery.

This research assessed the effects of the acoustic environment on patients' environmental appraisal in a healthcare facility, and the data show that healthcare soundscapes may have affected the participants' environmental appraisal. The music condition was perceived as being more positive than the three other experimental conditions for nine of the 11 evaluation dimensions and could significantly improve the patients' environmental perception in terms of the order, comfort, and stimulation. However, no significant difference was observed between the evaluation of mechanical, artificial, and ambient noise. Overall, the soundscape has less of an impact on patients' environmental appraisal than visual information, such as color, lighting, and spatial layout (Leather et al., 2003). The smaller effect may indicate that visual stimulation is a dominating factor affecting environmental appraisal in healthcare settings.

This study found that the soundscape could alter the patients' visual impression of the environment, such as their sense of light, order, and scale. This could be because people holistically perceive the environment, and audio and visual stimuli could drive multisensory environmental perception (Viollon et al., 2002; Pheasant et al., 2010). Attractive or meaningful visual contexts tend to increase people's tolerance to noise, causing less irritation in similar noisy acoustic environments (Bangjun et al., 2003; Iachini et al., 2012). However, studies have also observed a high correlation between audio information and an individual's visual experience and preference. In this study, the participants tended to regard the surrounding environment as orderly, comfortable, and stimulating under the music condition, which may be because the sound stimuli altered the participants' visual cognitive processing (Ren and Kang, 2015).

Generally, the objective data were relatively consistent with the subjective ratings (anxiety, perceived restorativeness state, and environmental appraisal), which could verify the validity of the method used in the study. However, when faced with audio stimuli, the psychological stress recovery indicator tended to be more sensitive than physiological parameters. The effect size also indicates that the soundscape could exert a greater effect on the outcome of psychological factors, supporting the results of previous studies. This may be because physiological stress recovery parameters, such as heart rate, skin conduction levels, or blood pressure, are indices of sympathetic arousal and cannot reflect the valence of emotion (Ward and Marsden, 2003). Therefore, physiological data cannot detect mild arousal responses coupled with positive emotional reactions. Therefore, interpreting a person's stress level using physiological data alone is insufficient, especially considering the stress reaction, and recovery is a complex process involving cognition and reflection (Bartlett, 1998).

CONCLUSION

Inpatients are more prone to anxiety and stress than healthy individuals; therefore, hospital wards must provide suitable acoustic environments to help them to relax and recover. The restorative effects of soundscapes have been investigated, but few studies have been conducted on patients and hospital environments. This study mainly explores and analyzes the influence of the acoustic environment on physiological/psychological stress indicators of patients in hospital wards.

The impact of soundscape on patients' physiological stress parameters was relatively modest. In this study, sounds did not significantly impact the patients' heart rate recovery rates (R_{HR}). However, the results demonstrate that the soundscape could significantly influence the patients' skin conductance level recovery rate (R_{SCL}). The recovery rate was faster under music than the mechanical or artificial noise conditions, though the difference fails to reach significance.

The acoustic environment could exert profound effects on the patients' psychological stress indicators, with both the patients' self-reported anxiety state and PRS significantly affected by the healthcare soundscapes. Patients continuously reported less anxiety and higher perceived restorativeness for the music soundscape than the ambient, mechanical, and artificial noise soundscapes. The reported anxiety levels were highest under the artificial sounds, and mechanical sounds were regarded as the least restorative. For the environmental appraisal of psychological parameters, the music condition was described as significantly more ordered, comfortable, and stimulating than the three other experimental conditions. There was no statistically significant difference between the environmental appraisal of mechanical, artificial, and ambient sounds. However, it was found that the acoustic environment could alter the patients' visual impression of the environment.

The interaction effects of gender, age, and acoustic environment were not significant. However, there were some environmental feedback tendencies for certain groups of participants, and future studies may consider the patients' social–economic characteristics. Hospital spaces are rather diverse;

thus, it would be interesting to consider other spaces, such as outpatient halls, waiting rooms, double beds, and dormitory bed wards, in future studies. While this study indicated that the acoustic environment of hospital wards influences the physiological and psychological indices of patients, and also demonstrated that VR is an effective method of analyzing the relative influences of different dominant sound sources, in future work, the absolute influence of the acoustic environment on the psychological and physiological indicators of patients could be examined in realistic environments.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants. Written informed consent was obtained from the individual for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

TZ, YW, QM, and JK contributed to conceptualization, methodology, and visualization. TZ contributed to investigation, formal analysis, and data curation. TZ and YW contributed to writing original draft preparation. TZ contributed to review and editing. YW, QM, and JK contributed to supervision. YW and QM contributed to administration and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the National Natural Science Foundation of China (NSFC) (51808160 and 51878210) and the Fundamental Research Funds for the Central Universities (HIT.NSRIF.2020035).

ACKNOWLEDGMENTS

The authors want to thank the Internal Medicine Department of the First Hospital of Harbin and the Second and Fourth Affiliated Hospitals of Harbin Medical University in Harbin, China, for their permission for soundscape investigation, respectively.

REFERENCES

Aletta, F., Axelsson, Ö., and Kang, J. (2017). Dimensions underlying the perceived similarity of acoustic environments. *Front. Psychol.* 8:1162. doi: 10.3389/fpsyg.2017.01162

Aletta, F., Kang, J., and Axelsson, Ö. (2016a). Soundscape descriptors and a conceptual framework for developing predictive soundscape models. *Landsc. Urban Plan.* 149, 65–74. doi: 10.1016/j.landurbplan.2016.02.001

Aletta, F., Masullo, M., Maffei, L., and Kang, J. (2016b). The effect of vision on the perception of the noise produced by a chiller in a common living environment. *Noise Control Eng. J.* 64, 363–378. doi: 10.3397/1/3763786

Aletta, F., Oberman, T., and Kang, J. (2018a). Associations between positive health-related effects and soundscapes perceptual constructs: a systematic review. *Int. J. Environ. Res. Public Health* 15:2392. doi: 10.3390/ijerph15112392

Aletta, F., Oberman, T., and Kang, J. (2018b). Positive health-related effects of perceiving urban soundscapes: a systematic review. *Lancet* 392:S3. doi: 10.3390/ijerph15112392

Aletta, F., Van Renterghem, T., and Botteldooren, D. (2018c). Influence of personal factors on sound perception and overall experience in urban green areas. A case study of a cycling path highly exposed to road traffic noise. *Int. J. Environ. Res. Public Health* 15:1118. doi: 10.3390/ijerph15061118

Allaouchiche, B., Duflo, F., Debon, R., Bergeret, A., and Chassard, D. (2002). Noise in the postanaesthesia care unit. *Br. J. Anaesth.* 88, 369–373. doi: 10.1093/bja/88.3.369

Alvarsson, J. J., Wiens, S., and Nilsson, M. E. (2010). Stress recovery during exposure to nature sound and environmental noise. *Int. J. Environ. Res. Public Health* 7, 1036–1046. doi: 10.3390/ijerph7031036

Annerstedt, M., Jönsson, P., Wallergård, M., Johansson, G., Karlson, B., Grahn, P., et al. (2013). Inducing physiological stress recovery with sounds of nature in a virtual reality forest — Results from a pilot study. *Physiol. Behav.* 118, 240–250. doi: 10.1016/j.physbeh.2013.05.023

Ba, M., and Kang, J. (2019). Effect of a fragrant tree on the perception of traffic noise. *Build. Environ.* 156, 147–155. doi: 10.1016/j.buildenv.2019.04.022

Baker, C. F. (1992). Discomfort to environmental noise: heart rate responses of SICU patients. *Crit. Care Nurs. Q.* 15, 75–90.

Baker, F., and Bor, W. (2008). Can music preference indicate mental health status in young people? *Australas. Psychiatry* 16, 284–288. doi: 10.1080/10398560701879589

Bangjun, Z., Lili, S., and Guoqing, D. (2003). The influence of the visibility of the source on the subjective annoyance due to its noise. *Appl. Acoust.* 64, 1205–1215. doi: 10.1016/s0003-682x(03)00074-4

Barry, M., and Friedli, L. (2010). "The influence of social, demographic and physical factors on positive mental health in children, adults and older people," in *Mental Capital and Wellbeing*, eds C. L. Cooper, J. Field, U. Goswami, R. Jenkins, and B. J. Sahakian (Hoboken, NJ: Wiley-Blackwell), 475–484.

Bartlett, D. (1998). *Stress: Perspectives and Processes*. Buckingham: McGraw-Hill Education.

Belojevic, G., Paunovic, K., Jakovljevic, B., Stojanov, V., Ilic, J., Slepcevic, V., et al. (2011). Cardiovascular effects of environmental noise: research in Serbia. *Noise Health* 13, 217–220. doi: 10.4103/1463-1741.80156

Berglund, B., Lindvall, T., Schwela, D. H., and Organization, W. H. (1999). *Guidelines for Community Noise*. Geneva: World Health Organization.

Beutel, M. E., Jünger, C., Klein, E. M., Wild, P., Lackner, K., Blettner, M., et al. (2016). Noise annoyance is associated with depression and anxiety in the general population-the contribution of aircraft noise. *PLoS One* 11:e0155357. doi: 10.1371/journal.pone.0155357

Bitten, T., Birgitte, B. H., Gunhild, P., and Inger, T. (2017). Specially selected music in the cardiac laboratory—an important tool for improvement of the wellbeing of patients. *Eur. J. Cardiovasc. Nurs.* 3, 21–26.

Buckles, E. (1990). Evaluation of patient satisfaction in A&E. *Nurs. Stand.* 4, 33–35. doi: 10.7748/ns.4.19.33.s40

Cacioppo, J. T., Tassinary, L. G., and Berntson, G. (2007). *Handbook of Psychophysiology*. Cambridge: Cambridge University Press.

Chamilothori, K., Chinazzo, G., Rodrigues, J., Dan-Glauser, E., Wienold, J., and Andersen, M. (2019a). Subjective and physiological responses to façade and sunlight pattern geometry in virtual reality. *Build. Environ.* 150, 144–155. doi: 10.1016/j.buildenv.2019.01.009

Chamilothori, K., Wienold, J., and Andersen, M. (2019b). Adequacy of immersive virtual reality for the perception of daylit spaces: comparison of real and virtual environments. *Leukos* 15, 203–226. doi: 10.1080/15502724.2017.1404918

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., et al. (2008). Audio-visual integration of emotion expression. *Brain Res.* 1242, 126–135. doi: 10.1016/j.brainres.2008.04.023

Deroy, O., and Spence, C. (2016). Crossmodal correspondences: four challenges. *Multisens. Res.* 29, 29–48. doi: 10.1163/22134808-00002488

Evans, G. W. (2003). The built environment and mental health. *J. Urban Health* 80, 536–555. doi: 10.1093/jurban/jtg063

Farrehi, P. M., Nallamothu, B. K., and Navvab, M. (2016). Reducing hospital noise with sound acoustic panels and diffusion: a controlled study. *BMJ Qual. Saf.* 25, 644–646. doi: 10.1136/bmjqs-2015-004205

Ferguson, E., Singh, A. P., and Cunningham-Snell, N. (1997). Stress and blood donation: effects of music and previous donation experience. *Br. J. Psychol.* 88(Pt 2), 277–294. doi: 10.1111/j.2044-8295.1997.tb02635.x

Fisher, J. D. (1974). Situation specific variables as determinants of perceived environmental aesthetic quality and perceived crowdedness. *J. Res. Pers.* 8, 177–188. doi: 10.1016/0092-6566(74)90019-1

Hamilton-Fletcher, G., Ward, J., and Wright, T. D. (2016). Cross-modal correspondences enhance performance on a colour-to-sound sensory substitution device. *Multisens. Res.* 29, 337–363. doi: 10.1163/22134808-00002519

Hammersen, F., Niemann, H., and Hoebel, J. (2016). Environmental noise annoyance and mental health in adults: findings from the cross-sectional German Health Update (GEDA) Study 2012. *Int. J. Environ. Res. Public Health* 13:954. doi: 10.3390/ijerph13100954

Hartig, T., Korpela, K., Evans, G. W., and Gärling, T. (1997). A measure of restorative quality in environments. *Scand. Hous. Plan. Res.* 14, 175–194. doi: 10.1080/02815739708730435

Hartig, T., and Staats, H. (2003). Guest Editors' introduction: restorative environments. *J. Environ. Psychol.* 23, 103–107. doi: 10.1016/s0272-4944(02)00108-1

Heitz, L., Symreng, T., and Scamman, F. (1992). Effect of music therapy in the postanesthesia care unit: a nursing intervention. *J. Post Anesth. Nurs.* 7, 22–31.

Heydarian, A., Carneiro, J. P., Gerber, D., Becerik-Gerber, B., Hayes, T., and Wood, W. (2015). Immersive virtual environments versus physical built environments: a benchmarking study for building design and user-built environment explorations. *Autom. Constr.* 54, 116–126. doi: 10.1016/j.autcon.2015.03.020

Hume, K., and Ahtamad, M. (2013). Physiological responses to and subjective estimates of soundscape elements. *Appl. Acoust.* 74, 275–281. doi: 10.1016/j.apacoust.2011.10.009

Iachini, T., Maffei, L., Ruotolo, F., Senese, V. P., Ruggiero, G., Masullo, M., et al. (2012). Multisensory assessment of acoustic comfort aboard metros: a virtual reality study. *Appl. Cogn. Psychol.* 26, 757–767. doi: 10.1002/acp.2856

Izso, L., Láng, E., Laufer, L., Suplicz, S., and Horváth, Á. (2009). Psychophysiological, performance and subjective correlates of different lighting conditions. *Light. Res. Technol.* 41, 349–360. doi: 10.1177/1477153509336798

Jensen, H. A., Rasmussen, B., and Ekholm, O. (2018). Neighbour and traffic noise annoyance: a nationwide study of associated mental health and perceived stress. *Eur. J. Public Health* 28, 1050–1055. doi: 10.1093/eurpub/cky091

Kang, J., Aletta, F., Gjestland, T. T., Brown, L. A., Botteldooren, D., Schulte-Fortkamp, B., et al. (2016). Ten questions on the soundscapes of the built environment. *Build. Environ.* 108, 284–294. doi: 10.1016/j.buildenv.2016.08.011

Kaplan, R., and Kaplan, S. (1989). *The Experience of Nature: A Psychological Perspective*. New York, NY: Cambridge University Press.

KnFerle, K., and Spence, C. (2012). Crossmodal correspondences between sounds and tastes. *Psychon. Bull. Rev.* 19, 1–15. doi: 10.3758/s13423-012-0321-z

Lacey, B. C., and Lacey, J. I. (1974). "Studies of heart rate and other bodily processes in sensorimotor behavior," in *Cardiovascular Psychophysiology: Current Issues in Response Mechanisms, Biofeedback and Methodology*, eds P. A. Obrist, A. H. Black, J. Brener, and L. V. DiCara (New Brunswick, NJ: Aldine Transaction), 538–564. doi: 10.4324/9781315081762-31

Leather, P., Beale, D., Santos, A., Watts, J., and Lee, L. (2003). Outcomes of environmental appraisal of different hospital waiting areas. *Environ. Behav.* 35, 842–869. doi: 10.1177/0013916503254777

Li, Z., and Kang, J. (2019). Sensitivity analysis of changes in human physiological indicators observed in soundscapes. *Landsc. Urban Plan.* 190:103593. doi: 10.1016/j.landurbplan.2019.103593

Liu, F., and Kang, J. (2018). Relationship between street scale and subjective assessment of audio-visual environment comfort based on 3D virtual reality and dual-channel acoustic tests. *Build. Environ.* 129, 35–45. doi: 10.1016/j.buildenv.2017.11.040

MacKenzie, D., and Galbrun, L. (2007). Noise levels and noise sources in acute care hospital wards. *Build. Serv. Eng. Res. Technol.* 28, 117–131. doi: 10.1177/0143624406074468

Marteau, T. M., and Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). *Br. J. Clin. Psychol.* 31, 301–306. doi: 10.1111/j.2044-8260.1992.tb00997.x

Medvedev, O., Shepherd, D., and Hautus, M. J. (2015). The restorative potential of soundscapes: a physiological investigation. *Appl. Acoust.* 96, 20–26. doi: 10.1016/j.apacoust.2015.03.004

Meehan, M., Razzaque, S., Insko, B., Whitton, M., and Brooks, F. P. (2005). Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments. *Appl. Psychophysiol. Biofeedback* 30, 239–258. doi: 10.1007/s10484-005-6381-3

Meng, Q., Zhao, T., and Kang, J. (2018). Influence of music on the behaviors of crowd in urban open public spaces. *Front. Psychol.* 9:596. doi: 10.3389/fpsyg.2018.00596

Mlinek, E. J., and Pierce, J. (1997). Confidentiality and privacy breaches in a university hospital emergency department. *Acad. Emerg. Med.* 4, 1142–1146. doi: 10.1111/j.1553-2712.1997.tb03697.x

Morsbach, G., McCulloch, M., and Clark, A. (1986). Infant crying as a potential stressor concerning mothers' concentration ability. *Psychologia* 29, 18–20.

Okokon, E. O., Turunen, A. W., Ung-Lanki, S., Vartiainen, A.-K., Tiittanen, P., and Lanki, T. (2015). Road-traffic noise: annoyance, risk perception, and noise sensitivity in the Finnish adult population. *Int. J. Environ. Res. Public Health* 12, 5712–5734. doi: 10.3390/ijerph120605712

Ozturk, O., Krehm, M., and Vouloumanos, A. (2013). Sound symbolism in infancy: evidence for sound–shape cross-modal correspondences in 4-month-olds. *J. Exp. Child Psychol.* 114, 173–186. doi: 10.1016/j.jecp.2012.05.004

Paunović, K., Stojanov, V., Jakovljević, B., and Belojević, G. (2014). Thoracic bioelectrical impedance assessment of the hemodynamic reactions to recorded road-traffic noise in young adults. *Environ. Res.* 129, 52–58. doi: 10.1016/j.envres.2014.01.001

Payne, S. R. (2013). The production of a perceived restorativeness soundscape scale. *Appl. Acoust.* 74, 255–263. doi: 10.1016/j.apacoust.2011.11.005

Pheasant, R. J., Fisher, M. N., Watts, G. R., Whitaker, D. J., and Horoshenkov, K. V. (2010). The importance of auditory-visual interaction in the construction of 'tranquil space'. *J. Environ. Psychol.* 30, 501–509. doi: 10.1016/j.jenvp.2010.03.006

Qin, X., Kang, J., and Jin, H. (2011). Subjective evaluation of acoustic environment of waiting areas in general hospitals. *Build. Sci.* 12, 53–60.

Rashid, M., and Zimring, C. (2008). A review of the empirical literature on the relationships between indoor environment and stress in health care and office settings: problems and prospects of sharing evidence. *Environ. Behav.* 40, 151–190. doi: 10.1177/0013916507311550

Recio, A., Linares, C., Banegas, J. R., and Díaz, J. (2016). Road traffic noise effects on cardiovascular, respiratory, and metabolic health: an integrative model of biological mechanisms. *Environ. Res.* 146, 359–370. doi: 10.1016/j.envres.2015.12.036

Ren, X., and Kang, J. (2015). Interactions between landscape elements and tranquility evaluation based on eye tracking experiments. *J. Acoust. Soc. Am.* 138, 3019–3022. doi: 10.1121/1.4934955

Shepley, M. M., Baum, M., Ginsberg, R., and Rostenberg, B. (2009). Eco-effective design and evidence-based design: perceived synergy and conflict. *HERD* 2, 56–70. doi: 10.1177/193758670900200305

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Atten. Percept. Psychophys.* 73, 971–995. doi: 10.3758/s13414-010-0073-7

Stamps, A. E. III (2007). Evaluating spaciousness in static and dynamic media. *Des. Stud.* 28, 535–557. doi: 10.1016/j.destud.2007.01.001

Tajadura-Jiménez, A., Larsson, P., Väljamäe, A., Västfjäll, D., and Kleiner, M. (2010). When room size matters: acoustic influences on emotional responses to sounds. *Emotion* 10, 416–422. doi: 10.1037/a0018423

Thorgaard, B., Henriksen, B. B., Pedersbaek, G., and Thomsen, I. (2004). Specially selected music in the cardiac laboratory—an important tool for improvement of the wellbeing of patients. *Eur. J. Cardiovasc. Nurs.* 3, 21–26. doi: 10.1016/j.ejcnurse.2003.10.001

Tobías, A., Recio, A., Díaz, J., and Linares, C. (2015). Noise levels and cardiovascular mortality: a case-crossover analysis. *Eur. J. Prev. Cardiol.* 22, 496–502. doi: 10.1177/2047487314528108

Ulrich, R. S., Simons, R. F., Losito, B. D., Fiorito, E., Miles, M. A., and Zelson, M. (1991). Stress recovery during exposure to natural and urban environments. *J. Environ. Psychol.* 11, 201–230. doi: 10.1016/s0272-4944(05)80184-7

Ulrich, R. S., Zimring, C., Zhu, X., DuBose, J., Seo, H.-B., Choi, Y.-S., et al. (2008). A review of the research literature on evidence-based healthcare design. *HERD* 1, 61–125. doi: 10.1177/193758670800100306

Viollon, S., Lavandier, C., and Drake, C. (2002). Influence of visual setting on sound ratings in an urban environment. *Appl. Acoust.* 63, 493–511. doi: 10.1016/s0003-682x(01)00053-6

Von Lindern, E., Hartig, T., and Lercher, P. (2016). Traffic-related exposures, constrained restoration, and health in the residential context. *Health Place* 39, 92–100. doi: 10.1016/j.healthplace.2015.12.003

Ward, R. D., and Marsden, P. H. (2003). Physiological responses to different WEB page designs. *Int. J. Hum. Comput. Stud.* 59, 199–212. doi: 10.1016/s1071-5819(03)00019-3

Watts, G., Khan, A., and Pheasant, R. (2016). Influence of soundscape and interior design on anxiety and perceived tranquillity of patients in a healthcare setting. *Appl. Acoust.* 104, 135–141. doi: 10.1016/j.apacoust.2015.11.007

Wu, Y., Meng, Q., Li, L., and Mu, J. (2019). Interaction between sound and thermal influences on patient comfort in the hospitals of China's Northern heating region. *Appl. Sci.* 9:5551. doi: 10.3390/app9245551

Xie, H., and Kang, J. (2012a). Sound field of typical single-bed hospital wards. *Appl. Acoust.* 73, 884–892. doi: 10.1016/j.apacoust.2012.03.005

Xie, H., and Kang, J. (2012b). The acoustic environment of intensive care wards based on long period nocturnal measurements. *Noise Health* 14, 230–236. doi: 10.4103/1463-1741.102960

Xie, H., Kang, J., and Mills, G. H. (2009). Clinical review: the impact of noise on patients' sleep and the effectiveness of noise reduction strategies in intensive care units. *Crit. Care* 13:208. doi: 10.1186/cc7154

Xie, H., Kang, J., and Mills, G. H. (2013). Behavior observation of major noise sources in critical care wards. *J. Crit. Care* 28, e1105–e1109. doi: 10.1016/j.jcrc.2013.06.006

Yin, J., Zhu, S., Macnaughton, P., Allen, J. G., and Spengler, J. D. (2018). Physiological and cognitive performance of exposure to biophilic indoor environment. *Build. Environ.* 132, 255–262. doi: 10.1016/j.buildenv.2018.01.006

Yost, W., and Zhong, X. (2015). Localizing sound sources when the listener moves: vision required. *J. Acoust. Soc. Am.* 137, 2373–2373. doi: 10.1121/1.4920624

Yu, C.-P., Lee, H.-Y., and Luo, X.-Y. (2018). The effect of virtual reality forest and urban environments on physiological and psychological responses. *Urban For. Urban Green.* 35, 106–114. doi: 10.1016/j.ufug.2018.08.013

Zhang, X., Lian, Z., and Ding, Q. (2016). Investigation variance in human psychological responses to wooden indoor environments. *Build. Environ.* 109, 58–67. doi: 10.1016/j.buildenv.2016.09.014

Zijlstra, E., Hagedoorn, M., Krijnen, W. P., Van der Schan, P., and Mobach, M. P. (2017). Motion nature projection reduces patient\"s psycho-physiological anxiety during ct imaging. *J. Environ. Psychol.* 53, 168–176. doi: 10.1016/j.jenvp.2017.07.010

Zwicker, E., and Fastl, H. (1999). Psychoacoustics. Facts and models. *Phys. Today* 54, 64–65. doi: 10.1063/1.1387599

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Sustainable Cities

ORIGINAL RESEARCH
published: 02 September 2020
doi: 10.3389/frsc.2020.00041

Social Media and Open Data to Quantify the Effects of Noise on Health

Luis Gasco[1], Rossano Schifanella[2,3], Luca Maria Aiello[4], Daniele Quercia[4,5], Cesar Asensio[1] and Guillermo de Arcas[1]*

[1] Universidad Politécnica de Madrid, Madrid, Spain, [2] Computer Science Department, University of Turin, Turin, Italy, [3] ISI Foundation, Turin, Italy, [4] Nokia Bell Labs, Cambridge, United Kingdom, [5] CUSP, King's College London, London, United Kingdom

OPEN ACCESS

Edited by:
Bert De Coensel, Ghent University, Belgium

Reviewed by:
Jin Yong Jeon, Hanyang University, South Korea
Edda Bild, McGill University, Canada
Pyoung Jik Lee, University of Liverpool, United Kingdom

****Correspondence:***
Rossano Schifanella
rossano.schifanella@unito.it

Specialty section:
This article was submitted to Health and Cities, a section of the journal Frontiers in Sustainable Cities

Received: *04 April 2020*
Accepted: *21 July 2020*
Published: *02 September 2020*

Citation:
Gasco L, Schifanella R, Aiello LM, Quercia D, Asensio C and de Arcas G (2020) Social Media and Open Data to Quantify the Effects of Noise on Health. Front. Sustain. Cities 2:41. doi: 10.3389/frsc.2020.00041

Noise is considered the second factor after air pollution to impact citizens' health and well-being in densely populated urban areas, as it takes a heavy toll on the health of the circulatory and nervous systems. Traditionally, research on urban noise was conducted through surveys with a limited temporal and spatial coverage, and focused on a subset of the wide spectrum of sounds sources present in an urban environment. To overcome these limitations, we use geo-referenced social media images from Flickr to characterize the soundscape of London at scale. We build a model that uses socioeconomic variables, official noise exposure levels, and the soundscape estimated from social media to predict at area level the prevalence of hypertension—a cardiovascular condition that is widely studied in connection to high noise exposure. We consistently observe that socioeconomic variables, such as age, gender, and income, play an important role in explaining hypertension rates. Official noise exposure levels add a relatively limited contribution in predicting the health outcome. On the contrary, the social media soundscape information considerably improves the model performance. This result speaks to the value of integrating social media data into strategic noise maps for enhancing their predictive power; it also hints at the fact that the presence (or absence) of specific types of sounds might be a better indicator of hypertension prevalence than noise levels themselves.

Keywords: noise, health, Flickr, hypertension, social media, city

INTRODUCTION

More than two-thirds of the world's population will live in cities by 2050 (United Nations, 2018). A significant portion of the population shift is directed to large metropolitan hubs in the global economic market, as they provide greater opportunities to their citizens for professional and social development (Sassen, 1991). Population growth is bringing key challenges to policy makers. For example, the rise in rental prices, partly due to the proliferation of short-term rentals (Wachsmuth and Weisler, 2018; Urquiaga et al., 2020) and the increasing cost of living in city centers (Andersson and Turner, 2014; Florida, 2017) have accelerated the process of gentrification. Over the years, the need of long-range commute within the city and the increased number of private vehicles on the streets heavily interfered with the implementation of effective policies for a better spatial organization of our cities and in the deployment of effective road infrastructures and public transportation services (Wallsten, 2015; Barthelemy, 2016). Consequently, traffic congestion and the increase in the environmental pollution have become one of the most important challenges for politicians and planners due to its connection to quality of life and health outcomes.

Noise is considered the second threat after air pollution that most affects our health and well-being in densely populated urban areas (European Environment Agency, 2014). Noise pollution is a health hazard that is connected to the circulatory and nervous systems. Cross-sectional studies based on surveys conducted on population samples in different cities around the world have shown a possible association between noise exposure and the prevalence of hypertension (Leon Bluhm et al., 2007; Belojević et al., 2008; Barregard et al., 2009; Bodin et al., 2009; Dratva et al., 2011). However, cross-sectional research is known to have some limitations. One of them is the low number of participants in the studies, which makes it difficult to show the effects of noise on population health (Stansfeld et al., 2011). Additionally, many of the studies use subjective measures, such as self-reported noise levels, which are not always aligned with objective noise measurements for cardiovascular diseases (Schmit and Lorant, 2009; Mosca et al., 2013). These reasons, together with other research biases, might be the reason why there are differences in the results of the studies of urban noise on people's health (Sørensen et al., 2011; Van Kempen et al., 2018).

The impact of noise on health has led to the development of laws and regulations to control and reduce its presence. In 1996, the European Union (EU) published the *Green Paper* (European Parliament, 1996), a document containing policy proposals on how to mitigate the unwanted effects on noise in European cities. This document was the basis for the *Environmental Noise Directive* (END), which was adopted in 2002 as the general regulation for environmental noise management in Europe (European Parliament, 2002). The END defines a general framework to produce noise maps by adopting two noise indicators: *night-time noise level* (L_{night}), which is the average sound pressure level during the night hours within the year, and *day-evening-night noise level* (L_{den}), which represents the average overall sound pressure level within the year. The current directive considers four main sources: industrial, aircraft, railway, and road traffic noise. The effect of these sources on population has been previously studied; however, there are a multitude of alternative sounds related to recreational activities that can be potentially linked to positive (Aletta et al., 2018) or negative (Asensio et al., 2018; Ottoz et al., 2018) effects on citizens well-being. The *World Health Organization* has recently published for the first time a document that provides guidelines to reduce the effect of leisure noise on citizens (World Health Organization, 2018).

Traditionally, research related to noise attitudes in cities was conducted through face-to-face surveys, in which citizens were asked about the presence of specific noise sources. Technological progress has made it possible to develop tools that facilitate this task, whether through online surveys (Silva et al., 2017) or large-scale crowdsourcing systems that allow noise levels to be measured and include questionnaires to characterize urban sounds (Radicchi, 2017). Advances in sound pattern recognition through deep neural networks and their incorporation into low-cost instrumentation have also enabled the detection of urban sounds through urban noise monitoring networks (Bello et al., 2019; Mydlarz et al., 2019). Additionally, with the widespread adoption of the Internet and social media, digital data has became a valuable source to characterize cities and to quantify their environmental dimension. By using geo-located picture tags from social media, Aiello et al. proposed a new methodology to capture the sensory layers of cities. By using social media data from 12 major cities around the world, they showed that it is possible to characterize at scale the *smellscape* (Quercia et al., 2015) and the *soundscape* (Aiello et al., 2016) of cities—namely the distribution of the typical categories of smells and sounds that a person would be likely to perceive in a given area. By capturing both pleasant and unpleasant perceptions, their approach expanded the negative perspective on sound and smell that was at the time predominant in urban planning, especially by contributing with knowledge that could lead to new approaches in the domain of noise and sound monitoring.

More recently, researchers have resorted to social media to detect and monitor urban phenomena. Gasco et al. proposed and developed a methodology to detect and classify noise complaints from Twitter by analyzing features from text, and they were able to measure the impact of events in cities in terms of noise perception (Gasco et al., 2017, 2019). Lorente et al. used location data from Online Social Networks to analyze how masses of people move during large events in cities (Lorente-Riverola and Ruiz-Sánchez, 2018).

Overall, research in this area has shown that one can use social media to model effectively the pulse of the urban life at a scale and granularity that would be hard to achieve with traditional methods. The contribution of this study is to assess the value of social media data in predicting hypertension rates in addition to traditional data sources of noise exposure and socio-economic factors. In contrast to cross-sectional studies, mostly based on surveys and small scale samples, we analyzed the whole territory of Greater London at the level of about one thousand census areas. We find that official noise exposure levels add a relatively limited contribution in predicting hypertension. On the contrary, the soundscape extracted from social media, understood as the presence of sound sources extracted from a visual platform such as Flickr, considerably improves the model performance. Our findings suggest that sound maps that incorporate social media information can better inform design policies than just considering maps of noise levels.

The rest of the manuscript is organized as follows. In Methods, we describe both the data sources we use in our analysis and the methodology to compute the hypertension rates, the noise exposure levels, the soundscape of an area and the socioeconomic confounding factors we exploit in a multivariate linear regression analysis. In Results, we present different models that combine noise exposure and area sound profiles estimated from social media data to predict hypertension at area level. In Discussion and Conclusions, we cover the impact and the limitations of our approach and lay out future directions of our work.

METHODS

In the following, we describe: (1) the data sources used in the study; the methodology to compute, (2) the hypertension prevalence rates, (3) the noise exposure, and (4) the soundscape of all the MSOAs in the Greater London region. Then, we present the multivariate linear regression model we use to discover the factors that are more strongly associated with hypertension, our target variable.

Data

Next, we describe the data sources we gathered from open data platforms and social media to model noise and hypertension prevalence in the Greater London area. The different data sources do not always overlap in terms of their temporal span. However, we do not expect the data distributions to change significantly over time. The social media data we use spans several years, which allowed us to estimate an average soundscape profile that discounts seasonality and special events. Hypertension rates and noise exposure data are stable overtime and highly correlated from year to year. The spatial unit of our study is the Middle layer Super Output Areas (MSOA), that are 983 geographical areas for use in tabulating census and other statistical data in UK, with an average population of 8,346 inhabitants.

Drugs Prescriptions

Our primary source for information on health outcomes is the National Health Service[1] (NHS), a collection of public healthcare providers and infrastructures that handle health care in UK. In this study, we focus on NHS England—one of the four agencies leading the healthcare system in each constituent country of the UK. To model drugs consumption, we refer to the general practice prescribing data[2] that contains all medicines, dressings, and appliances that are prescribed and dispensed each month by the set of general practices (GPs) in England. For each practice, we keep track of the total number of items prescribed and dispensed aggregated by BNF codes. The British National Formulary (BNF) is a pharmaceutical reference book that contains a wide spectrum of information and advice on prescribing directives and pharmacology; it provides a taxonomy in which all medicines are organized in classes according the disease that they are intending to treat (EBM, 2018). To characterize the prevalence of hypertension, we considered the prescriptions of the full year 2014 and focused on the drug category *2.5* of the BNF taxonomy ("Hypertension and Heart Failure"). To compute drug consumption rates across spatial units, we refer to open statistics on the patients registered at a GP[3]. The dataset provides information on the geographical provenance, i.e., where patients come from aggregated by MSOA, along with gender and age distributions.

Noise

The Department for Environment, Food, and Rural Affairs[4] (DEFRA) publishes strategic noise maps for urban areas with more than 100,000 inhabitants following the criteria specified in the Environmental Noise Directive (European Parliament, 2002). Strategic noise maps are calculated through by simulating how a noise source produced in different points of the city propagates in the surroundings. The simulation produces different measures of noise estimates for each cell of a mesh covering the full urban area. We considered the strategic noise maps available in the London area published in 2012, corresponding to road and rail sources[5]. We used the maps that represent the noise levels using the recommended descriptors of the European Union and defined in ISO 1996-2 (International Organization for Standardization, 2007). Specifically, we gathered the day-evening-night noise level (L_{den}), that quantify the equivalent noise level over the whole day, with a penalty of 5 dBA for evening noise and of 10 dBA for nighttime noise; and the night-time noise level (L_{night}), that represents the noise level during the night period (usually between 22.00 and 07.00 h).

Socio-Demographic Statistics

The Office for National Statistics[6] (ONS) is responsible for the census in England and Wales, and it is the provider of open data at several geographical aggregation levels on socio-economic, cultural and demographic variables as measured by the Census of Population, whose last update was performed in 2011[7]. In our analysis, we control for three sociodemographic confounding factors that have been linked to cardiovascular diseases and hypertension by previous literature: *age* (Pinto, 2007), *income* (Kaplan et al., 2010; Keenan et al., 2011), and *gender* (Hayes and Taler, 1998). We adopt the age organization in classes from the census, grouping the population in three buckets of 0–44, 45–64, and more than 65 years old, respectively. All the statistics are spatially aggregated at the level of MSOAs.

Social Media

To characterize the soundscape of London, we used to the Flickr[8] dataset published by the Chatty Maps project (Aiello et al., 2016). The dataset includes a random sample of 17M geo-referenced Flickr photos taken within the boundary of Greater London and uploaded between 2010 and 2015. Each photo in this sample is geo-referenced with the latitude and longitude of the place they have been taken, and comes with free-text tags added by the Flickr user who uploaded it. Users are denoted by an anonymized identifier.

Hypertension Rates

To characterize the incidence of hypertension we define the prescription rate $r_{hypertension}(m)$ as the number of items per patient prescribed in a timeframe of reference in each MSOA m.

[1] https://www.nhs.uk/

[2] https://digital.nhs.uk/data-and-information/publications/statistical/practice-level-prescribing-data

[3] https://digital.nhs.uk/data-and-information/publications/statistical/patients-registered-at-a-gp-practice

[4] https://www.gov.uk/government/organisations/department-for-environment-food-rural-affairs

[5] https://data.london.gov.uk/dataset/noise-pollution-in-london

[6] https://www.ons.gov.uk/

[7] https://www.ons.gov.uk/census/2011census/2011censusdata

[8] Flickr is a photo and video hosting service: https://www.flickr.com/

We assume that an area with a higher prescription rate for drugs curing hypertensions is a relevant proxy for the prevalence of that condition. The rate in an area m is defined as:

$$r_{hypertension}(m) = \frac{i_{hypertension}(m)}{p(m)}$$

where $i_{hypertension}(m)$ is the total number of items prescribed to residents of the MSOA m (regardless of which practices are prescribing those items), and $p(m)$ is the total number of patients living in MSOA m. We compute $i_{hypertension}(m)$ as:

$$i_{hypertension}(m) = \sum_{g \in g(m)} i_{hypertension}(g, m)$$

where $i_{hypertension}(g, m)$ is the number of items prescribed by the practice g to someone living in the MSOA m. Unfortunately, we can't directly measure the quantity $i_{hypertension}(g, m)$; however, we hypothesize that the number of items prescribed in a GP is uniformly distributed according to the patients geographical provenance. Therefore, we define:

$$i_{hypertension}(g, m) = r_{hypertension}(g) \cdot p(g, m)$$

where $p(g, m)$ represents the number of patients registered at the practice g and living in the MSOA m as derived from the patients provenance dataset. To compute the rate of drugs curing hypertension in a practice $r_{hypertension}(g)$ we use the relation:

$$r_{hypertension}(g) = \frac{i_{hypertension}(g)}{p(g)}$$

Let $p(g)$ be the total number of patients registered at the practice g:

$$p(g) = \sum_{m \in msoa(g)} p(g, m)$$

we are able to derive an estimate of the prescription rate at the level of a spatial unit m. **Figure 1** shows the spatial distribution and the probability distribution of the quantity $r_{hypertension}(m)$.

Noise Exposure

Data on the population's exposure to noise are usually presented in an aggregated format together with the noise maps. Nevertheless, we need that information for each area of London, hence we define a method for estimating it.

This method uses the geo-spatial layers of the London noise map, as well as the MSOAs boundaries and the residential buildings present in the city, both available in the London Datastore[9]. Considering those data sources, the methodology comprises three steps:

- Calculate the exposed area of residential buildings to different noise levels through a spatial intersection between the noise map and the residential terrain within each MSOA.
- From the total residential area in each MSOA and the areas exposed to each noise level, calculate the percentage of dwellings exposed to the different noise levels in each MSOA
- Assuming that the population of each MSOA lives equally distributed within residential areas, calculate the percentage of people exposed to noise levels. Although the population exposed to noise is usually given in numerical terms, we calculated it in percentage terms because it allows us to compare the impact of noise between areas regardless of their population.

Based on the WHO guidelines on the potential effects of noise on the cardiovascular system (Berglund et al., 1999; Hurtley, 2009), we considered the percentage of people exposed to a day-evening-night road track noise level over 55 dB (*RD.Lden.over55*) and the percentage of people exposed to a day-evening-night railway noise level over 55 dB (*RL.Lden.over55*). **Figure 2** shows the percentage of the population exposed to more than 55 dB for road and railway noise.

Noise exposure calculations refer to outdoor spaces. However, citizens spend part of their time in their homes, which, depending on the dwellings' quality of construction, will provide a better insulation from outside noise and therefore a possible decrease in hypertension. To account for this factor into our models, we calculated the buildings' *Energy Efficiency* index $EE_{building}$. This index is calculated using the domestic energy performance certificates provided by the UK Ministry of Housing, Communities and Local Government through a public API[10]. These certificates have a numerical index between 0 and 100 that indicates the energy efficiency of the property considering the type of window installed and the quality of construction of the façade. We gathered all the available certificates in Greater London, and we computed the average domestic energy performance index per MSOA with the certificates in each area. The spatial distribution of this index is shown in **Figure 3**.

Sound Profile

Strategic noise maps in cities capture noise from road and trains sources. Since the incidence of health conditions due to noise exposure could be potentially traced back to a wider spectrum of sound sources, we refer to social media to characterize the soundscape London areas. To estimate the presence of different types of sound sources from social media data, we use the approach proposed by Aiello et al. (2016). They first compiled a list of words that represent *sound sources* taken from Murray Schafer's seminal book "The Soundscape" (Schafer, 1993), an influential work that defined the concept of urban soundscape. Based on the co-occurrences of these words in picture tags from social media, they were able to arrange them in a taxonomy of urban sounds in which similar sounds are grouped together. This taxonomy has six top-level categories which match those discussed by Schafer: *transport* (e.g., sounds generated by cars, trains, and airplanes), *mechanical* (e.g., drills or other heavy mechanical devices), *human* (e.g., chatting or footsteps), *music*

[9]https://data.london.gov.uk/dataset/openstreetmap

[10]https://epc.opendatacommunities.org

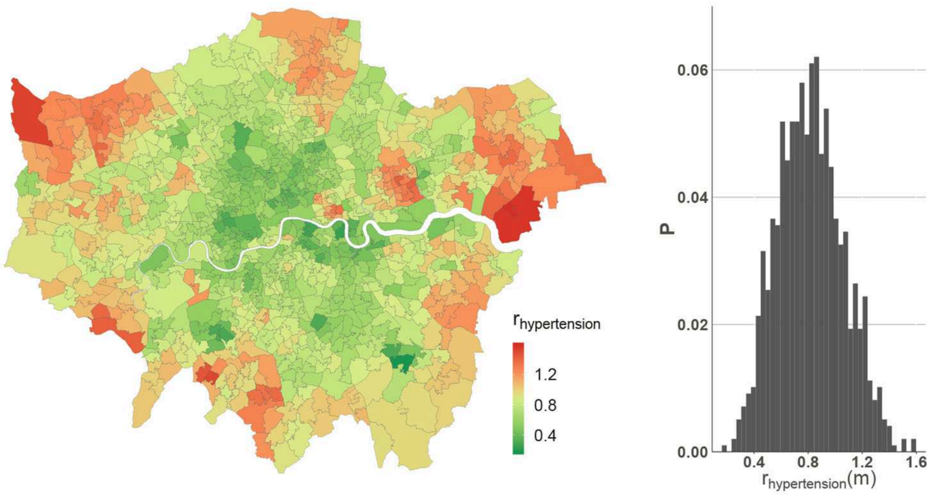


FIGURE 1 | Hypertension rates in London MSOAs. **(Left)** Spatial distribution; **(Right)** Probability distribution.

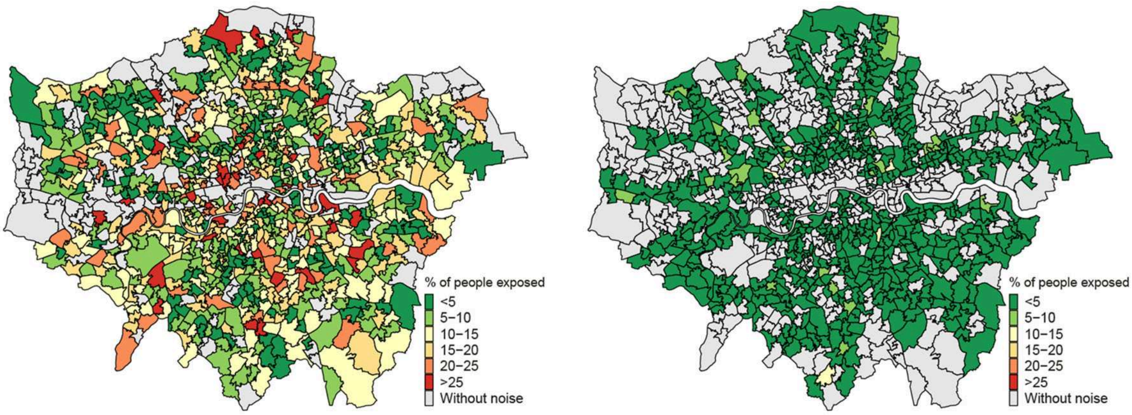


FIGURE 2 | Spatial distribution of percentage of people exposed to noise.

(e.g., street bands), *nature* (e.g., water, foliage, animals), and *indoor* (e.g., shower, office paper, or sounds typically generated inside buildings).

Specifically, sound words share the same taxonomic category if they have a high semantic similarity, estimated from the frequency of these sound words in Flickr pictures. Sometimes, this data-driven taxonomy groups sets of sounds by their context, rather by their source type. For example, the mechanical sounds from a typewriter or a printer are semantically more similar to "indoor" sounds (e.g., leafing through a paper document or flushing a toilet) than to other "mechanical" sounds that are mainly found in other contexts (e.g., the pounding sound of a jackhammer). Similarly, showering, playing an instrument or driving a car are all "human" activities, but none of them are categorized under the "human" class, as they are respectively "indoor," "music," and "transport"; instead, "human" sounds are mostly those that the human body can produce unaided (e.g., footsteps, talking, laughing). Naturally, alternative sound taxonomies are conceivable, but we decided to rely on Aiello et al.'s because it is theoretically-grounded and validated.

This taxonomy can then be applied to geo-referenced social media data to estimate the typical sounds of an area; the underlying idea is that if many pictures taken within an area are tagged with words belonging to a given sound category, that area will likely be characterized by that sound. In their experimental validation using Flickr picture tags in London, Aiello et al. provided evidence to support the validity of this estimation; for example, they showed that the vast majority of retrieved pictures do actually represent sound sources, and that the sound profiles they compute in bounding boxes around streets in London correlate with noise levels in expected ways. In recent years, this data-driven taxonomy has become a common reference for studies on noise and urban sounds (Kang et al., 2016; Zuo et al., 2016; Fairbrass et al., 2017).

This approach has a few working assumptions. First, the social media data considered should be *geo-salient*, meaning that it

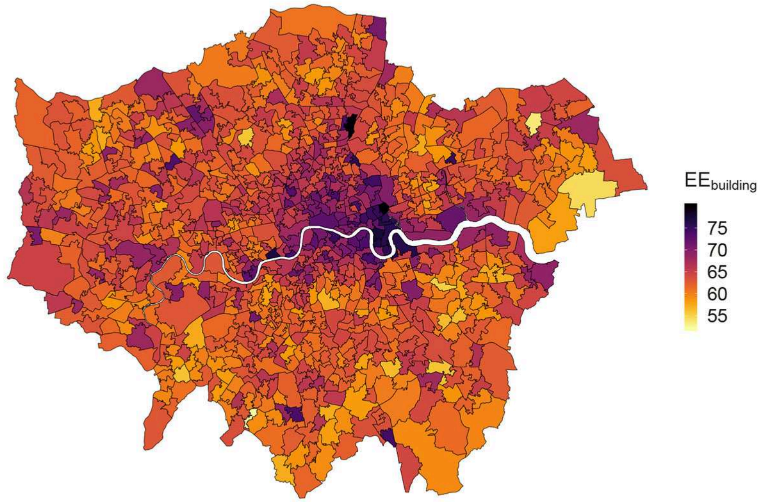


FIGURE 3 | Map of the average energy efficiency index in London MSOAs.

should be relevant to the geographical area corresponding to its geo-location. Some social media sources are by nature more geo-salient than others. For example, Twitter is not highly geo-salient because the content of a geo-referenced tweet might often unrelated with the location of the poster (e.g., people tweeting from their homes about a public demonstration happening in a different city). On the contrary, photo-sharing platforms tend to be geo-salient because the tags attached to the pictures are usually describing the picture itself, which is literally a depiction of the space around the geographic coordinates attached to the photo. This is why, like in the original approach (Aiello et al., 2016), we chose Flickr as a data source. Second, this method works effectively only on aggregate for areas that contain an abundant volume of data—as any approach based on collective intelligence applied to the urban context (Chatzigiannakis et al., 2011). This why, as we will detail next, we focused only on areas with large numbers of geo-referenced pictures.

We follow this methodology by first associating each Flickr photo to the MSOA whose boundaries contain its geographical coordinates. For each sound category c and MSOA m, we model the prevalence that sound category in the area as:

$$f(c, m) = \frac{\#\, pictures\ with\ sound\ tags\ in\ category\ c\ @\ m}{\#\, pictures\ with\ sound\ tags\ @\ m}$$

where the numerator represents the number of pictures that contain at least one tag from sound category c in the area m and the denominator counts the number of pictures that refer to any sound experience. The result of this step is creation of a 6-dimensional vector for each MSOA that models the prevalence of each sound category in an area. **Figure 4** shows the predominant sound type in a MSOA and their distribution. Note that natural and transport sounds are more predominant in the periphery and the inner city is characterized often by sounds related to humans and music. White areas are spatial units with a low coverage (<100 pictures related to sound categories) that consequently are filtered out from our analysis.

London is a global financial hub characterized by high population density and a tremendous tourists flow that is concentrated in specific areas of the city. A photo sharing platform as Flickr reflects this unequal spatial distribution of activity that results in a high heterogeneity between central and peripheral MSOAs. To take into account this effect, we estimate the social media platform penetration rate as:

$$pr_{Flickr}(m) = \frac{\#\, Flickr\ users\ @\ m}{\#\, residents\ @\ m}$$

where $pr_{Flickr}(m)$ is the ratio between the number of unique Flickr users who posted at least one photo in MSOA m and its population from the census. High photo density is an indicator of a place interestingness, for example because of its scenicness or historic value (Serdyukov et al., 2009). In this direction, a high penetration rate could also be linked to non-acoustic factors that affect noise annoyance (Asensio et al., 2017) and that may have an effect on health outcomes.

Analysis of Correlates

We use multivariate linear regression to determine to what extent the soundscape estimated from social media is related to health outcomes and how it may improve traditional models; in particular, we focus on hypertension that has been connected in the literature as a disease aggravated by noise. In **Table 1**, we summarize the list of dependent variables used in the study.

Previous work consistently showed how economic status might affect the prevalence of several types of diseases of the circulatory system (Kaplan et al., 2010; Keenan et al., 2011). To account for this confounding factor and to perform our analysis across homogeneous samples, we group London MSOAs in three economic classes following the approach implemented in several studies that looked at the relationship between pollutants

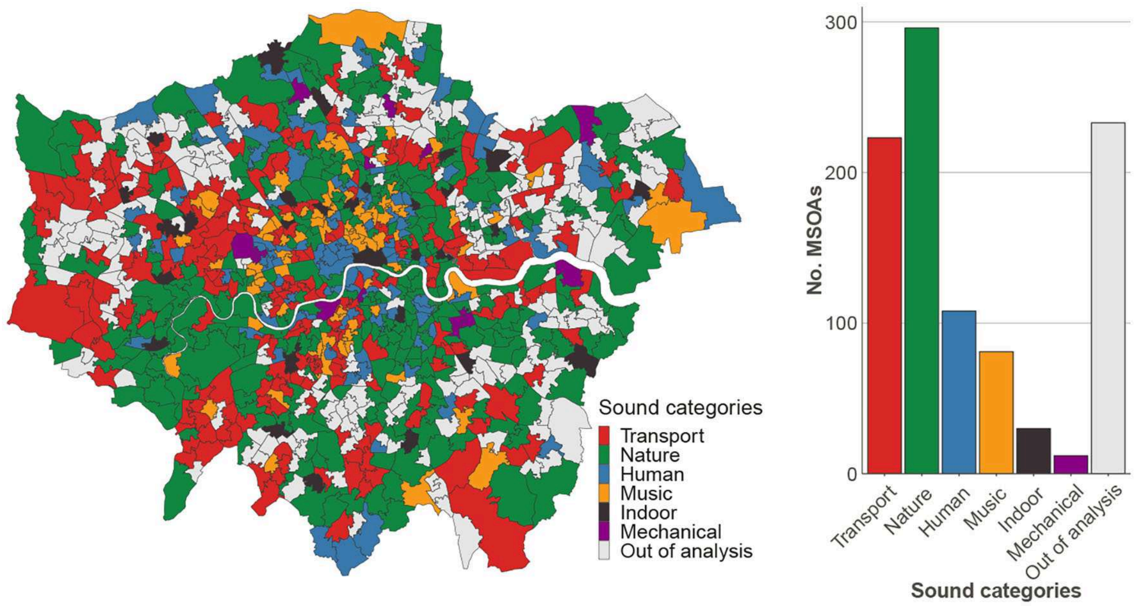


FIGURE 4 | (Left) Map of the prominent sound category in each MSOA. **(Right)** Frequency distribution of dominant sounds across areas. In gray, areas with low social media coverage.

TABLE 1 | Summary of the variables used in the multivariate linear regression model to characterize the prevalence of hypertension in London MSOAs.

Group	Variable name	Transformation	Mean(SD)	Explanation
Noise exposure	$RD.Lden.over55(m)$	Square root	0.362 (0.164)	Percentage of people exposed to a day-evening-night road traffic noise level over 55dB in MSOA *(m)*.
	$RL.Lden.over55(m)$		0.197 (0.225)	Percentage of people exposed to a day-evening-night railway noise level over 55dB in MSOA *(m)*.
	$EE_{building}(m)$	None	0.394 (0.148)	Average value of domestic energy efficiency certificates available in MSOA *(m)*.
Social Media	$f(transport, m)$	Square root	0.467 (0.218)	Fractions of pictures that are part of the category *transport* in MSOA *m*
	$f(nature, m)$		0.499 (0.225)	Fractions of pictures that are part of the category *nature* in MSOA *m*
	$f(human, m)$		0.443 (0.169)	Fractions of pictures that are part of the category *human* in MSOA *m*
	$f(music, m)$		0.300 (0.206)	Fractions of pictures that are part of the category *music* in MSOA *m*
	$f(building, m)$		0.306 (0.165)	Fractions of pictures that are part of the category *building* in MSOA *m*
	$f(mechanical, m)$		0.242 (0.161)	Fractions of pictures that are part of the category *mechanical* in MSOA *m*
	$pr_{Flickr}(m)$		0.0650 (0.0767)	Ratio between the number of unique Flickr users and the population living in MSOA *m*
Sociodemographics	$age_{0-44}(m)$	None	0.540 (0.196)	Percentage of people aged 0–44 in MSOA *m*
	$age_{45-64}(m)$		0.495 (0.185)	Percentage of people aged 45–64 in MSOA *m*
	$age_{over65}(m)$		0.344 (0.168)	Percentage of people aged more than 64 in MSOA *m*
	$income(m)$		0.311 (0.188)	Average household income in MSOA *m*
	$p_{female}(m)$	Squared	0.634 (0.135)	Percentage of women in MSOA *m*

and health outcomes (Richardson et al., 2013; Deguen et al., 2015; Fecht et al., 2015). To define the economic boundaries of these three classes we used the updated values defined at the Great British Class Survey (Savage et al., 2013). This study originally identified seven economic classes which we regrouped into three to comply with the methodological specifications of similar studies: a *High class* containing MSOAs with a yearly average household income greater than 68k pounds, a *Middle class* with average income between 33k and 68k, and a *Low class* with income <33k. **Figure 5** (left) shows the spatial distribution of the MSOAs color-coded according to their economic class. The distribution of different sound categories for each class is presented in **Figure 5** (right). Lower income areas are characterized by a predominant portion of transport-related sounds, whereas human, natural, and music sounds are more frequent in areas with higher economic status.

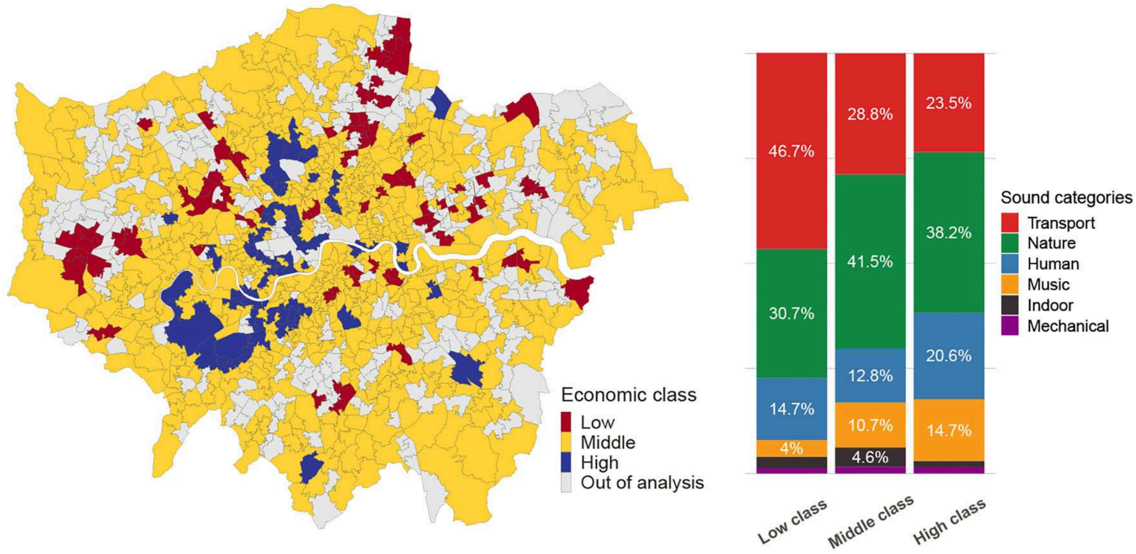


FIGURE 5 | (Left) Spatial distribution of the socioeconomic classes in London MSOAs. **(Right)** Distribution of the predominant sound categories for each social class.

We build a model for each economic class and we applied different transformation to our variables (**Table 1**) to make their distributions normal. We then apply a min-max normalization to all the features to ease the interpretation of the regression coefficients.

We perform our analysis in three steps. First, we calculate the control models including only the sociodemographic variables that we use as a baseline to measure the explanatory effect of the rest of the variables. Then, we recalculate the models incorporating the noise exposure variables to verify whether they are significant and their relevance in the model. Finally, we incorporate the social media variables to check if they allow us to better predict the outcome variable.

RESULTS

Next, we present the results of the regression tasks grouped by economic class.

Low-Income Areas

The 72 MSOAs belonging to the lowest economic income group are mainly located in the North and West part of London (**Figure 5**, left). Age, gender, and income are important explanatory variables. This is true in the low class model (**Table 2**) as well as for all the other models. In line with previous literature, areas with higher population aged 65 and over (Anderson et al., 1999; Buford, 2016) and with more males (Hayes and Taler, 1998) show higher hypertension prevalence. Income correlates with better healthcare and healthier habits, that have been shown to have a significant impact on cardiovascular diseases (Kaplan et al., 2010; Keenan et al., 2011; Aiello et al., 2019). When adding the noise exposure and energy efficiency variables, we observe a 9% increase in the adjusted R^2 (from 0.268 to 0.291). The only other significant variable is the energy efficiency $EE_{building}$. The higher construction quality of dwellings leads to an improved sound insulation technology and general construction standards that could explain the lower hypertension incidence rate. The noise variables are not significant and the absolute value of their coefficients is small, compared to others in the same model.

When adding soundscape variables from social media, the model achieves the best performance in terms of adjusted R^2 (0.363), with an increase of 35% over the socio-demographic baseline. The presence of natural sounds is negatively associated with hypertension prevalence, which is in line with the research hypotheses in the field of soundscapes (Aletta et al., 2018). This is also true for sounds typical of human and indoor activity. According the sound taxonomy from Aiello et al., indoor sounds correspond to activities typical of familiar contexts in either home or office—usually soft background sounds. Indoor sounds are not necessarily "relaxing" as those produced by natural elements; yet, if indoor sounds are predominant in an area, it might be an indicator that the area is denoted by a rather quiet sound ambiance that is not plagued by sounds that are more harmful to the human body.

To shed light on the relative importance of the regressors in the linear models, we apply the *lmg* method (Lindeman et al., 1980). This method provides the relative contribution of each predictor to the R^2 in a multivariate linear regression model. **Table 3** shows the results for the low-income areas. Socio-demographic variables explain the highest portion of variance across models. Among the social media variables, Flickr penetration, nature, and indoor the three most important ones.

Middle-Income Areas

The middle class represents the majority of London's MSOAs, for a total of 547 areas. **Table 4** summarizes the output

TABLE 2 | Low-income areas results.

	Dependent variable		
	Hypertension.per.capita		
	(1)	(2)	(3)
age_{over65}	0.674*** (0.156)	0.576*** (0.160)	0.631*** (0.160)
income	−0.357*** (0.090)	−0.375*** (0.089)	−0.377*** (0.095)
p_{female}	−0.496*** (0.138)	−0.417*** (0.140)	−0.489*** (0.143)
pr_{Flickr}			−0.294 (0.179)
f(*nature*)			−0.281** (0.108)
f(*human*)			−0.195* (0.115)
f(*music*)			−0.210 (0.144)
f(*mechanical*)			0.033 (0.105)
f(*transport*)			−0.207 (0.130)
f(*indoor*)			−0.283** (0.125)
*RD.Lden.over*55		−0.024 (0.099)	−0.034 (0.100)
*RL.Lden.over*55		0.025 (0.076)	0.023 (0.074)
$EE_{building}$		−0.266** (0.120)	−0.138 (0.123)
Observations	75	75	75
R^2	0.298	0.348	0.475
Adjusted R^2	0.268	0.291	0.363
Residual Std. Error	0.182 (df = 71)	0.180 (df = 68)	0.170 (df = 61)
F Statistic	10.040*** (df = 3; 71)	6.052*** (df = 6; 68)	4.242*** (df = 13; 61)

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

TABLE 4 | Middle income areas results.

	Dependent variable		
	Hypertension.per.capita		
	(1)	(2)	(3)
age_{over65}	0.555*** (0.034)	0.493*** (0.037)	0.424*** (0.038)
income	−0.047*** (0.005)	−0.048*** (0.005)	−0.028*** (0.006)
p_{female}	−0.023*** (0.006)	−0.023*** (0.006)	−0.026*** (0.006)
pr_{Flickr}			−0.042*** (0.008)
f(*nature*)			−0.001 (0.007)
f(*human*)			−0.011* (0.006)
f(*music*)			0.001 (0.007)
f(*mechanical*)			0.001 (0.006)
f(*transport*)			−0.005 (0.007)
f(*indoor*)			0.004 (0.005)
*RD.Lden.over*55		−0.012** (0.005)	−0.004 (0.005)
*RL.Lden.over*55		0.001 (0.006)	0.005 (0.006)
$EE_{building}$		−0.024*** (0.006)	−0.011* (0.006)
Observations	752	752	752
R^2	0.281	0.305	0.342
Adjusted R^2	0.278	0.299	0.331
Residual Std. Error	0.144 (df = 748)	0.142 (df = 745)	0.139 (df = 738)
F Statistic	97.618*** (df = 3; 748)	54.501*** (df = 6; 745)	29.529*** (df = 13; 738)

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

TABLE 3 | Relative variable importance for the Low-class models.

Group	Variables	Model 1 (%)	Model 2 (%)	Model 3 (%)
Sociodemographics	age_{over65}	42.21	29.24	21.52
	income	36.30	34.50	21.83
	p_{female}	21.49	13.91	10.85
Noise exposure	*RD.Lden.over*55		0.23	0.39
	*RL.Lden.over*55		0.47	0.38
	$EE_{building}$		21.65	11.15
Social media	pr_{Flickr}			13.34
	f(*nature*)			6.44
	f(*human*)			3.93
	f(*music*)			1.42
	f(*mechanical*)			0.83
	f(*transport*)			1.52
	f(*indoor*)			6.40

of the regression task across models. Similar to the low-income class, socioeconomic covariates of age, gender, and income are significant; the baseline model reaches an adjusted R^2 of 0.278. The addition of the noise exposure variables increases the R^2 to 0.305. The energy efficiency and the road noise are two significant predictors. Unexpectedly, the road noise has a negative coefficient, yet with a low absolute value.

The social media model increases the R^2 by 22% compared to the socio-demographic baseline. The Flickr penetration is the strongest significant variable. Like in the low class, human-related sounds are associated with arras characterized by lower hypertension levels.

The analysis of variable importance (**Table 5**) confirms the central role of socioeconomic regressors. Flickr penetration constitutes also a strong signal in the social media model.

HIGH-INCOME AREAS

In the 67 high-income London MSOAs, the addition of the noise variables on top of socio-economic factors slightly decreases the performance of the model and yields no new significant regressors (**Table 6**). The model that includes both noise exposure and social media variables increases the R^2 by 74% compared to the socio-demographic baseline, with significant coefficients for *indoor* sounds and, unlike previous models, for *mechanical* sounds too. The presence of mechanical sounds, e.g., industrial and work-related sounds emitted by tools and machinery performing tasks like hammering or drilling, is positively associated with higher prevalence of hypertension.

TABLE 5 | Relative variable importance for the Middle-class models.

Group	Variables	Model 1 (%)	Model 2 (%)	Model 3 (%)
Sociodemographics	age_{over65}	80.27	59.74	40.75
	income	15.23	15.13	6.71
	p_{female}	4.50	3.61	3.19
Noise exposure	*RD.Lden.over*55		2.84	1.43
	*RL.Lden.over*55		0.37	0.26
	$EE_{building}$		18.31	10.09
Social media	pr_{Flickr}			29.47
	f(*nature*)			0.45
	f(*human*)			2.04
	f(*music*)			3.48
	f(*mechanical*)			1.49
	f(*transport*)			0.16
	f(*indoor*)			0.48

TABLE 6 | High class models.

	Dependent variable		
	Hypertension.per.capita		
age_{over65}	−0.095 (0.169)	−0.088 (0.178)	0.051 (0.166)
age_{45-64}	0.561*** (0.183)	0.585*** (0.188)	0.590*** (0.168)
income	−0.181** (0.083)	−0.176** (0.085)	−0.103 (0.080)
p_{female}	−0.035 (0.113)	0.004 (0.143)	−0.252 (0.177)
pr_{Flickr}			−0.361* (0.209)
f(*nature*)			0.158 (0.145)
f(*human*)			0.089 (0.129)
f(*music*)			0.032 (0.126)
f(*mechanical*)			0.316*** (0.093)
f(*transport*)			0.163 (0.099)
f(*indoor*)			−0.209** (0.092)
*RD.Lden.over*55		−0.044 (0.104)	−0.159 (0.101)
*RL.Lden.over*55		0.029 (0.069)	−0.009 (0.064)
$EE_{building}$		0.134 (0.119)	0.108 (0.107)
Observations	68	68	68
R^2	0.272	0.289	0.521
Adjusted R^2	0.226	0.207	0.394
Residual Std. Error	0.159 (df = 63)	0.161 (df = 60)	0.141 (df = 53)
F Statistic	5.885*** (df = 4; 63)	3.491*** (df = 7; 60)	4.116*** (df = 14; 53)

$^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

TABLE 7 | Relative variable importance for the High-class models.

Group	Variables	Model 1 (%)	Model 2 (%)	Model 3 (%)
Sociodemographics	age_{over65}	20.13	19.11	11.21
	age_{45-64}	58.34	56.30	32.70
	income	17.92	16.55	5.67
	p_{female}	3.61	3.64	3.15
Noise exposure	*RD.Lden.over*55		0.77	2.24
	*RL.Lden.over*55		0.52	0.27
	$EE_{building}$		3.11	1.63
Social media	pr_{Flickr}			5.55
	f(*nature*)			2.24
	f(*human*)			0.49
	f(*music*)			3.61
	f(*mechanical*)			17.52
	f(*transport*)			3.89
	f(*indoor*)			9.83

The feature importance analysis shows that mechanical sounds are among the most important predictors of the outcome (**Table 7**).

There might be several possible reasons why the coefficients associated to noise levels are either not significant or even slightly negative for the middle-income class. One of the reasons might be the geographic granularity of the study. Relatively large areas such as MSOAs can be very diverse in terms of their noise exposure, land use, and socio-demographic characteristics. By considering MSOAs as homogeneous entities, our model misses out on important signals to relate high noise levels to hypertension prevalence. To gauge this intuition, we experimented by restricting our analysis to the noisiest areas, in which noise is likely to be perceived by people living in all parts of those areas. In particular, we focused on the 147 MSOAS above the 85th percentile of the RD.Lden.over55 distribution. On these areas only, we observe a positive Kendall rank correlation coefficient $\rho = 0.22$ ($p < 0.001$) between noise exposure and hypertension prevalence, as one would expect. The fact that the traditional noise exposure measures are not able to fully capture the relation with the health outcome does not go against the main goal of this paper, which is to show the benefits of adding the information of social media-data to study the relationship between noise and health.

DISCUSSION

Our results suggest that socioeconomic factors are consistently a primary source of information when studying health outcomes at population level. Although it is widely known that economic status affects the prevalence of some diseases, our study sheds light on the limits of the traditional noise exposure models in capturing the effects of noise on citizens well-being.

Exposure to railway noise is not significantly associated with hypertension in any of the models, despite this connection has been suggested extensively by previous studies (Sørensen et al., 2011). It has to be noted that some previous studies found that excluding from the sample participants exposed to the highest noise levels increased the association between exposure and hypertension (Lee et al., 2019). Also, areas that are potentially exposed to high levels of railway noise are often protected by the installation of noise barriers that reduce the impact on the population living closely to the rail tracks. On the contrary, we showed that the ability of estimate the presence of heterogeneous sound sources (e.g., natural sounds) using social media increases our ability of identifying sound elements that are significantly associated to health outcomes.

These findings could support the work of several stakeholders. Those include urban planners, who could save the cost of deploying large noise monitoring networks (Mydlarz et al., 2019) by using social media platforms to measure the presence of these sound sources, and medical researchers, who could complement their studies about health effects of noise exposure with social media sound maps, which have never been used before in the medical context.

Limitations

Our approach comes with a two main limitations.

Representativeness. Studies analyzing the effects of noise on population health are usually based on small cohorts of selected individuals exposed to specific acoustic conditions. In this work, we study the interaction between sound and health at an unprecedented scale, but at the cost of using a relatively coarse spatial aggregation (MSOAs). Large areas can be very diverse in terms of their noise exposure, land use, and socio-demographic characteristics. The predictive and descriptive power of our models is limited by not considering such heterogeneity. The representation of sound sources that we obtained from social media is also affected by a number of biases, including the uneven representation of location types and the mix between pictures taken by tourists and those taken by locals. Also, we average out the contribution from pictures taken across several years and both during day and night. This approach smooths out seasonal patterns and one-off events, thus yielding an average representation of an area's soundscape. This average representation does not capture the high dinamicity of the urban soundscape. Some previous studies have attempted to carry out longitudinal studies of sensory data from social media (Quercia et al., 2016), but to do that systematically, one would need to overcome several challenges including the data sparsity that slicing the data would entail, and the known inaccuracy of timestamps coming from photocameras (Thomee et al., 2014).

Causality. Our study is observational and its results do not necessarily speak to causality. The health variable we consider as outcome is concurrently influenced by a number of factors other than sound (such as nutrition and physical exercise) that are hard to control for because of the unavailability of data at area-level. Similarly, people's perception of the urban soundscape is mediated by several factors—such as quality of the facade insulation—that one cannot capture through publicly available data and that are therefore hard to control for. Last, it is challenging to disentangle the role of certain sound categories from other sensory factors that co-occur with those categories. For example, disentangling the contribution of the visual perception of greenery from the presence of nature-related sounds in explaining health outcomes is an arduous task when relying on purely observational studies.

CONCLUSIONS

We proposed a methodology for studying how urban sound is associated to health outcomes. Instead of conducting a survey-based cohort study, we used open and social media data to conduct an observational study to analyze the hypertension prevalence across areas in London. By grouping city areas by economic class, our study suggests that the use of social media constitutes a practical way of augmenting noise data with information of the presence of different types of sound sources that are currently not considered in the European strategic noise maps and that. In London, these additional social media variables augment the power of noise models to predict hypertension at area level. Also, this approach allowed us to find an inverse association between presence of nature sounds and prevalence of hypertension, which closely relates to existing hypotheses formulated soundscape researchers.

In the future, this type of study could be extended to other noise-related diseases such as effects on stress level, through tranquilizer prescriptions, and sleep quality losses, through hypnotic prescriptions. Additionally, it would be convenient to carry out studies with a lower level of data aggregation, or to quantify the biases produced by working with MSOAs, as well as those produced by the estimates made in the noise exposure calculations.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study and they can be downloaded at the following repositories: https://doi.org/10.1098/rsos.150690; https://digital.nhs.uk; https://data.london.gov.uk/dataset/noisepollution-in-london.

AUTHOR CONTRIBUTIONS

RS, LA, and DQ conceived, designed, and supervised the project. LG, CA, and GA contributed to the design of the study, writing the protocol, data preparation, and analysis. LG drafted the manuscript. RS, LA, CA, DQ, and GA performed the quality assessment and revised the manuscript. All authors have read and approved the submitted version.

FUNDING

RS has been partially supported by the project Countering Online hate speech through Effective on-line Monitoring funded by Compagnia di San Paolo. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. LG has been supported by the funding of EIT Digital Doctoral School.

REFERENCES

Aiello, L. M., Schifanella, R., Quercia, D., and Aletta, F. (2016). Chatty maps: constructing sound maps of urban areas from social media data. *R. Soc. Open Sci*. 3:150690. doi: 10.1098/rsos.150690

Aiello, L. M., Schifanella, R., Quercia, D., and Del Prete, L. (2019). Large-scale and high-resolution analysis of food purchases and health outcomes. *EPJ Data Sci.* 8:14. doi: 10.1140/epjds/s13688-019-0191-y

Aletta, F., Oberman, T., and Kang, J. (2018). Positive health-related effects of perceiving urban soundscapes: a systematic review. *Lancet* 392:S3. doi: 10.1016/S0140-6736(18)32044-0

Anderson, G. H. (1999). Effect of age on hypertension: analysis of over 4,800 referred hypertensive patients. *Saudi J. Kidney Dis. Transpl*. 10, 286–297.

Andersson, R., and Turner, L. M. (2014). Segregation, gentrification, and residualisation: from public housing to market-driven housing allocation in inner city Stockholm. *Int. J. Housing Policy* 14, 3–29. doi: 10.1080/14616718.2013.872949

Asensio, C., Gasco, L., and de Arcas, G. (2017). A review of non-acoustic measures to handle community response to noise around airports. *Curr. Pollut. Rep*. 3, 230–244. doi: 10.1007/s40726-017-0060-x

Asensio, C., Gasco, L., De Arcas, G., Løpez, J. M., and Alonso, J. (2018). Assessment of residents–exposure to leisure noise in Málaga (Spain). *Environments* 5. doi: 10.3390/environments5120134

Barregard, L., Bonde, E., and Öhrström, E. (2009). Risk of hypertension from exposure to road traffic noise in a population-based sample. *Occup. Environ. Med*. 66, 410–415. doi: 10.1136/oem.2008.042804

Barthelemy, M. (2016). A global take on congestion in urban areas. *Environ. Plann. B* 43, 800–804. doi: 10.1177/0265813516649955

Bello, J. P., Silva, C., Nov, O., DuBois, R. L., Arora, A., Salamon, J., et al. (2019). Sonyc: A system for monitoring, analyzing, and mitigating urban noise pollution. *Commun. ACM* 62, 68–77. doi: 10.1145/3224204

Belojević, G. A., Jakovljević, B. D., Stojanov, V. J., Slepcević, V. Z., and Paunović, K. Z. (2008). Nighttime road-traffic noise and arterial hypertension in an urban population. *Hypertens. Res*. 31:775. doi: 10.1291/hypres.31.775

Berglund, B., Lindvall, T., Schwela, D. H., World Health Organization, and Occupational and Environmental Health Team. H. (1999). *Guidelines for Community Noise*. Available online at: https://apps.who.int/iris/handle/10665/66217

Bodin, T., Albin, M., Ardö, J., Stroh, E., Östergren, P.-O., and Björk, J. (2009). Road traffic noise and hypertension: results from a cross-sectional public health survey in Southern Sweden. *Environ. Health* 8:38. doi: 10.1186/1476-069X-8-38

Buford, T. W. (2016). Hypertension and aging. *Ageing Res. Rev*. 26, 96–111. doi: 10.1016/j.arr.2016.01.007

Chatzigiannakis, I., Mylonas, G., and Vitaletti, A. (2011). Urban pervasive applications: challenges, scenarios and case studies. *Comput. Sci. Rev*. 5, 103–118. doi: 10.1016/j.cosrev.2010.09.003

Deguen, S., Petit, C., Delbarre, A., Kihal, W., Padilla, C., Benmarhnia, T., et al. (2015). Neighbourhood characteristics and long-term air pollution levels modify the association between the short-term nitrogen dioxide concentrations and all-cause mortality in Paris. *PLoS ONE* 10:e131463. doi: 10.1371/journal.pone.0131463

Dratva, J., Phuleria, H. C., Foraster, M., Gaspoz, J.-M., Keidel, D., Künzli, N., et al. (2011). Transportation noise and blood pressure in a population-based sample of adults. *Environ. Health Perspect*. 120, 50–55. doi: 10.1289/ehp.1103448

EBM, D. (2018). University of Oxford. Available online at: https://openprescribing.net/bnf/ (accessed November 26, 2018).

European Environment Agency (2014). *Noise in Europe 2014. Number 10/2014*. Publications Office of the European Union, Luxembourg.

European Parliament (1996). *Future Noise Policy. European Commission Green Paper*. Office for Official Publications of the European Communities, Luxembourg.

European Parliament (2002). *Directive 2002/49/EC of the European Parliament and the Council of 25 June 2002 Relating to the Assessment and Management of Environmental Noise*. Official Journal of the European Communities, Luxembourg.

Fairbrass, A. J., Rennert, P., Williams, C., Titheridge, H., and Jones, K. E. (2017). Biases of acoustic indices measuring biodiversity in urban areas. *Ecol. Indic*. 83, 169–177. doi: 10.1016/j.ecolind.2017.07.064

Fecht, D., Fischer, P., Fortunato, L., Hoek, G., de Hoogh, K., Marra, M., et al. (2015). Associations between air pollution and socioeconomic characteristics, ethnicity and age profile of neighbourhoods in England and the Netherlands. *Environ. Pollut*. 198, 201–210. doi: 10.1016/j.envpol.2014.12.014

Florida, R. (2017). *The New Urban Crisis: Gentrification, Housing Bubbles, Growing Inequality, and What We Can Do About It*. Oneworld Publications.

Gasco, L., Asensio, C., and De Arcas, G. (2017). "Towards the assessment of community response to noise through social media," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings, Vol. 255* (Hong Kong: Institute of Noise Control Engineering), 2209–2217.

Gasco, L., Clavel, C., Asensio, C., de Arcas, G. (2019). Beyond sound level monitoring: Exploitation of social media to gather citizens subjective response to noise. *Sci. Total Environ*. 658, 69–79. doi: 10.1016/j.scitotenv.2018.12.071

Hayes, S. N., and Taler, S. J. (1998). "Hypertension in women: current understanding of gender differences," in *Mayo Clinic Proceedings, Vol. 73* (Elsevier), 157–165. doi: 10.1016/S0025-6196(11)63649-5

Hurtley, C. (2009). *Night Noise Guidelines for Europe*. WHO Regional Office Europe.

International Organization for Standardization (2007). *ISO 1996-2:2007 Acoustics - Description, Measurement and Assessment of Environmental Noise - Part 2: Determination of Environmental Noise Levels*.

Kang, J., Aletta, F., Gjestland, T. T., Brown, L. A., Botteldooren, D., Schulte-Fortkamp, B., et al. (2016). Ten questions on the soundscapes of the built environment. *Build. Environ*. 108, 284–294. doi: 10.1016/j.buildenv.2016.08.011

Kaplan, M. S., Huguet, N., Feeny, D. H., and McFarland, B. H. (2010). Self-reported hypertension prevalence and income among older adults in Canada and the United States. *Soc. Sci. Med*. 70, 844–849. doi: 10.1016/j.socscimed.2009.11.019

Keenan, N. L., Rosendorf, K. A., for Disease Control, C., (CDC), P., et al. (2011). Prevalence of hypertension and controlled hypertension–United States, 2005-2008. *MMWR Surveill. Summ*. 60(Suppl.), 94–97. Available online at: https://www.cdc.gov/Mmwr/preview/mmwrhtml/su6001a21.htm?s_cid=su6001a21_wAuthor

Lee, P. J., Park, S. H., Jeong, J. H., Choung, T., and Kim, K. Y. (2019). Association between transportation noise and blood pressure in adults living in multi-storey residential buildings. *Environ. Int*. 132:105101. doi: 10.1016/j.envint.2019.105101

Leon Bluhm, G., Berglind, N., Nordling, E., and Rosenlund, M. (2007). Road traffic noise and hypertension. *Occup. Environ. Med*. 64, 122–126. doi: 10.1136/oem.2005.025866

Lindeman, R. H., Merenda, P., and Gold, R. Z. (1980). *Introduction to Bivariate and Multivariate Analysis, Glenview, Il*. Scott: Foresman and Company.

Lorente-Riverola, I., and Ruiz-Sánchez, J. (2018). "Spatial footprints of context-aware digital services. Eventual self-regulated alignments of dating apps with the urban shape," in *Smart and Sustainable Planning for Cities and Regions*, eds A. Bisello, D. Vettorato, P. Laconte, and S. Costa (Cham: Springer International Publishing), 133–150. doi: 10.1007/978-3-319-75774-2_10

Mosca, I., Bhuachalla, B. N., and Kenny, R. A. (2013). Explaining significant differences in subjective and objective measures of cardiovascular health: evidence for the socioeconomic gradient in a population-based study. *BMC Cardiovasc. Disord*. 13:64. doi: 10.1186/1471-2261-13-64

Mydlarz, C., Sharma, M., Lockerman, Y., Steers, B., Silva, C., and Bello, J. P. (2019). The life of a New York city noise sensor network. *Sensors* 19. doi: 10.3390/s19061415

Ottoz, E., Rizzi, L., and Nastasi, F. (2018). Recreational noise: Impact and costs for annoyed residents in Milan and Turin. *Appl. Acoust*. 133, 173–181. doi: 10.1016/j.apacoust.2017.12.021

Pinto, E. (2007). Blood pressure and ageing. *Postgrad. Med. J*. 83, 109–114. doi: 10.1136/pgmj.2006.048371

Quercia, D., Aiello, L. M., and Schifanella, R. (2016). "The emotional and chromatic layers of urban smells," in *Tenth International AAAI Conference on Web and Social Media* (Cologne).

Quercia, D., Schifanella, R., Aiello, L. M., and McLean, K. (2015). Smelly maps: the digital life of urban smellscapes. *arXiv [preprint] arXiv:1505.06851*. Available online at: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/viewFile/10572/10516

Radicchi, A. (2017). "Hush city: A new mobile application to crowdsource and assess–everyday quiet areas in cities," in *Proceedings of Invisible Places: the*

International Conference on Sound, Urbanism and the Sense of Place (Azores), 7–9.

Richardson, E. A., Pearce, J., Tunstall, H., Mitchell, R., and Shortt, N. K. (2013). Particulate air pollution and health inequalities: a Europe-wide ecological analysis. *Int. J. Health Geogr.* 12:34. doi: 10.1186/1476-072X-12-34

Sassen, S. (1991). *The Global City*. New York, NY; London; Tokyo; Princeton, NJ: Princeton University Press.

Savage, M., Devine, F., Cunningham, N., Taylor, M., Li, Y., Hjellbrekke, J., et al. (2013). A new model of social class? Findings from the BBC's great British class survey experiment. *Sociology* 47, 219–250. doi: 10.1177/0038038513481128

Schafer, R. M. (1993). *The Soundscape: Our Sonic Environment and the Tuning of the World*. Destiny Books.

Schmit, C., and Lorant, V. (2009). Noise nuisance and health inequalities in Belgium: a population study. *Arch. Public Health* 67:52. doi: 10.1186/0778-7367-67-2-52

Serdyukov, P., Murdock, V., and Van Zwol, R. (2009). "Placing flickr photos on a map," in *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Boston, MA), 484–491. doi: 10.1145/1571941.1572025

Silva, B., Santos, G., Eller, R., and Gjestland, T. (2017). Annoyance survey by means of social media. *J. Acoust. Soc. Am.* 141, 1019–1026. doi: 10.1121/1.4976055

Sørensen, M., Hvidberg, M., Hoffmann, B., Andersen, Z. J., Nordsborg, R. B., Lillelund, K. G., et al. (2011). Exposure to road traffic and railway noise and associations with blood pressure and self-reported hypertension: a cohort study. *Environ. Health* 10:92. doi: 10.1186/1476-069X-10-92

Stansfeld, S., and Crombie, R. (2011). Cardiovascular effects of environmental noise: research in the united kingdom. *Noise Health* 13:229. doi: 10.4103/1463-1741.80159

Thomee, B., Moreno, J. G., and Shamma, D. A. (2014). "Who's time is it anyway? Investigating the accuracy of camera timestamps," in *Proceedings of the 22nd ACM International Conference on Multimedia* (Orlando, FL), 909–912. doi: 10.1145/2647868.2655066

United Nations (2018). *World Urbanization Prospects: The 2018 Revision. Key Facts*. Technical report, United Nations.

Urquiaga, A. A., Lorente-Riverola, I. Sanchez, J. R. (2020). Platform-mediated short-term rentals and gentrification in Madrid. *Urban Stud.* doi: 10.1177/0042098020918154

Van Kempen, E., Casas, M., Pershagen, G., and Foraster, M. (2018). Who environmental noise guidelines for the European region: a systematic review on environmental noise and cardiovascular and metabolic effects: a summary. *Int. J. Environ. Res. Public Health* 15. doi: 10.3390/ijerph150 20379

Wachsmuth, D., and Weisler, A. (2018). Airbnb and the rent gap: gentrification through the sharing economy. *Environ. Plan. Econ. Space* 50, 1147–1170. doi: 10.1177/0308518X18778038

Wallsten, S. (2015). *The Competitive Effects of the Sharing Economy: How Is Uber Changing Taxis.* Technology Policy Institute, 22.

World Health Organization (2018). *Environmental Noise Guidelines for the European Region*. World Health Organization.

Zuo, J., Xia, H., Liu, S., and Qiao, Y. (2016). Mapping urban environmental noise using smartphones. *Sensors* 16:1692. doi: 10.3390/s161 01692

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 14 October 2020
doi: 10.3389/fpsyg.2020.570694

Anthropogenic Noise Source and Intensity Effects on Mood and Relaxation in Simulated Park Environments

Jacob A. Benfield[1], Gretchen A. Nurse Rainbolt[2], Lucy J. Troup[3] and Paul A. Bell[2]*

[1] Department of Psychological and Social Science, Pennsylvania State University, Abington, PA, United States, [2] Department of Psychology, Colorado State University, Fort Collins, CO, United States, [3] School of Education and Social Sciences, University of the West of Scotland, Paisley, SC, United Kingdom

OPEN ACCESS

Edited by:
PerMagnus Lindborg,
City University of Hong Kong,
Hong Kong

Reviewed by:
Kurt M. Fristrup,
National Park Service Fort Collins,
United States
Jochen Steffens,
University of Applied Sciences,
Germany

***Correspondence:**
Jacob A. Benfield
jab908@psu.edu

Specialty section:
This article was submitted to
Environmental Psychology,
a section of the journal
Frontiers in Psychology

Received: 08 June 2020
Accepted: 16 September 2020
Published: 14 October 2020

Citation:
Benfield JA, Rainbolt GAN,
Troup LJ and Bell PA (2020)
Anthropogenic Noise Source
and Intensity Effects on Mood
and Relaxation in Simulated Park
Environments.
Front. Psychol. 11:570694.
doi: 10.3389/fpsyg.2020.570694

Research on human caused sound has shown a wide range of effects in outdoor environments as well as laboratory simulations of those environments. Aircraft noise, ground traffic, and human voices have all been shown to lower scenic evaluation ratings and influence individual reports of affective state. However, previous research has relied entirely on pre-post measures of affect and psychological state rather than more momentary assessments. The current project utilized a time series of 15 measurements of overall mood and relaxation collected during a 30-min period during which participants (N = 229) were exposed to randomized volume levels of natural sounds, natural sounds with aircraft noise, natural sounds with ground traffic, or natural sounds with human voices added. Results supported previous findings with significant sound type X volume interactions showing differing rates of decline for both outcomes. Natural sounds did not relate to the diminishing effects observed for the three anthropogenic sound conditions.

Keywords: soundscapes, aircraft, transportation noise, national parks, stressors

INTRODUCTION

Noise, defined as *unwanted* or *harmful* sound, is often considered an ambient stressor by environmental psychologists because noise can place demands on us to cope or adapt while simultaneously influencing our psychological well-being (Seidman and Standring, 2010). For example, Evans et al. (2001) showed that children chronically exposed to noise sources such as road traffic at average sound levels greater than 60 dB had elevated systolic blood pressure and higher overnight urinary cortisol levels when compared to children with day-to-day exposure at lower sound intensity levels (<50 dB). The high-exposure children also self-reported higher perceived stress levels and demonstrated high physiological reactivity in laboratory manipulations, suggesting that noise exposure was interfering with relaxation while promoting stress and heightened physiological arousal. In addition to physiological stress responses, noise also impacts perceptual and psychological reactions to the environment. For example, research has found that positive affective states are compromised when specific sounds are perceived as "noise" (e.g.,Tarrant et al., 1995).

Noise and Natural Environments

The United States National Park Service (NPS) has jurisdiction over hundreds of natural, historical, and cultural sites throughout the United States of America and has been charged with both preserving the natural ecosystem for future generations and allowing for public visitation and recreation (Yellowstone Act of 1872, 1872; National Wilderness Preservation Act of 1964, 1964). Additional mandates requiring the federal government to assess and monitor soundscapes in National Parks demonstrate recognition of the ambient acoustic environment as one of the potential elements influencing both wildlife and visitor experiences (Noise Control Act of 1972, 1972). Moreover, other legislation (e.g., National Parks Air Tour Management Act of 2000, 2000) has been enacted to specifically target the management and study of specific noise sources such as aircraft overflights or recreational vehicles such as snowmobiles.

To facilitate these research and management goals, scientists have identified key indicators used to evaluate various aspects of soundscapes in recreational settings. Some well-established factors used to evaluate visitor experience have been based on outcomes related to stated preferences (Driver et al., 1987), landscape assessment (Daniel 2001), and affective responses to natural environments (Kaplan, 1995). For example, in the context of landscape assessment, Mace et al., 1999 found that the evaluations of scenic park landscapes along 8 dimensions were significantly lower in the presence of both 40 and 80 db helicopter overflight sounds. Follow-up research showed the effect to be salient across different noise source attributions (Mace et al., 2003) and additional noise source types such as human voices and automobile traffic (Benfield et al., 2010).

Noise and Affective Responses to Natural Environments

At the same time, the introduction of noise to a landscape not only impacts scene assessments, but also impacts the affective state of those engaged in the assessing. However, the results of those tests have been less consistent and harder to interpret collectively. For instance, Mace et al. (1999, 2003) and Benfield et al. (2010) each found that the addition of noise diminished landscape evaluations, but all also showed discrepant findings when it came to measures of affect. In those series of related studies, each found different combinations of changes in positive and negative affect, in spite of all three using comparable methodologies and stimuli.

Specifically, all three studies utilized the Positive and Negative Affect Schedule (PANAS; Watson and Clark, 1999) and a pre-post procedure for measuring change from exposure. All three studies required participants to evaluate various landscapes on a large screen, projected in a dark room with sounds presented at comparable intensity levels using surround sound speakers. Each study session lasted approximately one hour, with half of that time spent on the noise exposure and scenic evaluation task. Finally, all three studies demonstrated a decrease in positive affect in the presence of noise, although Benfield et al. (2010) showed the decrease in positive affect for all conditions, including the natural and control sound conditions, suggesting something other than the experimental manipulation.

In addition to some variance in positive affect findings, the results for negative affect were also inconsistent across conditions and studies. The original Mace et al. (1999) study showed no effect of noise on negative affect and attributed this finding, when paired with the loss of positive affect, as indicating a decrease in pleasure but not an increase in annoyance. However, the follow-up work in 2003 by Mace and colleagues did show increased negative affect, but not for all noise conditions. It was suggested that this effect may be an artifact of methodology or the result of different situational attributions assigned to the noise. Benfield et al. (2010) also failed to show an effect on negative affect but because of the global decrease in positive affect, suspected another factor at play. Specifically, Benfield et al. (2010) argued that the methodology, which was identical to Mace et al. (1999) and very similar to Mace et al. (2003), was causing participant fatigue and boredom which would explain a lowering of positive affect as well as the null effect on negative affect. Such a possibility has not been directly tested in laboratory soundscape research, yet could alter the interpretation of previous work and better inform future simulation work on this key management indicator.

The Current Study

Though previous research showed a relationship between noise exposure and mood and stress, the temporal evolution of these responses has not been documented. Substantial intervals between affective assays introduced potential confounding explanations, and may explain some of the disparate results. Those previous effects are often shown cross-sectionally or with simple pre-post measures of mood separated by up to an hour of time. The current study aimed to rectify that confound in methodology by measuring affective valence and arousal levels at regular intervals, throughout the scenic evaluation task, rather than in a pre-post framework. Additionally, the current study included direct self-report measures of both fatigue and effort, also recorded at regular intervals throughout the task, to control for additional factors that could explain prior findings in this domain.

Moreover, laboratory based research in this area is not always in agreement concerning the size, cause, or direction of effect (e.g., Mace et al., 1999 compared to Benfield et al., 2010) and often fails to test the effect of sound intensity alongside sound type. Specifically, even though participants in previous studies were exposed to different sound intensities during the evaluation task, the pre-post measurement procedure prevented testing the effect of intensity on mood. As such, those previous projects demonstrated clear connections between sound intensity level and scenic evaluations, but never to changes in affective state. Therefore, it was also the purpose of the current study to examine the interaction of sound intensity and sound type on affective valence and arousal. That is, the purpose of the current project was to test the robustness of the previous findings on natural soundscapes' effect on mood and stress with a previously unutilized methodology and set of measurements.

MATERIALS AND METHODS

Overall Design

A 4 (sound type) x 3 (sound intensity) mixed factorial timeseries design was utilized in this study. One of four soundscape conditions (between-subjects factor) were randomly administered to participants who each experienced three randomly presented sound intensity levels for that soundscape (within-subjects factor). Assessments of mood, relaxation, fatigue, and effort were made every two minutes throughout the 30-min landscape evaluation task for a total of 15 individual measurements, five for each of the sound intensity levels.

Participants

A total of 229 undergraduates (140 females; 89 males) participated in the research as partial fulfillment of a compulsory course research requirement. Participants were about 19 years old (M = 19.30, SD = 2.20, Range = 18 – 43) and reported visiting an average of 4 or 5 United States national parks in their lifetimes (M = 4.78, SD = 3.56, Range = 0 – 19). The majority of participants were of European descent (87%).

Materials and Measures

Predictor and Outcome Measures

The Weinstein Noise Sensitivity Scale (WNS) measures individual sensitivity to unwanted sounds or noises (Weinstein, 1978). It consists of 21 items rated on a 6-point scale ranging from 'strongly disagree' to 'strongly agree.' Summation of items creates a single score for overall noise sensitivity (α = 0.84). Previous research has shown noise sensitivity to be an important covariate when assessing the effects of noise on humans (Ellermeier et al., 2001; Miedema and Vos, 2003) and to be important in measuring recreation noise acceptability (Benfield et al., 2014).

Visual analog scales (VAS) were used for the measurement of both outcome variables (overall mood and overall level of relaxation) as well as two control variables (participant level of fatigue and overall effort). These VAS measures consisted of a five-inch (12.7 cm) horizontal line representing the continuum from "very low" to "very high" (or "very negative" to "very positive" for the mood rating). Participants responded by making a vertical mark along the continuum, and scores were calculated by measuring the distance of the vertical mark from the left edge of the continuum in $1/8^{th}$ inch increments (0.3 cm). This created a range of 0–40 points which provided a continuous, ratio distribution of the response data while helping to reduce range restriction sometimes observed when responding with discrete values (e.g., the typical 1–7 range used in Likert-type response measures). Research on VAS rating scales shows them to be a sensitive, valid, and reliable technique for obtaining participant responses across a range of phenomena including pain, attractiveness, self-esteem, and mood, especially when the time between responses is limited (Folstein and Luria, 1973; Price et al., 1983; Brumfitt and Sheeran, 1999; Grant et al., 1999; Rankin and Borah, 2003; Hasson and Arnetz, 2005).

Soundtracks

Using an acoustics database maintained by the NPS, four different sound recordings were used for the auditory manipulation. The natural sound condition (wind through foliage with mixed bird calls) was used as a baseline for the other three sound conditions: natural with voices, natural with ground traffic, and natural with air traffic. Each sound clip was created by adding isolated recordings of the noise (e.g., voices) to the natural baseline. For these sound conditions, the added element was present on an almost continuous basis with the longest gap between noise events being less than 10 *s*.

Visual Stimuli

A set of 30 scenes was assembled as the visual stimuli presented for rating. The first 5 scenes were practice slides to familiarize the participants with the procedure, rating sheet, and slide timings. The remaining 25 slides were target slides representing five scenes each from five national parks—Yellowstone, Olympic, Saguaro, Grand Canyon, and Everglades—which were chosen from a set of high-resolution photographs taken within each park. While the order of each park was presented randomly, the five scenes within each park were shown consecutively in a non-randomized order.

Procedure

Participants signed up for the study using an online recruitment website associated with an introductory psychology course and attended a single, one-hour research session. After completing an initial informed consent sheet, participants were then randomly assigned to one of the four sound conditions. All experimental sessions were conducted with participants seated 10 ft (3 m) away from a 6 × 6 ft (1.8 m × 1.8 m) screen mounted at the front of an 18 × 18 ft (5.5 m × 5.5 m) room. Scenes were presented on the screen via computer projector, and sounds were presented using a 4-channel surround sound system placed in the corners of the room.

A brief demographic questionnaire containing the WNS and other measures was presented at the beginning of the larger research packet. Upon completion, participants were then given instructions concerning the scenic evaluation task. The scenic evaluation task required participants to rate five practice slides and then view the set of 25 target slides three separate times in each session (20 *s* per slide; 80 slide ratings total). Each of the three runs of the 25 target slides was accompanied by one of the three sound levels: (1) control, or no added sounds to the 40–45 dB(A) background from the room; (2) low volume added sounds of 40–45 dB(A); or (3) high volume added sounds of 60–65 dB(A). The three sound levels were presented in one of four random orders with gradual changes in sound intensity occurring over a 20 *s* period at the change of conditions.

The VAS ratings for mood, relaxation, fatigue, and effort were presented after every set of five slides starting after the five practice slides. Because of the number of slide sets and the timing of the sets, a total of 15 VAS measurements were taken over the course of the 30-min scenic evaluation task with measurements spaced 2 min apart. At the conclusion of the evaluation task, participants were fully debriefed with regard to the purposes and methods of the project and given course research credit for compensation.

Data Analysis Strategy

Growth modeling is a multilevel data analysis technique that allows researchers to examine longitudinal or time series data for both intraindividual change (how do people change over time) and interindividual differences in change (how do people differ in how they change over time; Henry and Slater, 2007). In a multilevel framework, level 1 consists of the multiple measurement occasions recorded for an individual; level 2 is made up of the individuals themselves. Within the current study, growth modeling allows for the examination of change within individuals based on the length of time they have been participating, the changes in volume level that have occurred, or their individual level of fatigue/effort (level 1). It also allows for the simultaneous assessment of differences in change between sound exposure conditions or noise sensitivity scores (level 2).

TABLE 1 | Model summaries for VAS mood outcome scores.

	Model A (Means)	Model B (Linear)	Model C (Quadratic)	Model D (Experimental)	Model E (WNS)	Model F (Fatigue/Effort)
Fixed Effects (Mood)						
Initial Status						
Intercept	27.966**	28.974**	30.111**	29.778**	29.748**	20.708**
Automobile				2.088	2.200	1.590
Auto * Noise sensitivity					−0.124	−0.086
Aircraft				2.786**	2.802**	2.375**
Aircraft * Noise Sensitivity					−0.106	−0.130
Voices				1.167	1.101	1.051
Voices * Noise Sensitivity					−0.224*	−0.197*
Volume Order (C-65-45)				−0.464	−0.298	−0.194
Volume Order (45-65-C)				−0.368	−0.399	−0.280
Volume Order (65-45-C)				−0.201	−0.446	−0.547
Volume(45 dBA)				0.520	0.553	0.618
Volume(65 dBA)				−1.113*	−1.104*	−1.103*
Vol(45 dBA) * Auto				−2.574**	−2.571**	−2.325**
* Aircraft				−3.357**	−3.392**	−3.312**
* Voices				−3.225**	−3.327**	−3.365**
* WNS					−0.055*	−0.048*
Vol(65 dBA) * Auto				−3.306**	−3.307**	−2.792**
* Aircraft				−4.639**	−4.645**	−4.247**
* Voices				−4.400**	−4.445**	−4.386**
* WNS					−0.049	−0.053
Fixed Effects (Cont.)						
Noise Sensitivity					0.016	0.011
Fatigue (Level 1)						−0.093**
Effort (Level 1)						0.287**
Fatigue (Level 2)						−0.175**
Effort (Level 2)						0.372**
Rate of Change						
Time		−0.145**	−0.669**	−0.254*	−0.248*	−0.279**
Quadratic Time						
$Time^2$			0.037**	0.010	0.010	0.014*
Variance Components						
Level 1 – Within	29.883**	29.622**	26.107**	25.994**	25.986**	23.782**
Variance Explained	2.45%	6.62%	1.82%	6.30%	19.82%	
Level 2 – Between	45.070**	43.496**	42.171**	41.042**	36.830**	26.581**
Variance Explained	3.49%	3.07%	2.67%	10.24%	27.77%	
Model Fit						
Log Likelihood (CF for MLR)	−11940.076(4.762)	−11473.485(2.678)	−11343.224(2.154)	−11100.654(1.532)	−11086.625(1.468)	−10933.910(1.537)
Parameters	3	6	10	33	39	43
Chi-square Change (TRD)		1571.013	190.4401	384.554	25.14158	138.2183
p-value		< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

$^*p < 0.05$; $^{**}p < 0.01$. Referent Groups: Natural Sounds; Control-45–65 dBA Volume Order; 0 dB Volume (control).

For both outcomes—overall mood and relaxation—a series of six growth models were created. The first three (Models A – C in **Tables 1**, **2**) represent unconditional models designed to provide baseline measures of variance and model fit as well as to assess the best way to model the effect of time (i.e., measurement occasion) onto the data. The remaining three models (D – F in **Tables 1**, **2**) represent conditional models in which the effects of the experimental manipulations (Model D), individual noise sensitivity (Model E), and participant fatigue/effort (Model F) on the outcome variable are modeled. For the purposes of brevity, only conditional models are discussed extensively. It suffices to say testing of unconditional models A-C showed that significant amounts of both within- and between-person variance existed (ICC = 0.601 and 0.608 for mood and relaxation, respectively), and that quadratic models were preferable to linear models of change over time.

All analyses were conducted using the SAS PROC MIXED procedure (Singer, 1998; Singer and Willett, 2003). Tests for

TABLE 2 | Model summaries for VAS relaxation outcome scores.

	Model A (Means)	Model B (Linear)	Model C (Quadratic)	Model D (Experimental)	Model E (WNS)	Model F (Fatigue/Effort)
Fixed Effects (Relax)						
Initial Status						
Intercept	24.841**	25.427**	26.879**	27.061**	26.988**	17.597**
Automobile				1.086	1.185	0.585
Auto * NoiseSensitivity					−0.111	−0.072
Aircraft				1.380	1.405	0.920
Aircraft * NoiseSensitivity					−0.106	−0.118
Voices				1.179	1.044	0.845
Voices * NoiseSensitivity					−0.259**	−0.240*
Volume Order (C-65-45)				0.409	0.692	0.939
Volume Order (45-65-C)				1.044	1.017	1.110
Volume Order (65-45-C)				0.680	0.607	0.583
Volume(45 dBA)				−0.388	−0.387	−0.329
Volume(65 dBA)				−2.555**	−2.553**	−2.558**
Vol(45 dBA) * Auto				−2.118**	−2.092**	−1.951**
* Aircraft				−4.337**	−4.327**	−4.310**
* Voices				−4.108**	−4.132**	−4.167**
* WNS					−0.040	−0.034
Vol(65 dBA) * Auto				−3.118**	−3.118**	−2.763**
* Aircraft				−6.088**	−6.082**	−5.787**
* Voices				−4.436**	−4.453**	−4.364**
* WNS					−0.036	−0.040
Fixed Effects (Relax)						
WNS					0.005	−0.001
Fatigue (Level 1)						−0.101**
Effort (Level 1)						0.232**
Fatigue (Level 2)						−0.177**
Fatigue (Level 2)						0.384**
Rate of Change						
Intercept		−0.086	−0.755**	−0.147	−0.143	−0.160
Quadratic Time						
$Time^2$			0.048**	0.006	0.006	0.009
Variance Components						
Level 1 – Within	37.546**	37.431**	33.947**	33.677**	33.675	31.706**
Variance Explained	1.67%	4.34%	1.43%	5.10%	14.31%	
Level 2 – Intercept	58.282**	56.798**	56.191**	55.171**	50.632**	40.536**
Variance Explained	2.54%	1.09%	1.81%	8.19%	19.90%	
Model Fit						
Loglikelihood (CF for MLR)	−12366.300(3.317)	−11938.664(2.328)	−11790.389(1.759)	−11501.677(1.317)	−11490.211(1.287)	−11393.431(1.373)
Parameters	3	6	10	33	39	43
Chi-square Change (TRD)		638.7394	327.4986	513.3451	20.4385	87.5243
p-value		< 0.001	< 0.001	< 0.001	0.002	< 0.001

$^*p < 0.05$; $^{**}p < 0.01$. Referent Groups: Natural Sounds; Control-45–65 dBA Volume Order; 0 dB Volume (control).

multicollinearity showed variance inflation factor (VIF) values within an acceptable range (VIF = 1.06–2.49). Variance explained values are based on pseudo-R^2 statistics for both levels of the model (Singer and Willett, 2003).

RESULTS

Overall Mood

For all three conditional models, the inclusion of the added parameters significantly improved model fit and supported previous research on both sound type and sound volume. The inclusion of the experimental variables (i.e., sound type, sound volume, sound order, and a sound type X sound volume interaction term) showed no main effect for sound order, sound type, or the low-volume condition. However, the model did demonstrate a negative effect for the high-volume condition, and all three noise types significantly interacted with both the high- and low-volume conditions (Model D in **Table 1**). Noise significantly decreased mood ratings, but the size of the detriment varied depending on the type of sound and the volume level of the sound. High-volume exposure was always more detrimental than low-volume exposure with human voices (β = −3.23 for low volume; −4.40 for high volume) and aircraft noises (β = −3.36 for low volume; −.64 for high volume) having a larger effect than automobile traffic noise (β = −2.57 for low volume; -3.31 for high volume). The inclusion of the experimental variables explained an additional 1.82% of variance in scores across time points and 2.67% of the variance in individual average mood scores.

The pattern of results shown in the first conditional model persisted through the other two conditional models. Significant sound type X volume level interactions showed high volume levels to be more problematic than low volume levels with human voices and aircraft noise being more bothersome than automobile noise. The addition of the individual noise sensitivity covariate explained 6.30% of variance across measurement occasions and 10.24% of average mood score variance (Model E in **Table 1**). While no main effect for noise sensitivity was shown, a significant interaction with the low volume condition (β = 0.06) showed that higher sensitivity to noise related to a larger negative effect in the low volume condition. That interaction was not significant for the high-volume condition. A similar interaction existed between noise sensitivity and the human voices condition. Greater sensitivity to noise related to lower mood scores when exposed to human voices (β = −0.22); the effect was not shown for aircraft or automobile noises. The fatigue and effort covariates also explained large portions of remaining variance both across measurements and in overall scores (19.82% and 27.77%, respectively), with both covariates having significant main effects on mood (Model F in **Table 1**). Higher than average fatigue related to decreased mood scores (β = −0.09 within; -0.18 between) while above average effort related to improved mood scores (β = 0.28 within; 0.37 between). All totaled, the addition of the experimental parameters, noise sensitivity, fatigue, and effort into the model (Model F) explained 26.24% of the variance observed across measurement occasions, and 36.90% of the variance in average mood scores across individuals compared to the unconditional quadratic model (i.e., compared to changes occurring only due to time; Model C). **Figure 1** displays prototypical mood trajectories based on the average score of all covariates—WNS, fatigue, and effort—across time, volume levels, and sound conditions based on the parameters given in the full model (**Table 1**, Model F).

Level of Relaxation

Similar to the findings related to overall mood scores, level of relaxation was significantly affected by the experimental conditions (see Model D in **Table 2**). Once again, no main effect for sound order, sound type, or low volume was shown but significant sound type X volume level interactions mirrored the pattern of results for mood ratings. Human voices (β = −4.11 for low volume; −4.44 for high volume) and aircraft noise (β = −4.34 for low volume; −6.09 for high volume) were more detrimental than automobile traffic noise (β = 2.12 for low volume; 3.12 for high volume); high volume conditions were more problematic than low volume conditions. The experimental variables explain 1.43% of variability across measurement occasions and 1.81% of variability in average overall scores while significantly improving model fit, X^2 (23) = 513.35, $p < 0.001$.

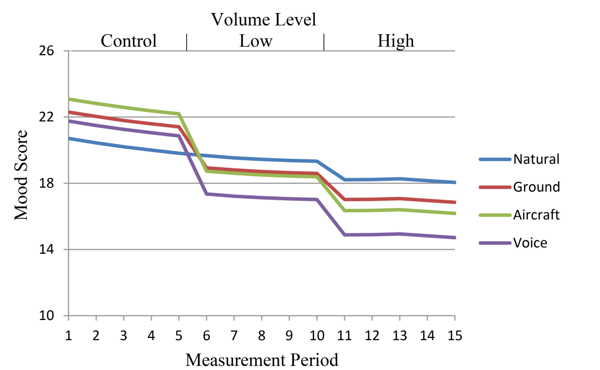


FIGURE 1 | Prototypical change in mood score based on sound type, volume level, and measurement occasion assuming average levels of noise sensitivity, fatigue, and effort.

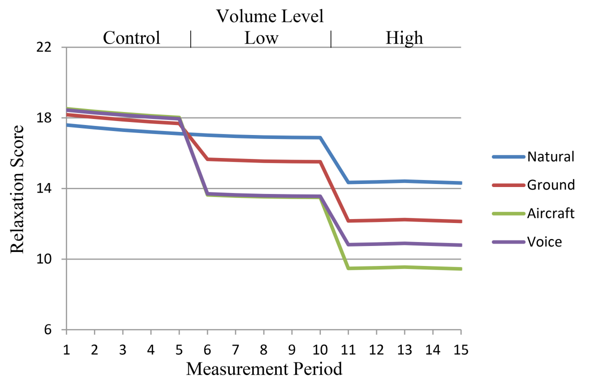


FIGURE 2 | Prototypical change in relaxation score based on sound type, volume level, and measurement occasion assuming average levels of noise sensitivity, fatigue, and effort.

Subsequent models that included the noise sensitivity covariate (Model E in **Table 2**) and the fatigue and effort covariates (Model F in **Table 2**) also significantly improved model fit and explained additional within person variance (5.10% for Model E; 14.31% for Model F) and between person variance (8.19% for Model E; 19.90% for Model F). The effects of the experimental variables shown in Model D persisted in the fuller models showing that volume and sound type combine to lower relaxation scores even after controlling for the effects of noise sensitivity and participant fatigue and fluctuation in effort. **Figure 2** displays prototypical trajectories for the full model (Model F in **Table 2**) assuming average levels of noise sensitivity, fatigue, and effort; the full model explained 19.85% of the variability in relaxation across time and 27.81% of the variability in average individual relaxation scores in comparison to the unconditional quadratic model (Model C in **Table 2**).

DISCUSSION

Anthropogenic noise, especially at high intensity, decreases individual mood and relaxation while natural sounds have a lessened or null effect depending entirely on sound intensity levels. That noise type by intensity detriment persists even after controlling for individual change across time as well as noise sensitivity, experimental fatigue, and task effort. This finding is consistent with previous research (e.g., Mace et al., 1999; 2003) and builds upon that evidence by showing that the negative effect upon mood or relaxation is not solely an artifact of study design or participant fatigue as others have suggested (e.g., Benfield et al., 2010). It is noteworthy, however, that fatigue and effort were able to account for much more variability in overall mood and relaxation scores than either noise sensitivity or the actual acoustic stimuli alone, suggesting that Benfield and colleagues' concern about laboratory fatigue or overall effort driving mood findings was not unreasonable. Rarely are such variables directly measured and controlled for in laboratory simulations on noise, and the current data suggest that each can make a substantial impact on findings, especially when measured regularly and alongside certain outcomes. Future research should explore the varied impact of such confounds on similar environmental research and also regularly include controls for them in most, if not all, laboratory simulations and other studies in which exposure is prolonged and participant motivation is potentially less than optimal.

Additionally, noise sensitivity has a measurable influence at low-volume sound levels. High-sensitivity individuals were more disturbed than low-sensitive individuals at low volume levels, but sensitivity had minimal impact at high-volume levels. This information has implications for policy and research. For instance, researchers can better anticipate contexts in which noise sensitivity measurement is more or less crucial to accounting for differences. High intensity sound exposure may not show noise sensitivity effects and therefore may not require controlling for such variables. Likewise, low intensity sounds may only elicit effects when interacting with noise sensitivity and such studies should include measures for sensitivity. In the context of management policy, protected areas visited regularly by persons with higher noise sensitivity (e.g., locations known for unique or subtle sound qualities) may consider more stringent noise abatement strategies, even for sounds that may be physically less intense but reported as problematic. In other words, effective management strategies may require prioritizing subjective visitor ratings of acceptability over objective acoustic measurements of intensity. Similarly, individuals with higher noise sensitivity may be made more aware of how that trait interacts with their perception of sounds as a way to reframe experiences and potentially mitigate conflict with others.

The testing of complex interactions or nuanced effects, such as the role of fatigue on mood across time based on differing sound intensity, is best accomplished under controlled laboratory conditions. The ability to generate causal conclusions regarding naturally occurring phenomenon provides soundscape researchers with a strong foundation on which to build future projects. However, soundscape research and the problems surrounding noise exposure are often more applied and practical in nature. As such, the current project tells us a lot about the role of noise on affective state, arousal, and fatigue in highly controlled situations, but more ecologically valid and intensive field-based studies will be necessary to fully understand the practical implications of these effects.

Ultimately, the repeated measures design of this study provided stronger controls over temporal effects and improved capacity to address differences among subjects. All the effects on mood were shown to occur within very short timeframes (i.e., 2-min evaluation intervals) and can be reversed by lowering sound intensity or removing the stimulus. This has implications for future research and management policy alike. Likewise, these findings demonstrate that growth modeling can detect effects of noise that may be difficult to demonstrate with more traditional statistical techniques. Social science noise researchers may find that more intensive, time-interval based techniques provide stronger evidence of noise impact and can lend itself to more ecologically valid, non-laboratory assessment of noise impacts such as ecological momentary assessments (EMA; Steffens et al., 2017). Such approaches are regularly used by health researchers but less often by environmental psychologists or others within the social science-oriented noise research community. As the current data demonstrate, this relative lack of time-series or momentary assessment is influencing our understanding of the nuance found within human perceptions of soundscape and noise research.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Colorado State University Institutional

Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by Cooperative Agreement No. H2380040002, Metrics of Human Responses to Natural Sound Environments from the National Park Service. Grantees undertaking projects under government sponsorship are encouraged to express freely their findings and conclusions. Points of view or opinions do not, therefore, necessarily represent official National Park Service policy.

REFERENCES

Benfield, J. A., Bell, P. A., Troup, L. J., and Soderstrom, N. C. (2010). Aesthetic and affective effects of vocal and traffic noise on natural landscape assessment. *J. Environ. Psychol.* 30, 103–111. doi: 10.1016/j.jenvp.2009.10.002

Benfield, J. A., Nurse, G. A., Jakubowski, R., Gibson, A. W., Taff, B. D., Newman, P., et al. (2014). Testing noise in the field: a brief measure of individual noise sensitivity. *Environ. Behav.* 46, 353–372. doi: 10.1177/0013916512454430

Brumfitt, S. M., and Sheeran, P. (1999). The development and validation of the visual analogue self-esteem scale (VASES). *Br. J. Clin. Psychol.* 38, 387–400. doi: 10.1348/014466599162980

Driver, B. L., Nash, R., and Haas, G. (1987). "Wilderness benefits: a state-of-knowledge view," in *Proceedings of the National Wilderness Research Conference: Issues, State of Knowledge, Future Directions. USDA Forest Service General Technical Report INT-220*, Ogden, UT, 294–319.

Ellermeier, W., Eigenstetter, M., and Zimmer, K. (2001). Psychoacoustic correlates of individual noise sensitivity. *J. Acoust. Soc. Am.* 109, 1464–1473. doi: 10.1121/1.1350402

Evans, G. W., Lercher, P., Meis, M., Ising, H., and Kofler, W. W. (2001). Community noise exposure and stress in children. *J. Acoust. Soc. Am.* 109, 1023–1027. doi: 10.1121/1.1340642

Folstein, M. F., and Luria, R. (1973). Reliability, validity, and clinical application of the visual analog mood scale. *Psychol. Med.* 3, 479–486. doi: 10.1017/s0033291700054283

Grant, S., Aitchison, T., Henderson, E., Christic, J., Zare, S., McMurray, J., et al. (1999). A comparison of the reproducibility and the sensitivity to change of visual analog scales, Borg scales, and Likert scales in normal subjects during submaximal exercise. *Chest* 116, 1208–1217. doi: 10.1378/chest.116.5.1208

Hasson, D., and Arnetz, B. B. (2005). Validation and findings comparing VAS vs. Likert scales for psychosocial measures. *Int. Electron. J. Health Educ.* 8, 178–192.

Henry, K. L., and Slater, M. D. (2007). "Assessing change and intraindividual variation: longitudinal multilevel and structural equation modeling," in *The SAGE Sourcebook of Advanced Data Analysis Methods for Communication Research*, eds A. F. Hayes, M. D. Slater, and L. B. Snyder (Thousand Oaks, CA: Sage), 55–87. doi: 10.4135/9781452272054.n3

Kaplan, S. (1995). The restorative benefits of nature: toward an integrative framework. *J. Environ. Psychol.* 15, 169–182. doi: 10.1016/0272-4944(95)90001-2

Mace, B. L., Bell, P. A., and Loomis, R. J. (1999). Aesthetic, affective, and cognitive effects of noise on natural landscape assessment. *Soc. Nat. Resour.* 12, 225–242. doi: 10.1080/089419299279713

Mace, B. L., Bell, P. A., Loomis, R. J., and Haas, G. E. (2003). Source attribution of helicopter noise in pristine national park landscapes. *J. Park Recreat. Admi.* 21, 97–119.

Miedema, H. M. E., and Vos, H. (2003). Noise sensitivity and reactions to noise and other environmental conditions. *J. Acoust. Soc. Am.* 113, 1492–1504. doi: 10.1121/1.1547437

National Parks Air Tour Management Act of 2000 (2000). *National Parks Air Tour Management Act of 2000. SEC. 801 [49 U.S.C. 40128].*

National Wilderness Preservation Act of 1964 (1964). *National Wilderness Preservation Act, 40 U.S.C. §573.*

Noise Control Act of 1972 (1972). *Noise Control Act of 1972.*SEC. 2 [42 U.S.C. 4901].

Price, D. D., McGrath, P. A., Rafii, A., and Buckingham, B. (1983). The validation of visual analog scales as ratio scale measures for chronic and experimental pain. *Pain* 17, 45–56. doi: 10.1016/0304-3959(83)90126-4

Rankin, M., and Borah, G. (2003). Perceived functional impact of abnormal facial appearance. *Plast. Reconstr. Surg.* 111, 2140–2146. doi: 10.1097/01.prs.0000060105.63335.0c

Seidman, M. D., and Standring, R. T. (2010). Noise and quality of life. *Int. J. Environ. Res. Public Health* 7, 3730–3738.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *J. Educ. Behav. Stat.* 23, 323–355. doi: 10.3102/10769986023004323

Singer, J. D., and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford: Oxford University Press.

Steffens, J., Steele, D., and Guastavino, C. (2017). Situational and person-related factors influencing momentary and retrospective soundscape evaluations in day-to-day life. *J. Acoust. Soc. Am.* 141, 1414–1425. doi: 10.1121/1.4976627

Tarrant, M. A., Haas, G. E., and Manfredo, M. J. (1995). Factors affecting visitor evaluations of aircraft overflights of wilderness. *Soc. Nat. Resour.* 8, 351–360. doi: 10.1080/08941929509380927

Watson, D., and Clark, L. A. (1999). *The PANAS-X: Manual for the Positive and Negative Affect Schedule-Expanded Form*. Ames, IA: The University of Iowa.

Weinstein, N. (1978). Individual differences in reactions to noise: a longitudinal study in a college dormitory. *J. App. Psychol.* 63, 458–466. doi: 10.1037/0021-9010.63.4.458

Yellowstone Act of 1872 (1872). *Yellowstone Act, 17 U.S.C. §32.*

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 23 October 2020
doi: 10.3389/fpsyg.2020.570761

Modeling Evaluations of Low-Level Sounds in Everyday Situations Using Linear Machine Learning for Variable Selection

Siegbert Versümer*, Jochen Steffens, Patrick Blättermann and Jörg Becker-Schweitzer

Institute of Sound and Vibration Engineering, University of Applied Sciences Düsseldorf, Düsseldorf, Germany

Human sound evaluations not only depend on the characteristics of the sound but are also driven by factors related to the listener and the situation. Our research aimed to investigate crucial factors influencing the perception of low-level sounds as they—in addition to the often-researched loud-level sounds—might be decisive to people's quality of life and health. We conducted an online study in which 1,301 participants reported on up to three everyday situations in which they perceived low-level sounds, resulting in a total of 2,800 listening situations. Participants rated the sounds' perceived loudness, timbre, and tonality. Additionally, they described the listening situations employing situational eight dimensions and reported their affective states. All sounds were then assigned to the categories natural, human, and technical. Linear models suggest a significant difference of annoyance ratings across sound categories for binary loudness levels. The ability to mentally fade-out sound was the most crucial situational variable after valence, arousal, and the situation dimensions positivity and negativity. We ultimately selected the most important factors from a large number of independent variables by applying the percentile least absolute shrinkage and selection operator (Lasso) regularization method. The resulting linear regression showed that this novel machine-learning variable-selection technique is applicable in hypothesis testing of noise effects and soundscape research. The typical problems of overfitting and multicollinearity that occur when many situational and personal variables are involved were overcome. This study provides an extensive database of evaluated everyday sounds and listening situations, offering an enormous test power. Our machine learning approach, whose application leads to comprehensive models for the prediction of sound perception, is available for future study designs aiming to model sound perception and evaluation.

Keywords: machine learning, variable selection, human perception, situation, Lasso, environmental sound, online-survey

OPEN ACCESS

Edited by:
Bert De Coensel,
Ghent University, Belgium

Reviewed by:
Catherine Lavandier,
Université de Cergy-Pontoise, France
PerMagnus Lindborg,
School of Creative Media, City University of Hong Kong, Hong Kong

****Correspondence:***
Siegbert Versümer
s.versuemer@hs-duesseldorf.de

Specialty section:
This article was submitted to Environmental Psychology, a section of the journal Frontiers in Psychology

Received: *08 June 2020*
Accepted: *07 September 2020*
Published: *23 October 2020*

Citation:
Versümer S, Steffens J, Blättermann P and Becker-Schweitzer J (2020) Modeling Evaluations of Low-Level Sounds in Everyday Situations Using Linear Machine Learning for Variable Selection. Front. Psychol. 11:570761. doi: 10.3389/fpsyg.2020.570761

INTRODUCTION

Myriad research has shown that annoyance reactions to unpleasant sounds can cause psychological stress (Gunn et al., 1975; Wolsink et al., 1993; Lercher, 1996; Stallen, 1999) that consequently affects cognition and health (Serrou, 1995; Babisch, 2002; World Health Organization, 2011; Beutel et al., 2016; Klatte et al., 2017). While the majority of studies have focused on the perception, evaluation, and effects of medium or loud sounds generated by road traffic (Aletta et al., 2018; Riedel et al., 2019) and aircraft noise (Kroesen et al., 2008; Schreckenberg et al., 2016), annoyance has also been found in response to low-level sounds—for example, noise from wind turbines (Wolsink et al., 1993; Crichton et al., 2015; Klatte et al., 2017; van Renterghem, 2019). Since research to date in the field of wind turbine noise has focused on low frequencies, we aimed to investigate the evaluation of low-level day-to-day sounds in general and to establish comprehensive models including situational, sound-related, and person-related factors to predict the perception of environmental sounds in both low- and mid/high-level scenarios. Moreover, we investigated which influencing factors had a substantial impact on the evaluation of low-level sounds when taking into account multiple variables. To address these research aims, we conducted an online study wherein 1,301 participants reported on up to three everyday situations (including 32 relevant sound-related, situational, and person-related variables) in which they perceived low-level sounds. To handle this large number of variables, we implemented the percentile least absolute shrinkage and selection operator (Lasso) method, a linear machine learning approach, to select the crucial variables associated with annoyance ratings and to establish comprehensive models which overcome problems associated with overfitting and can predict annoyance for new data that were not involved in the model training and validation.

Previous research on soundscape perception and reactions to noise has identified several influencing factors related to sound, situation, and perception. Besides exposure level (Wolsink et al., 1993; Basner et al., 2014; Guski et al., 2017), these factors include many non-auditory variables, such as sensitivity to noise (Fields, 1993; Job, 1996; Schreckenberg et al., 2010; Hill et al., 2014; Shepherd et al., 2015; Park et al., 2016; Kim et al., 2017), extraversion and neuroticism with contradictory evidence for the relevance of this factor (Lercher, 1996), attitude toward the source or the authorities that operate the sound source (Stallen, 1999; Job et al., 2007; Kroesen et al., 2008), perceived disturbance (Stallen, 1999; Kroesen et al., 2008), fear of the noise source (Miedema and Vos, 1999), and the failure to cope with the environment which leads to stress (Guski, 1999). Many objective situational variables, including the presence of other people, the location of the perceiver, the sound insulation of dwellings, the visibility of the source, economic benefit through the source, exposure time, or ambient noise level (Fields, 1993; Wolsink et al., 1993; Bangjun et al., 2003; Janssen et al., 2011; Steffens et al., 2017) have also been identified as relevant factors.

Psychological Situations and Situational Characteristics

Situations can be seen as "fluctuating, dynamic, and dependent upon different perspectives" (Rauthmann, 2015, p. 177). Since situational factors are known to be essential predictors of human perception and behavior, they have been the subject of many studies. Nevertheless, these factors, interpreted as "situational," have mostly been physical, objective, easily measurable, and (in a laboratory setting) controllable quantities: exposure time, noise insulation of dwellings, and ambient noise (Fields, 1993); age, benefit, and visibility of the source (Bangjun et al., 2003; Janssen et al., 2011); or exposure level, buildings, trees, and fences (Wolsink et al., 1993). Situations may be defined by the actual objective environment (*E*) and the momentary mental and affective state of the person (*P*) perceiving the specific situation (*S*). Lewin (1936) described a person's behavioral states (*B*) driven by a function of the perception of that situation as $B = f(P, E) = f(S)$. Following this theory, situations can be split up into cues, characteristics, and classes (Rauthmann et al., 2015). The objective physical quantities mentioned above can be seen as the situational cues from which people derive situational characteristics and psychological meaning during the evaluation processes. Finally, situational classes group situations that have similar characteristics or cues.

This view of situations is in line with the model of the "cognitive–motivational–emotive system" discussed by Smith and Lazarus (1990, p. 622). In that model, objective conditions—the cues—are individually interpreted by the person through imprinting his or her personality, including individual needs, commitments, goals, knowledge, attitudes, and beliefs. The resulting subjective situational construal—the characteristics—ultimately serves as the basis for the subsequent appraisal processes that mediate a person's emotional response. For example, imagine a bike path parallel to a highly frequented 8-lane road surrounded by tall trees in full leaf. Cyclists who were highly skeptical of the greenery's capability to attenuate traffic noise, improve air quality, or enhance health reported lower soundscape quality (Aletta et al., 2018).

The importance of taking psychological and situational characteristics into account is evident, as they reflect situational social aspects and people's cognitive and emotional perceptions of their environments. To propose a taxonomy for measuring and describing psychological situations, Rauthmann et al. (2015) developed the DIAMONDS model through measuring individual differences in situation perception. This model consists of eight situational dimensions: "*Duty* (does something need to be done?), *Intellect* (is deep information processing required?), *Adversity* (is someone being overtly threatened?), *Mating* (is the situation sexually and/or romantically charged?), *pOsitivity* (is the situation pleasant?), *Negativity* (do negative things taint the situation?), *Deception* (is someone deceptive?) and *Sociality* (is social interaction and relationship formation possible, desired, or necessary?)" (Rauthmann et al., 2015, p. 364). The DIAMONDS model follows the principle of personality research that "individuals [may] think about situational characteristics in much the same way they think about personal characteristics" (Halevy et al., 2019, p. 4). Interestingly, to the best of our knowledge, such a model has not yet been used to investigate sound evaluation in terms of differences in individual situation perception. Therefore, we included the assessment of psychological situations in our study and hypothesized that psychological situation characteristics would significantly be associated with annoyance ratings of environmental sounds.

Stress and Its Precursors as Pivotal Points in Human Perception

In addition to the psychological situations that might play an essential role in human perception of sound, perceived control was assumed to be "the most important non-acoustical determinant of environmental noise annoyance" in the stress–annoyance model developed by Stallen (1999, p. 77), which interpreted annoyance as stress. He hypothesized that annoyance is driven by three main factors: perceived disturbance, perceived control, and coping with annoyance.

The first factor, perceived disturbance, depends on the sound of the sources and an initial cognitive–emotive appraisal process (Lazarus, 1966; Stallen, 1999). Disturbance occurs when people are hindered from achieving their goals (e.g., concentration, relaxation, sleep, communication). It is linked to annoyance both directly (Stallen, 1999; Kroesen et al., 2008) and indirectly through a mediated path via coping strategies (Park et al., 2016). Disturbance is also influenced by the personality trait noise sensitivity (Park et al., 2016).

The second factor, perceived control, is not associated with disturbance or noise (Stallen, 1999). Instead, it has been hypothesized that perceived control is driven by the noise management of the source—not the source itself—and that it directly affects annoyance through a secondary path (Lazarus, 1966; Stallen, 1999; Park et al., 2016). Kroesen et al. (2008) followed the approach of Stallen's stress model in investigating annoyance induced by aircraft noise (Stallen, 1999). Perceived control and coping capacity were together shown to be the most important variables after concerns about adverse health effects and perceived disturbance.

Coping can be defined as "constantly changing cognitive and behavioral efforts to manage specific external and/or internal demands that are appraised as taxing or exceeding the resources of the person" (Lazarus and Folkman, 1984, p. 141). Coping is driven by "the belief and confidence of an affected person that he/she will somehow manage the problem" (Guski, 1999, p. 51). Coping with stress in general or annoyance in particular can be seen as a reappraisal of a person's environment (Gunn et al., 1975; Smith and Lazarus, 1990; Stallen, 1999). Botteldooren and Lercher (2004), as well as Glass et al. (1972), assumed that annoyance is a prerequisite for coping. Park et al. (2016) reported an additional mediation effect of coping on the relationship between disturbance and annoyance that was not present in the model by Kroesen et al. (2008).

Personality Traits and Demographic Factors

In contrast to the aforementioned dynamic situational factors, stable personality traits change little in adulthood. The Big Five dimensions of personality, for example, were derived through a lexical approach, meaning that all relevant aspects of personality will develop and be found in the language of a community: *Neuroticism*, *Extraversion*, *Openness to experience*, *Agreeableness*, and *Conscientiousness* have been consistently reported to be important and sufficient descriptors of human personality (for an overview, see Digman, 1990; Costa and McCrae, 2008). Though *Extraversion* and *Neuroticism* (as well as all demographic variables) have often been discussed in relation to human sound evaluation, the results have been controversial (see the review by Fields, 1993). Even when all Big Five dimensions are considered together, a recent study by Lindborg and Friberg (2016) showed only small, albeit significant, effects.

Noise sensitivity seems to play an essential role in moderating or mediating the effect of sound on annoyance (Miedema and Vos, 2003; van Kamp et al., 2004) and health (Job, 1996). Shepherd et al. (2015) analyzed the effect of (other) personality traits on sensitivity to noise and revealed that extraversion acted as a major predictor. In their study, all Big Five dimensions showed linear and independent effects on noise sensitivity and together accounted for 33% of variance. Similarly, Lindborg and Friberg (2016) reported that noise sensitivity can be predicted by extraversion and conscientiousness. Belojević and Jakovjlevic (2001) also investigated factors influencing sensitivity to noise and found that neuroticism was the only significant person-related factor in noisy environments but had no significant effect in quiet areas. Since noise sensitivity plays a vital role in sound perception and since extraversion and neuroticism may influence noise sensitivity, extending existing findings by investigating these variables for low-level sounds seems worthwhile.

Demographic variables have often been investigated in noise annoyance research, showing only a small or generally insignificant effect (Yu and Kang, 2008). Miedema and Vos (1999) reported that people between 20 and 70 years of age showed higher annoyance compared to younger or older people. Gender was not significant, but education level showed a small effect of increased annoyance with increasing years of education. The hypothesis that people with higher education, and thus higher income, experience less annoyance by seeking less noisy living environments seems to apply only to residents of small or medium-sized cities, with income not significantly moderating annoyance (Fyhri and Klćboe, 2006).

Aims and Hypotheses

Many of the studies mentioned above have focused on a small number of variables associated with sound evaluations and annoyance reactions. Our study, in contrast, combined a high number of relevant sound-related, situational, and person-related variables in a comprehensive model to predict low- and mid/high-level sounds in everyday life. We therefore attempted to identify the most relevant predictors. Based on previous research, we assumed that situational variables, as opposed to person-related factors, would have higher explanatory potential in predicting annoyance ratings of both low- and mid/high-level sounds. We further hypothesized that the category of a sound (natural vs. technical vs. human) would play a decisive role in evaluating environmental sounds. We included demographic factors to investigate the extent to which previous results are reproducible in a retrospective online study. Finally, we explored which low-level sounds participants perceived as particularly pleasant or annoying and how often these sounds occurred in day-to-day life.

MATERIALS AND METHODS

Participants

Initially, we defined 18 quotas of 100 participants each. The quotas were established by combining three *Age Class*es (20–40, 41–60, and 61–80 years) with two *Gender*s (*female* and *male*) and three *Education Level*s (International Standard Classification of Education (ISCED) levels 0–2 (up to lower secondary education), level 3 (upper secondary education), and levels 4–8 (university-level education); United Nations Educational, Scientific and Cultural Organization (UNESCO) UNESCO Institute for Statistics, 2015. For the application to the German education system (see Schneider, 2008). We commissioned the Cologne-based commercial market research company respondi AG[1] to provide suitable participants from its online panel according to our quota targets defined above.

Of the 12,000 persons invited by email, 4,087 started the online questionnaire; 1,815 (54%) completed the questionnaire and reported and evaluated 5,445 sound situations. Of these, 514 (28%) reported no sound situations or gave implausible answers. Consequently, 2,645 datasets were excluded from the evaluation. We ultimately analyzed the data of 1,301 participants (630 men, mean age = 49.9, *SD* = 15.5; 671 women, mean age = 49.6, *SD* = 16.4), who reported on 2,800 sound situations. *Education Level* and *Gender* were evenly distributed (men: 223 level 3, 185 below, 222 above; women: 228 level 3, 196 below, 247 above). In addition to German subjects, the sample also included a few respondents of non-German *Nationality* (12 men; 13 women). Participants with any type of *Hearing Impairment* (n = 213; 16.4%) were included in the final dataset since the mean *Annoyance* ratings of all their reports did not differ significantly from the ratings by persons without a known hearing disability [Wilcoxon (W) = 530,871; p = 0.749; calculated on raw data]. Additionally, their *Noise Sensitivity* (M = 15.16; *SD* = 3.76; calculated on raw data) did not differ significantly (W = 107,813; p = 0.107) from participants without a known *Hearing Impairment* (M = 14.70; *SD* = 4.13).

Procedure

To address our research topics regarding the occurrence of low-level environmental sounds in everyday life and person-related and situational factors influencing their evaluations, we conducted a large-scale online study using the LimeSurvey[2] software (see the questionnaire in the **Supplementary Material**). After reporting on sociodemographic and person-related variables, participants described and evaluated up to three sound situations they had experienced in the past, with no acoustic stimuli provided by us. (If a participant reported less than three sound situations, we provided one to three preset sound situations, so that each participant had three to evaluate. The situations we added were not taken into account in this analysis.) We let the participants decide how they understood "quiet" or "low-level" (in the sense of "not loud"). We further used the term "sound" to avoid bias toward negatively perceived sounds classified as "noise." After finishing the questionnaire, the participants were automatically redirected to the panel operator respondi AG to receive monetary compensation for their participation.

Design and Questionnaire

In our online study, we asked the participants to remember and evaluate low-level sounds they had heard. Thus, we focused on sound immission and not on sound emission. Since perceived sound level decreases with increased distance from the source and can be changed in terms of frequency components, we believe that evaluating sound sources from a greater distance—e.g., through closed windows—will lead to biased, experience-driven judgments. The questionnaire we used in our study is provided as **Supplementary Material**.

Person-Related Variables

Besides the sociodemographic variables (i.e., *Age*, *Gender*, *Education Level, and Nationality*), participants reported other temporal stable variables, such as whether they were aware of having any *Hearing Impairment*. They further rated their living environment by answering the question, "How would you describe your living environment?" using the five-level bipolar item *Liveliness*, ranging from *very lively* (1) to *very calm* (5). They also reported on the number of *Persons* living in the household (*1* to "*6 or more*") and their household's monthly disposable *Net Income* (German Federal Statistical Office, 2018). Moreover, since the person-related factors *Extraversion*, *Neuroticism*, and *Noise Sensitivity* have shown associations with sound evaluations in previous studies (Fields, 1993; Job, 1996; Lercher, 1996; Schreckenberg et al., 2010; Hill et al., 2014; Shepherd et al., 2015; Kim et al., 2017), we obtained those factors using the German 10 Item Big Five Inventory (BFI-10; Rammstedt and John, 2007; Rammstedt et al., 2012) and a nine-item *Noise Sensitivity* questionnaire ("Kurzform des Fragebogens zur Lärmempfindlichkeit," LEF-K, developed by Zimmer and Ellermeier, 1998). Participants also answered the question "Are you generally able to mentally fade out sounds (even loud ones)?" using a five-level scale ranging from *does not apply at all* (0) to *is absolutely right* (4) for the item *General Fade-out*. To group the reports for each participant (see section "Statistical Analyses" for the random effect), we assigned a unique *ID* to each participant.

Sound-Related Variables

In addition to the temporally stable person-related variables, the following variables are assumed to change over time depending on the sound and its embedding situation. Participants responded to the item "Please remember sounds that you have classified as low-level in your environment in the past." by reporting sounds in free-form text descriptions. They also rated the perceived *Loudness* of their sounds ("How do you rate the sound?") on a five-level scale ranging from *scarcely audible* (1) to *low-level* (3) to *middle and louder* (5). The *Loudness* levels 4 and 5 were intended to check whether participants had indicated a low-level sound. The *Timbre* of the sound was assessed on a five-level bipolar scale ranging from *deep, dull* (1; German: "tief, dumpf") to *high, shrill* (5; German: "hoch, schrill") as well as the item *Tonality* based on

[1]https://www.respondi.com/

[2]https://www.limesurvey.org/

the levels *broadband noise* (1; German: “rauschartig”) to *tonal* (5; German: “tonhaltig”).

They also used a German translation of the standard soundscape dimensions by responding to the question “Please indicate how much you consider the following characteristics to be a description of the sound.” These eight dimensions—namely *Pleasant* (“angenehm”), *Vibrant/Exciting* (“lebendig/pulsierend”), *Eventful* (“ereignisreich”), *Chaotic* (“chaotisch”), *Annoying/Distracting* (“lästig/störend”), *Monotonous* (“monoton”), *Uneventful* (“ereignislos”), and *Calm* (“ruhig”)—are measured on Likert scales ranging from *strongly agree* (1) to *strongly disagree* (5) and can be arranged in a circumplex model of soundscape perception (Axelsson et al., 2010; Lindborg and Friberg, 2016; following ISO/TS 12913-2, ISO, 2018). To obtain the dependent variable *Annoyance* that was relevant for our analyses, we computed the arithmetic mean across the ratings of *Pleasant* (inversed) and *Annoying*.

Situational Variables

Concerning the time-varying situational variables, participants responded to “For each sound, please mention a situation in which you have experienced this sound.” using free-form text descriptions. Participants’ affective state (“Please assess how you feel in this sound situation.”) was obtained in terms of *Valence* (*negative-positive*), *Arousal* (*calm-excited*), and the perceived *Control* over the sound situation (*weak-strong*). Here, we used the Self-Assessment Manikin (SAM), which consists of three sets of nine pictograms (see the questionnaire in **Supplementary Material**) representing the different states of the three affective dimensions (Lang, 1980; Bradley and Lang, 1994; PXLab: Irtel, 2007). The SAM has been shown to be quickly and consistently answerable by people of various nationalities and languages, by adults and children, and by people with language disorders (Bradley and Lang, 1994; Bynion and Feldner, 2017). It has also been demonstrated to be applicable to the evaluation of acoustic stimuli (IADS: Bradley and Lang, 2000) and therefore seemed suitable for our study. The use of the SAM can activate responses in any part of the emotional system, like physiological, behavioral, and emotional (Suk, 2006). Thus, the SAM seems to be a more profound measurement method than written scales, which must be processed via cognition. The pictograms can be modified or replaced with signs to achieve similar results (Affective Slider: Betella and Verschure, 2016). In many studies, the original SAM was adapted in terms of number of levels, number of pictograms, and manipulation of the pictograms (Bynion and Feldner, 2017; Bartosova et al., 2019). The semiotics of the pictograms and signs, although self-explanatory, are usually explained at the beginning of a test (Lang and Bradley, 1997; Suk, 2006). Võ et al. (2009) used the SAM to avoid the German translation for arousal (“Erregung”), which could have sexual associations. Since the SAM pictograms can be used for many attributes other than valence, arousal, and dominance (Suk, 2006), we believed that additional descriptions of the three affective variables were necessary. Because slightly modified words were successfully used in most studies, we added the adjectives given above to clarify the two anchors of each of these scales.

Participants further responded to the question “Can you mentally fade out this sound?” for the *Specific Fade-out* variable using a five-level Likert scale ranging from *does not apply at all* (0) to *does fully apply* (4). To assess participants’ *Active Coping* response to the sound situation, we asked “Suppose you feel disturbed by the sound 1 in situation a[3]. Would you take action to reduce the disturbing effect?” to which respondents answered *yes* or *no*. For a more detailed description of the situation and its psychological characteristics, we utilized an ultra-brief German measure of the situational eight-factor DIAMONDS model (S8-II; Rauthmann, 2018). Finally, participants reported the *Frequency* of occurrence of the described situation using six levels: *less than once a year* (0); *once to four times per year* (1); *five to 11 times per year* (2); *once to three times monthly* (3); *once to three times weekly* (4); *four to seven times weekly* (5); and *more than once a day* (6).

Data Analysis

Data Preparation

We first analyzed all sound descriptions and classified them into the three *macro-level sound categories* of *natural*, *human*, and *technical* sounds that have already been applied in previous studies (Axelsson et al., 2010; Bones et al., 2018) as well as the soundscape standard (ISO 12913-2, ISO, 2018).

We further established 38 *micro-level sound categories* (see **Figure 3**) in the course of a more detailed qualitative analysis. This categorization was mainly carried out using two processing loops. In the outer loop, an audio expert looked through the sound descriptions and searched for an often-mentioned sound or word. A category was then established for the sounds described by that word. This definition was based on the knowledge of sound properties, sound sources, and theories of sound perception. For example, the two categories *Dogs and insects with possible threats* and *Dogs, insects, and other animals without possible threats* were created, as an individual’s attitude to the sound source can change the perception of sound; e.g., a fly might be less annoying than a mosquito because one expects a possible painful mosquito bite. Another example was the *Signals* category, which includes all types of signals—such as ring tones, alarm clocks, and doorbells—that have a concrete meaning for the participant and urge the participant to take action. Some sounds have been combined, such as sounds caused by garbage collection and construction site noises, since participants usually have no direct influence on these sound sources. As a result, reduced perceived control and limited coping may emphasize annoyance.

In the inner loop, the data was then filtered by this word or iteratively for a part of the word (e.g., by omitting the word ending). The word was also modified or replaced by a synonym. If the context derived from the sound and situation descriptions of each filtered sound situation matched the noise category, all these reports were assigned to the selected category and excluded from further handling. By this exclusion, the number of remaining reports was reduced successively. If no further observation could be assigned to this category, the process was resumed

[3]Instead of “1” and “a,” the sound and situation descriptions given by the participants were inserted here.

TABLE 1 | Categories used for the variables *Location* and *Activity*.

Description	Variable	N_{obs}
Location		
At home, indoors	*Home (indoors)*	1,659
At home outdoors, incl. garden, and nature	*Garden/Nature*	463
Undefined	*Undefined*	259
Other, indoors	*Other (indoors)*	212
At work/office	*Work/Office*	129
Other (outdoors)	*Other (outdoors)*	78
Activity		
Undefined		978
Relaxing, falling asleep, awakening		864
Being on the move, transportation		331
Working, studying, cognitive work		220
Entertainment (TV, radio, movie, theater, gaming, internet surfing)		116
Housework		88
Social activities		74
Taking a meal		50
Personal hygiene		30
Exercise, sport, leisure activities, hobbies		17
Making a call		17
Pure music listening and entertainment (TV, books/news reading)		8
Coping with emotions and stress		7

N_{obs} *Number of observations in the mentioned category.*

with the outer loop and the next category was defined. These loops were repeated until all sound situations were categorized. Accordingly, there were numerous categories and no "undefined" group. Datasets which included responses with nonsensical terms (e.g., "fff") or in a language other than German were excluded from the evaluation.

We also derived categories for the *Location* where participants experienced the described situations (see **Table 1**). In addition, we applied the *Activity* categories introduced by Greb et al. (2018), which we adapted slightly for our data, as shown in **Table 1**. Namely, we removed *Making music*, added *Making a call*, and assigned new activities to the existing categories when a similar evaluation distribution existed. Of the 13 *Activity* categories, the category *Undefined* was the largest due to 978 situation descriptions that contained no information about activities. We therefore excluded the *Activities* from further detailed analysis.

To test the inter-rater reliability, a second rater assigned the sound situation descriptions to the *micro-level sound categories*. For 190 descriptions, none of these categories seemed reasonable. Again, reports with nonsensical sound and situation descriptions were marked for removal. The point estimate of Krippendorff's alpha of 0.782 with bootstrapped 95% confidence intervals (CIs) [0.767, 0.796][4] were obtained by bootstrapping 1,000 samples (Zapf et al., 2016). This reliability lies between α = 0.667 and α = 0.800, which is why the *micro-level sound categories* should only be used for "drawing tentative conclusions" (Krippendorff, 2004, p. 241). According to Krippendorff, an α above 0.800 is considered reliable, which we observed for the *macro-level* sound categories (α = 0.877, CI [0.863, 0.891][5]). Values for the *Locations* (α = 0.625, CI [0.599, 0.650][6]) are below that threshold, which allows only tentative conclusions.

As we were particularly interested in the perception and evaluation of *low-level* sounds, we grouped all datasets according to their *Loudness* rating as *low-level* (*Loudness* levels [1–3]) or *mid/high-level* ([4–5]). The person-related variable describing the perceived *Liveliness* of the living environment was grouped into the category *calm* or *lively* through a median split ($Mdn_{\text{environment}} = 4$). Finally, we assigned to each participant the mean value of the reported *Net Income* interval and combined two variables—number of *Persons* living in the household and monthly disposable household *Net Income*—to form a new variable: net *Income per Person*.

Statistical Analyses

All statistical analyses were performed with R 3.6.3 (R Core Team, 2020) and R-Studio (RStudio Team, 2019). To predict *Annoyance* assessments by person-related, situational, and sound-related variables, we calculated several hierarchical linear mixed-effect models, as such models can handle non-normally distributed data and take into account dependencies of the three observations (level 1) within the participants (level 2) while allowing for the inclusion of time-varying (i.e., situation-related) predictors. The participants, represented by their grouping *ID*, were included as a random factor in all models we used in this paper (see **Table 2** for an overview of all models).

To calculate these models, we used two different approaches. First, we used the *lme4* (Bates et al., 2015) and *performance*[7] R packages to calculate the marginal and conditional coefficients of determination (R^2_m and R^2_c) as effect size measures (models *32SFF/32SFFA*, *Age*Edu*, and *CMSFF1/2/3*). R^2_m addresses the variance of *Annoyance* that is explained by fixed factors, whereas R^2_c represents the variance that is explained by both fixed and random factors (Nakagawa and Schielzeth, 2013). We derived probability values for each implemented variable and factor-level dummy using the *lmerTest* R package (Kuznetsova et al., 2017). To assess the influence of each variable of interest (see subsections of section "Design and Questionnaire") on *Annoyance*, we built one single-fixed-factor model for each variable and dummy (**Table 4**, model *32SFF*). We also derived one probability value for each variable (or, in the case of a factor, including the dummies of a factor) using ANOVA (**Figure 5**, model *32SFFA*). Several publications have shown that ANOVA may be successfully applied to non-normally distributed data (Glass et al., 1972; Harwell et al., 1992; Lix et al., 1996). The probability values

[4] *N* (number of subjects with two or more ratings) = 3,039; *n* (number of ratings) = 2; *k* (number of categories) = 38 + 2.

[5] *N* (number of subjects with two or more ratings) = 3,039; *n* (number of ratings) = 2; *k* (number of categories) = 3 + 2; *B* (number of bootstrap samples) = 1,000.

[6] *N* (number of subjects with two or more ratings) = 2,778; *n* (number of ratings) = 2; *k* (number of categories) = 6; *B* (number of bootstrap samples) = 1,000. The *N* here is smaller than the *N* for the sound categories because locations have not been clustered for nonsensical reports.

[7] https://easystats.github.io/performance/

TABLE 2 | Overview about the Models used in this contribution.

Model	Fixed factors	Data	Presented information	Used in
Macro1	*Macro-level sound categories*	All data	Estimated marginal means and CI	Sections "Statistical Analyses" and "Sound Categories: Macro-Level"
Macro2	*Macro-level sound categories* + *Binary loudness* levels			Section "Statistical Analyses" **Figure 2**
Micro1	*Micro-level sound categories*	All data	Estimated marginal means and CI	Sections "Statistical Analyses" and "Influence of Micro-Level Sound Categories"
Micro2	*Micro-level sound categories* + *Binary loudness*			Section "Statistical Analyses" **Figure 3**
Micro3a	*Micro-level sound categories*	*Low-level* subset		**Figure A1** in the Annex
Micro3b	*Micro-level sound categories*	*Mid/high-level* subset		
Location	*Location* + binary *Loudness*	All data	Estimated marginal means and CI	Sections "Statistical Analyses" and "Influence of Location" **Figure 4**
Liveliness	*Liveliness* + binary *Loudness*	All data	Estimated marginal means and CI	Sections "Statistical Analyses" and "Living Environment" **Table 3**
32SFF	32 single-fixed-factor models	All data	β, CI, *p*, df, $R^2{}_m$, $R^2{}_c$	Sections "Statistical Analyses" and "Single-Fixed-Factor Models" **Table 4**
32SFFA	32 single-fixed-factor models (ANOVA)		β, $R^2{}_m$	Section "Statistical Analyses" and "Single-Fixed-Factor Models" **Figure 5**
*Age*Edu*	*Age * Education*	All data	β, CI, *p*, df, $R^2{}_m$, $R^2{}_c$	Sections "Statistical Analyses" and "Role of Person-Related Factors" **Table 5**
CML1 *CML2* *CML3*	LASSO selected variables	All data *Low-level* subset *Mid/high-level* subset	β, CI, *p*, df for all LASSO selected variables and loudness subsets	Sections "Percentile Lasso Regression Parameter Selection Method" and "Comprehensive Models" **Table 6**
CMSFF1 *CMSFF2* *CMSFF3*	Relevant variables from single-fixed-factor models	All data *Low-level* subset *Mid/high-level* subset	β, CI, *p*, df, $R^2{}_m$, $R^2{}_c$	Sections "Statistical Analyses" and "Comprehensive Models" **Table 7**

SFF singe-fixed-factor; CM comprehensive model.

were calculated using Satterthwaite's approximation of degrees of freedom. This approximation combined with restricted maximum likelihood estimation produces "the most consistent Type 1 error rates, being neither anti-conservative nor overly sensitive to sample size" (Luke, 2017, p. 1500).

Second, we used bootstrapping—drawing 50,000 samples—with the *clusterBootstrap* R package (Deen and de Rooij, 2020) to calculate the marginal means of *Annoyance*, including the 95% CIs for non-normally distributed data (models *Macro1/2, Micro1/2/3a/b*, *Location*, and *Liveliness*). This method uses linear models, is relatively free of assumptions, and is particularly well suited for hierarchical data. Non-normally distributed data were considered by resampling the observations at the individual level (within persons). This means that if one observation of a person was selected randomly, all other observations of this person were also included in the calculation (which is also the case for the Lasso regression method described in the next section).

We used β as a standardized regression coefficient only for independent variables that did not represent a physical quantity (such as age, income, and persons in the household). The factor levels represented by their dummy variables were also standardized. The dependent variable *Annoyance* was not standardized for a more intuitive interpretation.

When calculating CIs or probabilities, we accepted the inflation of Type I errors because applying a correction to the confidence levels and *p*-values for all models used in this paper would have resulted in many different confidence levels, complicating the interpretation. Additionally, reducing the family-related error rate using a correction method would have increased the probability of Type II errors and reduced the validity of the test. More importantly, the discussion of which correction should be used and how the family might be defined would be beyond the scope of this paper, as this is a very controversial topic in the research community (Rothman, 1990; Perneger, 1998; Bender and Lange, 2001). The same applies to the general use of probabilities in the context of linear mixed-effects regression models (Luke, 2017). Thus, all the probabilities and CIs given here should be interpreted in this context and should not be seen as a hard cut-off condition.

Percentile lasso regression parameter selection method

One of our aims was to establish a comprehensive model predicting the perceived *Annoyance* of *low-level* sounds by utilizing the most essential sound-related, situational, and person-related factors. To this end, we followed the work of Greb et al. (2018) and used the percentile-Lasso regression method (Roberts and Nowak, 2014) for multilevel linear regression modeling based on the measured data (models CML1/2/3). The Lasso method was first described by Tibshirani (1996) and has become a popular shrinkage method in the field of statistical learning algorithms. It adds a ℓ^1 regularization term with a tuning parameter λ to linear regression models that controls the amount of shrinkage applied to the regression coefficients. Choosing a high λ value potentially sets all coefficients to zero, while λ = 0 results in a linear regression model without penalty (see **Figure 1**). This form of regularization can thus be used to extract important features from the data and reduce overfitting by excluding less important predictors from the model and therefore lowering its complexity. To achieve these advantages, the optimal compromise between retaining all contributing factors in the model (λ = 0) and excluding all variables (at $\lambda_{\max}$) must be found. Therefore, the loss function, minimized within the Lasso, is defined in Eq. 1.

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\left(y_i - \sum_{j=0}^{p}\beta_j \times x_{ij}\right)^2 + \lambda\sum_{j=0}^{p}|\beta_j| \qquad (1)$$

We chose five-fold cross-validation to find the optimal λ value that results in a parsimonious and generalized model with small prediction error. The technique of *K*-fold cross-validation involves randomly splitting the data into *K* nearly equally sized folds and using *K*−1 folds as training data. The remaining fold is used for validating the previously estimated statistical model and calculating the mean squared error (MSE) of the prediction on unseen data that was not involved in the training. This routine is repeated *K* times until every fold has been used as a validation set, resulting in a cross-validation error (CV) as the mean MSE calculated from all *K* repetitions. Our decision that *K* = 5 resulted from the number of observations—considering computational costs and a sensible amount of data in the validation sets—since research has noted that 5- and 10-fold cross-validation can be viewed as equally efficient with regards to the bias–variance tradeoff (Krstajic et al., 2014). The random fold assignment was—respecting the two-level structure of the measured data—based on the level of the participants to assure that all measurements of one participant were assigned to either the training or validation set for all repeatitions within the cross-validation.

A set of 100 λ values (grid) was used to build and validate models within every cross-validation cycle. As suggested by Greb et al. (2018), the grid had an exponential form to achieve a higher resolution of values toward zero. The value $\lambda_{\max}$ was determined in advance by successively increasing λ until all regression coefficients were set to zero. As proposed by Hastie et al. (2009), the 1-SE[8] rule was then applied to calculate the optimal λ value for every cross-validation cycle and to choose the most parsimonious model whose MSE was within one SE of the minimum cross-validation error.

To overcome the sensitivity of finding the optimal λ value to the cross-validation fold assignment (Krstajic et al., 2014), we repeated the process of cross-validation 100 times and selected the 95th percentile as the optimal λ value for the final fit. As reported by Roberts and Nowak (2014), the 95th percentile produces good and reliable results.

We used the *glmmLasso* R package (Groll, 2017) to implement the percentile-Lasso regression method. The package allowed us to calculate the generalized linear mixed effect models using a group Lasso estimator, as proposed by Groll and Tutz (2014), which applies the same amount of shrinkage to all dummy variables that constitute one factor variable. All factor variables in the dataset were coded as dummy variables. All predictor variables, including the dummy variables, were z-standardized to ensure a fair penalization and to compare their relative contributions to the *Annoyance* ratings. The factor levels containing the most observations were selected as the reference category for the dummy creation, as depicted in the caption of **Table 6**.

RESULTS

Descriptive Statistics

Sound Categories: Macro-Level

Participants reported 904 *natural* (32%), 552 *human* (20%), and 1,344 *technical* sounds (48%) as well as 1,260 (45%) *mid/high-level* and 1,540 (55%) *low-level* sounds (separated by the binary perceived *Loudness* variable). The results of a linear mixed-effects model (*Macro1*) revealed that predicted mean *Annoyance* of the three *macro-level sound categories* differed significantly according to the bootstrapped CIs (model *Macro1*): M_{natural} = 1.87, CI [1.79, 1.95]; M_{human} = 3.06, CI [2.94, 3.19]; $M_{\text{technical}}$ = 3.41, CI [3.33, 3.48]. **Figure 2** shows the estimated marginal means and CIs for both levels of perceived *Loudness*, indicating significant differences between all means (model *Macro2*). The differences of the estimated marginal mean values in relation to the *Loudness* levels were the same for all three sound categories due to the addition of *Loudness* levels as a second fixed factor. **Figure 2** also displays the underlying distributions of the measured data for the subsets shown. The distributions for the *macro-level sound categories human* and *technical* differed across *Loudness* levels. As expected, more *mid/high*-level sounds were reported at higher levels of *Annoyance*. The *low-level human* sounds were in the opposite direction, and *low-level technical* sounds were normally distributed.

[8] SE, Standard Error.

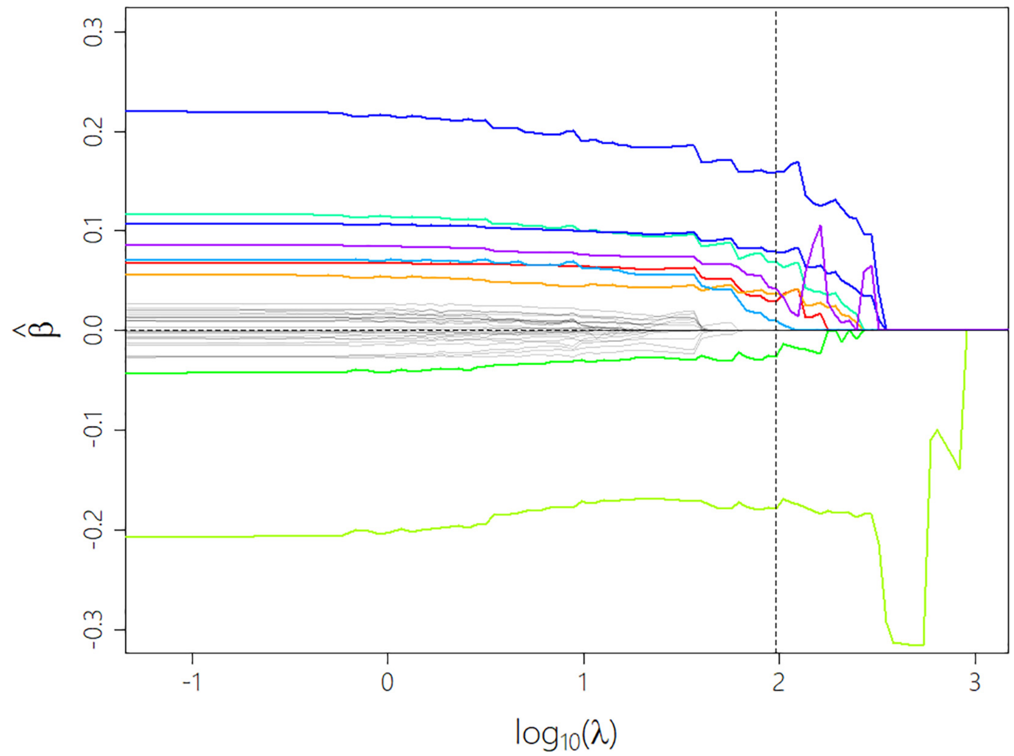


FIGURE 1 | Progression of the coefficients as a function of the tuning parameter λ during the shrinking process. The colored lines show predictors that don't get eliminated until the optimal λ (vertical dotted line) is reached. Dummy variables that constitute one factor variable share the same color. Most coefficients follow the expected decreasing trend while some (see light green curve) show a completely unexpected and sometimes even strongly transient progression which can be considered as regularization artifacts.

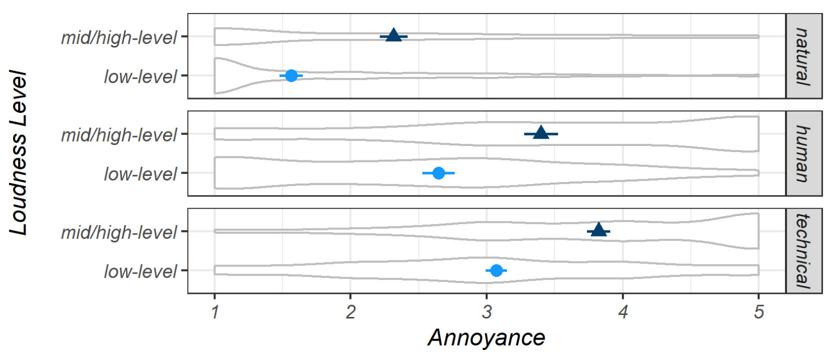


FIGURE 2 | Estimated marginal means for *Annoyance* for *natural*, *human*, and *technical* sounds, separated by binary *Loudness* levels, displayed with 95% confidence intervals, both determined by bootstrapping. *Very pleasant* = 1, *very annoying* = 5. Distributions of the underlying measured *Annoyance* judgments are presented in gray. Model *Macro2*.

Influence of Micro-Level Sound Categories

Participants reported a total of 2,800 sounds that were merged into 38 *micro-level sound categories* (see **Figure 3** with data from model *Micro2*). Similar to the sound categories at the *macro-level* shown in **Figure 2**, the estimated mean values for the categories at the *micro-level* presented in **Figure 3** were equally spaced between the two *Loudness* levels. Since the distributions of the measured data (not shown here) differed for the *micro-level* categories even more than for the *macro-level* categories, the estimated marginal mean values and CIs must be interpreted with the information given above. To provide a more realistic view of these differences, we calculated two different models for both *Loudness* subsets that are not discussed in detail here (models *Micro3*a/b; **Figure A1** in the **Appendix**).

In addition to the *Loudness*-dependent results shown in **Figure 3**, we now discuss the Loudness-independent estimated

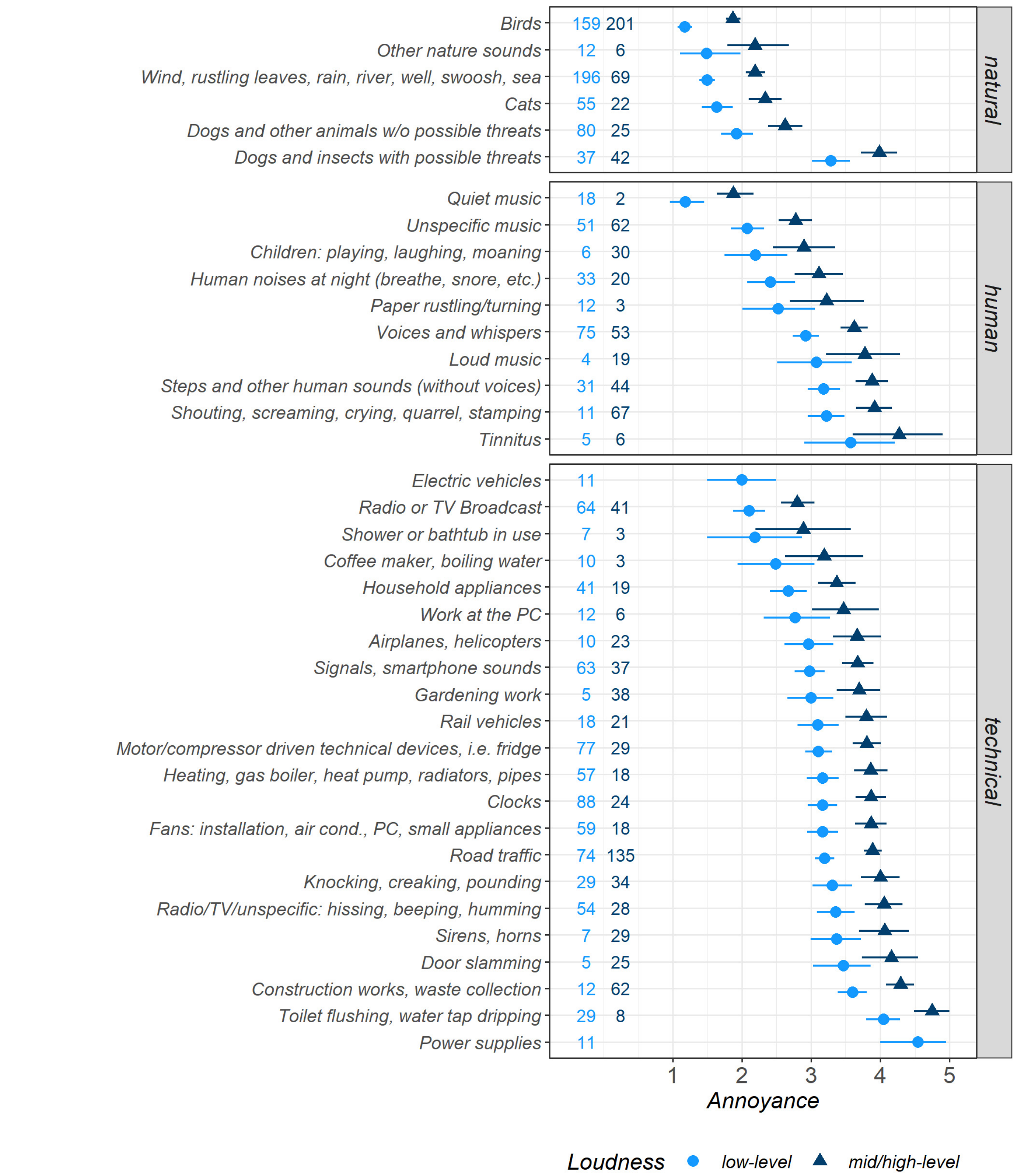


FIGURE 3 | Estimated marginal means for *Annoyance* for sounds from 38 *micro-level sound categories*, separated by binary *Loudness* levels, displayed with 95% confidence intervals (both determined by bootstrapping) and the numbers of observations per category and *Loudness* level subsets. *Very pleasant* = 1, *very annoying* = 5. Model *Micro2*.

marginal means (model *Micro1*). The reports comprised 360 sounds from *Birds*, which constituted the most pleasant natural category (i.e., that with the lowest *Annoyance* ratings), and overall the category with the highest number of reports (*Annoyance M* = 1.56, CI [1.47, 1.66]). With low *Loudness* values, 44% of these sounds were classified as *low-level*. By contrast, *Dogs and insects with possible threats* were the category of natural sounds that participants found most annoying on average (*M* = 3.66, CI [3.38, 3.93]; 47% *low-level* sounds). The remaining natural categories fell close together between the *Birds* value and the *neutral/ambiguous Annoyance* value 3. The second-highest occurrence was reports of noisy, non-tonal sounds like *Wind, rustling leaves, rain, [. . .], sea* (M = 1.68, CI [1.56, 1.80]; 74% *low-level*), and *Cats* (M = 1.84, CI [1.69, 2.09]; 71% *low-level*), which they perceived as pleasant.

The *human* sound category *Quiet music* was the most pleasant of all *micro-level sound* categories, with 20 reports (M = 1.25, CI [1.05 1.50]; 90% *low-level*). The remaining *human* sound categories hovered around the neutral/ambiguous *Annoyance* value 3 between the second most common *human* category *Unspecific music* (M = 2.46, CI [2.21, 2.71]; 45% *low-level*; making music, radio play, movie) and the most annoying *human* category *Tinnitus* with 11 reports (M = 3.95, CI [3.21, 5.62]; 46% *low-level*). The category *Human noises at night* (M = 2.68, CI [2.29, 3.08]; 33% *low-level*) was followed by *Voices and whispers* with the highest number of *human* sounds (M = 3.21, CI [3.03, 3.40]; 59% *low-level*).

The most pleasant *technical* category consisted of 11 sounds from *Electric vehicles* (M = 2.00, CI [1.50, 2.50]; 100% *low-level*). For *neutral/ambiguous Annoyance* ratings (value 3), there were reports of *Household appliances*, such as washing machines and dishwashers (M = 2.89, CI [2.60, 3.20]; 68% *low-level*). Most other *technical* sound categories mainly fell above the *Annoyance* value 3, comprising 112 reports of *Clocks* (M = 3.32, CI [3.09, 3.54]; 77% *low-level*) and 106 sounds from motor- or compressor-driven fridges and freezers (M = 3.30, CI [3.10, 3.50]; 73% *low-level*). *Road traffic* was the most frequently mentioned *technical* category (M = 3.64, CI [3.50, 3.79]; 35% *low-level*).

Influence of Location

We evaluated the situation descriptions regarding the categorization of the *Location* in which participants experienced the sound (see **Table 1** for the numbers of observations per *Location*) and established six categories: *Garden/Nature*; *Home (indoors)*; *Work/Office*; *Other (indoors)*, such as driving in a car or being in a cinema; *Other (outdoors)*, including walking or riding a bike; and *Undefined*.

Figure 4 shows the marginal means for *Annoyance* separated by the binary *Loudness* levels for the six locations given above (model *Location*). *Mid/high-level* sounds were only rated as pleasant for the *Location Garden/Nature* (M = 2.33, CI [2.20, 2.47]). In contrast, *mid/high-level* sounds in all other locations were, on average, reported as neutral or slightly annoying (*Home (indoors)*: M = 3.48, CI [3.38, 3.57]; *Work/Office*: M = 3.64, CI [3.42, 3.86]; *Other (indoors)*: M = 3.38, CI [3.20, 3.56]; *Other (outdoors)*: M = 3.12, CI [2.83, 3.42]; and *Undefined*: M = 3.46, CI [3.30, 3.62]). Unsurprisingly, the average estimated *Annoyance* means for *low-level* sounds from all locations showed less *Annoyance* and were rated from neutral (*Work/Office*: M = 2.85, CI [2.64, 3.06]; *Other (indoors)*: M = 2.59; CI [2.41, 2.76]; *Home (indoors)*: M = 2.68, CI [2.60, 2.76]; *Undefined*: M = 2.67, CI [2.51, 2.83]) to pleasant (*Other (outdoors)*: M = 2.33, CI [2.04, 2.62]), with *Garden/Nature* having the most pleasant ratings on average by far (M = 1.54, CI [1.42, 1.65]). The indoor locations and the *Undefined* category showed similar patterns.

Most of the estimated *Annoyance* means for all *Loudness* levels together (not displayed in **Figure 4** for better readability) were rated as neutral or ambiguous (*Other (outdoors)*: M = 2.61, CI [2.13, 3.20]; *Home (indoors):* M = 3.04, CI [2.96, 3.12; *Work/Office*: M = 3.15, CI [2.94, 3.36]; *Other (indoors):* M = 2.92, CI [2.75, 3.08]); *Undefined*: M = 3.15, CI [2.98, 3.31]) except for sounds from *Garden/Nature*, which had the only pleasant mean value (M = 1.87, CI [1.76, 1.99]).

Living Environment

Of all participants, 63.6% stated that they lived in a *calm* or *very calm* area. They reported 63.3% of all assessed sound situations and 64.4% of all *low-level* sounds, as depicted in **Table 3**. The bootstrapped estimated marginal means of all subsets differed significantly regarding both *Loudness* and *Liveliness* levels, as indicated by their CIs (model *Liveliness*).

Single-Fixed-Factor Models

In this section, we present the results of several linear mixed-effects models, each including only one fixed factor, to investigate the effect sizes and directions of the single bivariate relationships related to perceived *Annoyance* of *low-level* sounds (see **Table 4**, model *32SFF*). **Figure 5** (model *32SFFA*) depicts the R^2_m and probability values for all variables assessed (except for the soundscape dimensions, which were correlated with our target variable *Annoyance* and thus would lead to tautological findings). Among the crucial variables that explained a substantial amount of variance (own criterion of $R^2_m \geq 0.05$) were eight situational factors (the affective *Valence*, *Arousal*, *Control*, the *Positivity* and *Negativity* DIAMONDS dimensions, the *Specific Fade-out* ability, the *Location*, and *Active Coping* reaction) and three sound-related factors (*macro-level* and *micro-level sound category* as well as *Loudness*) but no *person-related* factor. These relevant variables will be examined in more detail in the following sections.

Role of Situational Factors

Of the situational variables, *Valence* explained the most variance in the *Annoyance* evaluations ($\beta = -0.87$; $R^2_m = 0.411$), followed by *Arousal*, which had a lower but still substantial explanation of variance ($\beta = 0.63$; $R^2_m = 0.213$). While positive *Valence* was associated with higher pleasantness (less *Annoyance*), high *Arousal* was related to higher *Annoyance* judgments. Among the situational DIAMONDS dimensions, *Positivity* ($\beta = -0.60$; $R^2_m = 0.191$) and *Negativity* ($\beta = 0.52$; $R^2_m = 0.145$) revealed the most substantial associations with *Annoyance*. The *Specific Fade-out* ability ($\beta = -0.46$; $R^2_m = 0.112$) was followed by several minor effects: Concerning the *Location* ($R^2_m = 0.078$) variable, being in the garden or nature ($\beta = -0.38$) instead of staying at home (reference level) was associated with more pleasant sounds. *Active Coping* reactions ($\beta = 0.37$; $R^2_m = 0.070$) and perceived *Control* ($\beta = -0.37$; $R^2_m = 0.071$) showed similar variance explanations and similar effects but in opposite directions: More *Active Coping* was associated with greater *Annoyance*, while higher levels of perceived *Control* were linked to less annoying sound evaluations. The other situation dimensions, in contrast, revealed R^2_m values below 0.050.

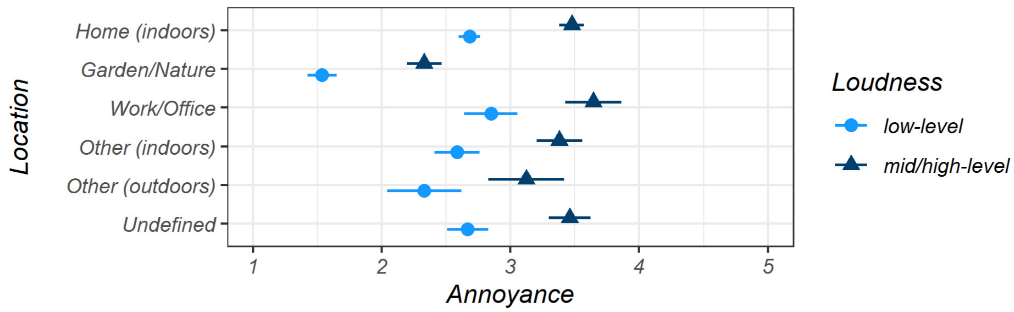


FIGURE 4 | Estimated marginal means for *Annoyance* for all *Location* categories, separated by binary *Loudness* levels, displayed with 95% confidence intervals, both determined by bootstrapping. *Very pleasant* = 1, *very annoying* = 5. Model *Location*.

Role of Sound-Related Factors

Among sound-related variables, both the three *macro-level* (R^2_m = 0.207) and 38 *micro*-level sound categories (R^2_m = 0.330) explained a substantial amount of variance in the *Annoyance* ratings. Both *natural* (β = −0.65) and *human* (β = −0.14) *macro*-level sound categories had more pleasant sound evaluations compared to the *technical* category. They were followed by perceived *Loudness* (β = 0.33; R^2_m = 0.058). Whereas the sound characteristic *Tonality* was significant (β = −0.08; R^2_m = 0.003; p = 0.001), *Timbre* was not (β = −0.03; R^2_m < 0.001; p = 0.257). However, both variables showed a negligible effect size (|β| ≤ 0.08; R^2_m ≤ 0.003).

Role of Person-Related Factors

The *General Fade-out* ability showed the highest, albeit still quite small, variance explanation (R^2_m = 0.029) and was negatively associated with *Annoyance* judgments (β = −0.24). The *Net Income* of the household as well as the *Income per Person* revealed very little variance explanation but similar negative effects, suggesting that higher income was associated with more pleasant sound situations. By contrast, the personality traits *Noise Sensitivity* (β = 0.14) and *Neuroticism* (β = 0.12) were significant positive predictors of *Annoyance* judgments but showed minimal R^2_m values below 0.010. Moreover, *Extraversion* (β = 0.04) did not show a significant effect at all (p = 0.274; R^2_m = 0.001). All other demographic variables and the *Liveliness* of the living environment showed minimal R^2_m values (<0.010) and/or were insignificant. For *Liveliness* (β = 0.12), a more lively living environment was associated with higher *Annoyance* in sound evaluation. A model for the interaction effect of *Age Class* and *Education Class* on our measured data (model *Age*Edu*, **Table 5**) confirmed the findings of Miedema and Vos (1999) with these significant influences on annoyance: Younger people (20–40 years) reported less annoying sound situations (β = −0.26) than older people (61–80 years; reference dummy level). Participants with no or up to a lower secondary–level education were slightly less annoyed (β = −0.12) than people with a university-level education (reference dummy level). Finally, only one of the four interactions was significant, showing that young people with an upper secondary–level education experience less annoying sound situations than older people with a university-level education (β = −0.22).

Comprehensive Models

In this section, we present the comprehensive model *CML1* predicting *Annoyance* ratings derived from the Lasso regression method for variable selection. Some variables had not been processed due to missing values (*Net Income* and *Income per Person*)—which were not allowed for the regularization method—or having too many factor levels (*Micro Sound Categories*). The minimum cross-validation error (CV = 0.81) was reached at a tuning parameter value of λ_{opt} = 96.0 (see dashed line in **Figure 1**). This optimal compromise between a model in which all our variables were retained as contributors and a model without any fixed factors left over (at a λ_{max} = 2,256) incorporated the most important variables shown in **Table 6**. This prediction model explained over half of the variance of the *Annoyance* evaluations (R^2_m = 0.570).

The most crucial fixed factor was the affect *Valence* (β = −0.47, CI [−0.51, −0.42]) followed by the *natural* sound category (β = −0.38, CI [−0.41, −0.34]) and the *Specific Fade-out* ability (β = −0.20, CI [−0.24, −0.16]). The situational variable *Positivity* (β = −0.19, CI [−0.22, −0.15]) was included in the model, whereas *Negativity* was excluded. Lower *Annoyance* ratings were therefore related to higher (more positive) *Valence* scores, *natural* sounds as opposed to *technical* ones, and the stronger ability of respondents to fade out sounds. In contrast to the aforementioned negative effects, the fifth most important variable was the positive effect *mid/high-level Loudness* (β = 0.17, CI [0.14, 0.21]), followed by the positive effects *Active Coping* (β = 0.13, CI [0.09, 0.16]) and *Arousal* (β = 0.13, CI [0.09, 0.18]). That is, higher *Annoyance* values were related to higher *Loudness*, higher *Arousal*, and higher *Active Coping* scores. The sound characteristic *Tonality* had the smallest significant effect (β = −0.05, CI [−0.09, −0.02]). Finally, the *General Fade-out* capability (β = −0.01, CI [−0.05, 0.03]) was included as the only variable with a non-significant p-value (p = 0.526), whereas all other effects had highly significant p-values (p < 0.010).

Concerning the random effect *ID*, the residual (within-subject) variance σ^2 = 0.60 and the random intercept (between-subject) variance τ_{00} = 0.21 were observed. The quite high within-subject variance of the *Annoyance* ratings may be due to high variation in the characteristics of the sounds and situations reported by participants. This variation is slightly smaller for the subset of sounds reported as low-level (σ^2 = 0.55). An

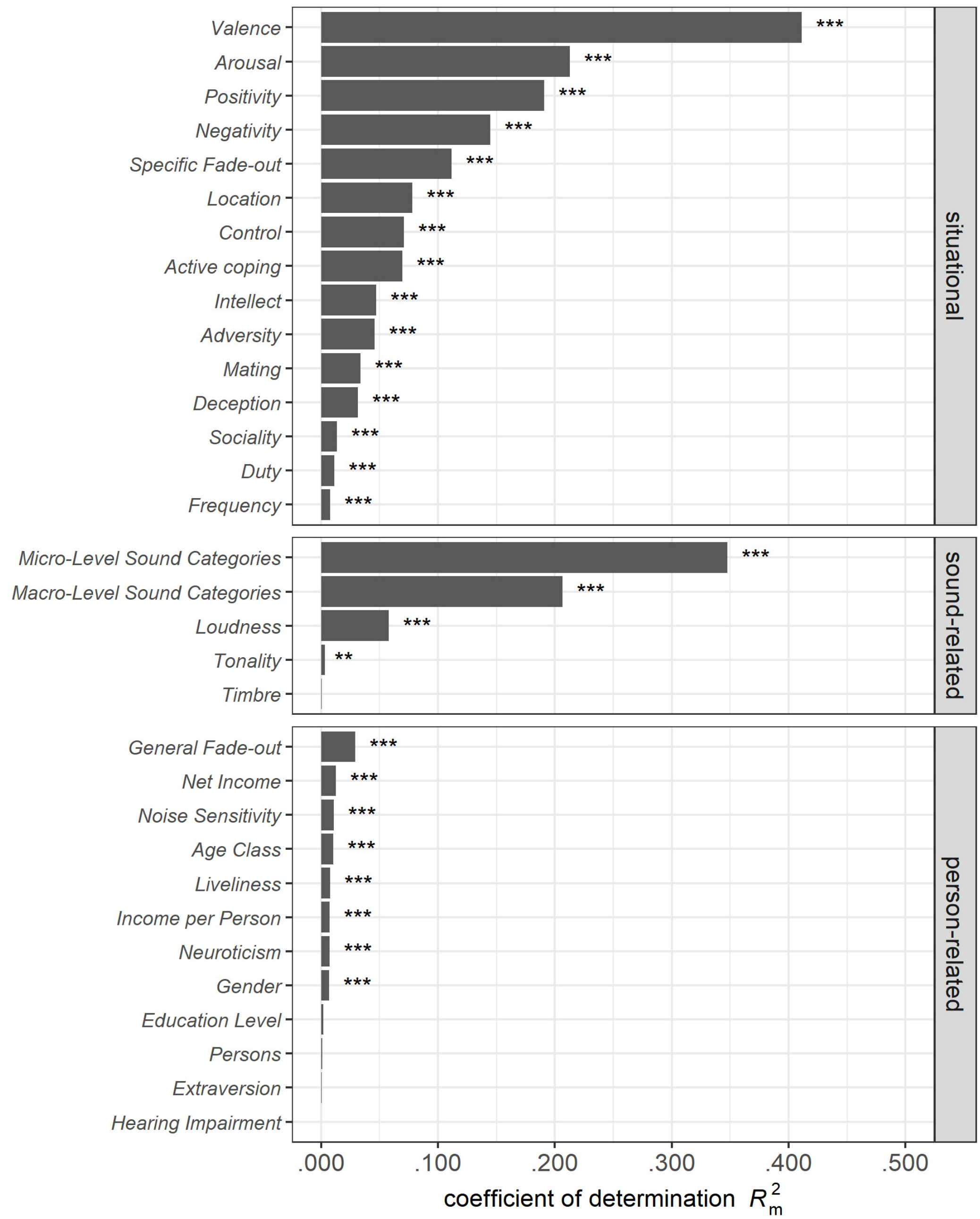


FIGURE 5 | $R^2{}_m$ and probabilities for the assessed variables determined by bivariate analyses of the single-fixed-factor models. Probabilities are given as $^{***}p < 0.001$; $^{**}p < 0.010$; $^{*}p < 0.050$. Model *32SFFA*.

TABLE 3 | Frequencies of observations and estimated marginal means for *Annoyance* judgments, differentiated by the *Liveliness* of the living environment.

	Lively				Calm				
	N_{ID}	N_{obs}	*M*	CI	N_{ID}	N_{obs}	*M*	CI	$N_{obs\ sum}$
Low-level	474	544	2.62	[2.51, 2.74]	827	984	2.39	[2.30, 2.47]	1,528
High-level		484	3.43	[3.31, 3.55]		788	3.20	[3.10, 3.30]	1,272
$N_{obs\ sum}$		1,028				1,772			2,800

With bootstrapped confidence intervals for each level of Liveliness of the living environment and each perceived Loudness level. N_{ID} *number of participants;* N_{obs} *number of observations; model Liveliness.*

TABLE 4 | *Annoyance* estimates of bivariate single-fixed-factor models including dummy variables of each factor.

	All *Loudness* levels; factors from single-fixed-factor models					
Predictors	**Estimates**	**[CI]**	***p***	***df***	R^2_m	R^2_c
Situational						
Valence	−0.87	−0.91, −0.83	**<0.001**	2717.6	0.411	0.563
Arousal	0.63	0.58, 0.68	**<0.001**	2772.3	0.213	0.476
8D: *Positivity*	−0.60	−0.64, −0.55	**<0.001**	2797.1	0.191	0.481
8D: *Negativity*	0.52	0.47, 0.57	**<0.001**	2789.5	0.145	0.438
Specific Fade-out	−0.46	−0.51, −0.41	**<0.001**	2795.2	0.112	0.444
Location[5]: *Garden/Nature*	−0.38	−0.43, −0.33	**<0.001**	2721.4	0.078	0.431
Location[5]: *Work/Office*	0.01	−0.04, 0.06	0.600	2680.7		
Location[5]: *Other (indoors)*	−0.01	−0.06, 0.03	0.565	2495.3		
Location[5]: *Other (outdoors)*	−0.06	−0.11, −0.02	**0.006**	2623.4		
Location[5]: *Undefined*	0.03	−0.02, 0.07	0.279	2740.1		
Control	−0.37	−0.42, −0.32	**<0.001**	2797.8	0.071	0.444
Active Coping	0.37	0.32, 0.41	**<0.001**	2665.3	0.070	0.461
8D: *Intellect*	−0.30	−0.35, −0.25	**<0.001**	2752.3	0.047	0.434
8D: *Adversity*	0.30	0.24, 0.35	**<0.001**	2781.0	0.046	0.427
8D: *Mating*	−0.25	−0.30, −0.20	**<0.001**	2790.1	0.034	0.435
8D: *Deception*	0.24	0.19, 0.30	**<0.001**	2768.7	0.031	0.420
8D: *Sociality*	−0.16	−0.21, −0.11	**<0.001**	2760.1	0.013	0.422
8D: *Duty*	0.15	0.10, 0.20	**<0.001**	2693.8	0.011	0.415
Frequency[7]: less than once a year	0.11	0.06, 0.17	**<0.001**	2724.4	0.008	0.422
Frequency[7]: 1.4 times a year	0.06	0.01, 0.11	**0.028**	2628.4		
Frequency[7]: 5.11 times a year	0.03	−0.02, 0.09	0.241	2638.1		
Frequency[7]: 1.3 times a month	0.07	0.01, 0.13	**0.013**	2546.8		
Frequency[7]: 1.3 times a week	0.03	−0.02, 0.09	0.256	2485.0		
Frequency[7]: more than once a day	0.02	−0.04, 0.08	0.526	2547.3		
Sound-related						
Macro-level sound category[8]: *natural*	−0.65	−0.70, −0.61	**<0.001**	2615.9	0.207	0.512
Macro-level sound category[8]: *human*	−0.14	−0.18, −0.10	**<0.001**	2461.3		
Loudness: mid/high-level (ref.: *low-level*)	0.33	0.28, 0.38	**<0.001**	2767.0	0.058	0.412
Tonality	−0.08	−0.12, −0.03	**0.001**	2476.9	0.003	0.426
Timbre	−0.03	−0.07, 0.02	0.257	2527.7	<0.001	0.419
Person-related						
General Fade-out	−0.24	−0.30, −0.18	**<0.001**	1240.4	0.029	0.417
Net Income of the household[1]	−0.16	−0.22, −0.09	**<0.001**	1174.1	0.013	0.416
Noise Sensitivity	0.14	0.08, 0.21	**<0.001**	1237.2	0.011	0.418
Liveliness: *lively* (Ref.: *calm*)	0.12	0.06, 0.19	**<0.001**	1248.7	0.008	0.418
Neuroticism	0.12	0.05, 0.18	**<0.001**	1238.8	0.007	0.419
Income per Person[1,2]	−0.12	−0.18, −0.05	**<0.001**	1161.6	0.007	0.415
Gender: male. (Ref.: *female*)	0.11	0.05, 0.18	**<0.001**	1257.6	0.007	0.419
Education Class[4]: ISCED level ≤ 2	0.07	−0.00, 0.14	0.054	1265.96	0.002	0.418
Education Class[4]: ISCED level 3	0.02	−0.05, 0.09	0.567	1228.29		
Persons living in the household	−0.04	−0.11, 0.02	0.202	1255.8	0.001	0.418
Extraversion	0.04	−0.03, 0.10	0.274	1254.6	0.001	0.418
Age Class[3]*:* 20 to 40 Years	0.16	0.09, 0.24	**<0.001**	1268.6	0.011	0.419
Age Class[3]*:* 41–60 Years	0.07	−0.01, 0.14	0.070	1237.0		
Hearing Impairment: yes. (Ref.: *no*)	−0.01	−0.07, 0.06	0.802	1273.4	<0.001	0.418
N_{ID}	1,301					
N_{obs}	2,800					

[1]N_{obs} 2612, N_{ID} 1215. Reference levels: [2] [3200; 4500[EUR, [3]61–80 years, [4]ISCED level 4.8, [5]Home (indoors), [6]Undefined, [7]4.7 times a week, [8]macro-level sound category: technical. N_{obs} number of observations. N_{ID} number of participants, 8D situational eight DIAMONDS. CI confidence intervals and p-values were not corrected. The 38 micro-level sound categories are not shown for better readability. Model 32SFF. p-values that are significant at the level of $\alpha < 0.050$ are shown in bold.

TABLE 5 | *Annoyance* estimates of the *Age Class* and *Education Class* interaction effect model including all dummy variables of each factor.

	All *Loudness* levels					
Predictors	**Estimates**	**[CI]**	***p***	***df***	R^2_m	R^2_c
Intercept	2.85	2.79, 2.91	**<0.001**	1257.3	0.016	0.421
Age Class[1]: 20–40 Years	−0.26	−0.37, −0.15	**<0.001**	1269.3		
Age Class[1]: 41–60 Years	0.03	−0.08, 0.14	0.601	1245.2		
Education Class[2]: ISCED level ≤ 2	−0.12	−0.23, −0.00	**0.042**	1265.6		
Education Class[2]: ISCED level 3	0.04	−0.07, 0.15	0.504	1248.6		
*Age Class[1]: 20–40 Years * Education Class[2]: ISCED level ≤ 2*	−0.01	−0.21, 0.19	0.925	1274.4		
*Age Class[1]: 41–60 Years * Education Class[2]: ISCED level ≤ 2*	0.01	−0.18, 0.20	0.922	1256.5		
*Age Class[1]: 20–40 Years * Education Class[2]: ISCED level 3*	−0.22	−0.41, −0.03	**0.025**	1264.0		
*Age Class[1]: 41–60 Years * Education Class[2]: ISCED level 3*	−0.01	−0.20, 0.18	0.886	1233.4		
N_{ID}	1,301					
N_{obs}	2,800					

*Reference levels: [1]61–80 years, [2]ISCED level 4.8. N_{obs} number of observations. N_{ID} number of participants, CI bootstrapped confidence intervals were given at α = 0.05. Model Age*Edu. p-values that are significant at the level of $\alpha < 0.050$ are shown in bold.*

interpretation of this might be that the low-level sounds reported by participants were generally less annoying, whereas mid/high-level sounds ($\sigma^2 = 0.65$) can be very annoying or even very pleasant—for example, imagine playing your favorite music or the *Bird* sounds that were reported as being *mid/high-level*. This can be seen in **Figure 2**, which shows the distribution for the two sound level subsets (additionally separated by *macro-level sound category*, model *Macro2*). A similar relationship between the sound level subsets can be observed for the between-subject variation τ_{00} of the random effect. These values were 0.3–0.4 times the within-subject variances. Unsurprisingly, the relationship between the sound level subsets mentioned above can also be found in the standard deviations of the raw data ($SD_{\text{all}} = 1.39$; $SD_{\text{low-level}} = 1.24$; $SD_{\text{mid/high-level}} = 1.42$; see also the distributions of the reported raw-data in **Figure 2**). Finally, an adjusted (i.e., conditional) intraclass-correlation coefficient for the full dataset—$ICC_{\text{adj}} = \tau_{00}/(\tau_{00} + \sigma^2) = 0.26$—described the proportion of explained variance to total variance (including the fixed effects) due to differences between participants which were represented by the random effect *ID*. From a critical perspective, all of the above differences in variances and their interpretation may be strongly influenced by the huge variety in the sounds reported by participants due to the fact that each participant reported individual sounds, as no audio was played back and no grouped listening (as in sound walks) was performed.

In addition to the aforementioned model computed over the full dataset, we derived two further models using the Lasso regression method for the subsets of *low-level* (model *CML2*) versus *mid/high-level* (model *CML3*) observations. This was done to investigate whether evaluations of both *low-* and *mid/high-level* sounds would follow similar patterns. Both models showed similar but slightly smaller marginal R^2-values ($R^2_{\text{m\_low-level}} = 0.532$; $R^2_{\text{m\_mid/high-level}} = 0.543$) compared to the overall model. The *Loudness* variable was no longer included in the sub-models, presumably because it served as the grouping variable. The two variables *Tonality* and *General Fade-out* were also excluded by the Lasso regression at the optimal tuning parameters (*low-level*: $\lambda_{\text{opt}} = 90.8$, $\lambda_{\text{max}} = 1{,}231$; *mid/high-level*: $\lambda_{\text{opt}} = 117.0$, $\lambda_{\text{max}} = 1{,}124$). Compared to the overall model, the model for the *low-level* subset showed a slightly smaller cross-validation error ($CV = 0.72$). In contrast, the error of the *mid/high-level* model was slightly higher ($CV = 0.93$). When comparing the predictor estimates of the two models based on their CIs, no significant differences were observed, and both showed the same selected variables.

A model (*CMSFF1*) containing all variables that were significant in the bivariate analyses (subsections of section "Design and Questionnaire" and **Figure 5**) and showed an $R^2_m \geq 0.050$ is shown in **Table 7**. The variables selected in this way confirmed the variable selection by the Lasso regularization method. The variable *Control*, which was meaningful in the bivariate analysis ($\beta = -0.37$; $R^2_m = 0.071$; $p < 0.001$), became unimportant in the comprehensive model ($\beta = 0.00$; $p = 0.910$). Some *Location* levels were inconsistently significant within each *Loudness* subset as well as between subsets. Although the Lasso variable selection method—in a misleading manner—selected *General Fade-out*, which was not significant, no *person-related* variable we assessed achieved an R^2_m of 0.050 in the single-fixed-factor models, and such variables were therefore excluded in this comprehensive linear model.

DISCUSSION

Summary

In this online study, we investigated the human perception of low-level environmental sounds and the influencing effects of sound-related, situational, and person-related factors. Moreover, we investigated whether variable-selection methods from linear machine-learning algorithms can aid noise effects and soundscape research by creating comprehensive models which can reliably predict and explain a considerable amount of variance in unseen data which was not used in the training when the model was built.

TABLE 6 | Estimations of Lasso-selected parameters for the full dataset and two *Loudness* subsets.

	All levels		**Low-level**		**Mid/high-level**	
λ_{opt}	96.0		90.8		117.0	
λ_{max}	2,256		1,231		1,124	
CV	0.81		0.72		0.93	
Predictors	**Estimates [CI]**	***p* *df***	**Estimates [CI]**	***p* *df***	**Estimates [CI]**	***p* *df***
(Intercept)	2.84	**<0.001**	2.47	**<0.001**	3.27	**<0.001**
	2.80, 2.88	1193.8	2.42, 2.51	833.9	3.22, 3.33	735.1
Situational						
Valence	−0.47	**<0.001**	−0.49	**<0.001**	−0.43	**<0.001**
	−0.51, −0.42	2787.7	−0.54, −0.43	1517.5	−0.49, −0.36	1252.0
Arousal	0.13	**<0.001**	0.14	**<0.001**	0.11	**0.001**
	0.09, 0.18	2769.5	0.09, 0.19	1472.8	0.05, 0.18	1259.8
Positivity	−0.19	**<0.001**	−0.13	**<0.001**	−0.24	**<0.001**
	−0.22, −0.15	2778.9	−0.18, −0.08	1506.6	−0.30, −0.18	1263.9
Specific Fade−out	−0.20	**<0.001**	−0.14	**<0.001**	−0.23	**<0.001**
	−0.24, −0.16	2788.6	−0.19, −0.10	1508.0	−0.29, −0.17	1262.4
Active Coping	0.13	**<0.001**	0.11	**<0.001**	0.15	**<0.001**
	0.09, 0.16	2785.4	0.07, 0.16	1509.6	0.10, 0.20	1257.5
Sound-related						
Natural sounds	−0.38	**<0.001**	−0.35	**<0.001**	−0.43	**<0.001**
	−0.41, −0.34	2693.2	−0.40, −0.31	1496.3	−0.49, −0.37	1229.7
Human sounds	−0.15	**<0.001**	−0.17	**<0.001**	−0.14	**<0.001**
	−0.18, −0.11	2635.1	−0.21, −0.12	1455.7	−0.19, −0.08	1173.2
Mid/high-level	0.17	**<0.001**	(grouping variable)		(grouping variable)	
	0.14, 0.21	2787.3				
Tonality	−0.05	**0.002**				
	−0.09, −0.02	2700.9				
Person-related						
General Fade-out	−0.01	0.526				
	−0.05, 0.03	1325.9				
Random effects						
σ^2	0.60		0.55		0.65	
τ_{00}	0.21_{ID}		0.16_{ID}		0.26_{ID}	
ICC_{adj}	0.26		0.22		0.29	
N_{ID}	$1,301_{ID}$		930_{ID}		798_{ID}	
N_{obs}	$2,800_{obs}$		$1,528_{obs}$		$1,272_{obs}$	
Marginal R^2	0.570		0.532		0.543	
Conditional R^2	0.680		0.637		0.674	

Reference micro-level sound category: technical. Reference Loudness: low-level. λ_{max} highest value of Lasso tuning parameter, λ_{opt} optimal value of Lasso tuning parameter, CV cross-validation error, i.e., the mean of the mean square errors of all cross-validation folds, σ^2 residual (within-subject) variance, τ_{00} random intercept (between-subject) variance (i.e., variation between individual intercepts and average intercept), ICC_{adj}, adjusted intraclass-correlation coefficient = $\tau_{00}/(\tau_{00} + \sigma^2)$ describes the variance—including the fixed-effects variance—between participants; N_{ID}, number of persons; N_{obs}, number of observations; CI, confidence intervals were given at $\alpha = 0.05$. Models CML1/2/3. p-values that are significant at the level of $\alpha < 0.050$ are shown in bold.

The results of our study corroborate previous findings suggesting that sound evaluations are dependent on myriad influencing factors, in particular situational factors (Fields, 1993; Wolsink et al., 1993; Stallen, 1999; Job et al., 2007; Kroesen et al., 2008; Steffens et al., 2017). Moreover, we demonstrated that linear mixed-effects models combined with novel machine learning variable-selection techniques are applicable in hypothesis testing in noise effects and soundscape research. Furthermore, they can overcome problems associated with overfitting and multicollinearity when many situational and person-related variables are included in the course of a multiple regression. The feasibility of these techniques is further supported by our extensive and time-consuming bivariate analyses of the single variables, which overall led to similar results.

To the best of our knowledge, this is the first study in the realm of sound perception that takes into account such a large number of psychological variables and utilizes linear machine learning to overcome the aforementioned statistical problems. In addition, the established models derived from the percentile-Lasso method maintain interpretability due to the linear, additive

TABLE 7 | Comprehensive model of all parameters from the bivariate analyses that reached an $R^2_m \geq 0.050$, respectively; with the full dataset and two *Loudness* subsets.

	All levels		Low-level		Mid/high-level	
Predictors	***Estimates* [CI]**	***p* *df***	***Estimates* [CI]**	***p* *df***	***Estimates* [CI]**	***p* *df***
(Intercept)	2.84	**<0.001**	2.47	**<0.001**	3.27	**<0.001**
	2.80, 2.88	1188.3	2.42, 2.51	820.2	3.22, 3.33	722.1
Situational						
Valence	−0.46	**<0.001**	−0.47	**<0.001**	−0.42	**<0.001**
	−0.51, −0.41	2779.7	−0.53, −0.42	1510.9	−0.49, −0.35	1231.2
Arousal	0.13	**<0.001**	0.12	**<0.001**	0.11	**0.001**
	0.08, 0.17	2770.8	0.07, 0.18	1469.9	0.04, 0.17	1254.8
Positivity	−0.17	**<0.001**	−0.12	**<0.001**	−0.23	**<0.001**
	−0.21, −0.13	2768.3	−0.17, −0.06	1489.0	−0.30, −0.17	1256.7
Negativity	0.06	0.007	0.06	0.013	0.04	0.173
	0.02, 0.10	2778.5	0.01, 0.11	1502.4	−0.02, 0.11	1254.7
Specific Fade-out	−0.19	**<0.001**	−0.14	**<0.001**	−0.23	**<0.001**
	−0.23, −0.16	2772.0	−0.19, −0.10	1497.9	−0.29, −0.17	1255.0
Location: Garden/Nature	−0.02	0.299	−0.03	0.333	−0.03	0.398
	−0.06, 0.02	2762.6	−0.08, 0.03	1511.4	−0.09, 0.03	1256.4
Location: Work/Office	0.00	0.840	0.02	0.298	−0.03	0.276
	−0.04, 0.03	2771.0	−0.02, 0.07	1512.2	−0.08, 0.02	1226.9
Location: Other (indoors)	−0.05	**0.003**	−0.02	0.370	−0.09	**0.001**
	−0.08, −0.02	2659.1	−0.06, 0.02	1427.7	−0.14, −0.04	1200.9
Location: Other (outdoors)	−0.03	0.052	−0.06	**0.005**	0.02	0.515
	−0.07, 0.00	2738.1	−0.10, −0.02	1494.9	−0.03, 0.07	1228.0
Location: Undefined	−0.03	0.126	0.01	0.680	−0.07	**0.013**
	−0.06, 0.01	2782.2	−0.03, 0.05	1502.3	−0.12, −0.01	1255.1
Control	0.00	0.910	−0.01	0.758	0.02	0.442
	−0.04, 0.04	2736.7	−0.06, 0.04	1453.5	−0.04, 0.08	1257.0
Active Coping	0.12	**<0.001**	0.10	**<0.001**	0.14	**<0.001**
	0.09, 0.16	2781.1	0.06, 0.15	1505.4	0.08, 0.19	1249.9
Sound-related						
Natural sounds	−0.38	**<0.001**	−0.35	**<0.001**	−0.44	**<0.001**
	−0.42, −0.34	2652.8	−0.40, −0.29	1487.9	−0.50, −0.37	1210.1
Human sounds	−0.15	**<0.001**	−0.17	**<0.001**	−0.13	**<0.001**
	−0.18, −0.11	2636.7	−0.21, −0.12	1455.9	−0.19, −0.08	1168.0
Mid/high-level	0.16	**<0.001**	(Grouping variable)		(Grouping variable)	
	0.12, 0.19	2781.7				
Random effects						
σ^2	0.61		0.56		0.64	
τ_{00}	0.20_{ID}		0.15_{ID}		0.26_{ID}	
ICC	0.25		0.21		0.29	
N_{ID}	$1{,}301_{ID}$		930_{ID}		798_{ID}	
N_{obs}	$2{,}800_{obs}$		$1{,}528_{obs}$		$1{,}272_{obs}$	
Marginal R^2	0.573		0.538		0.549	
Conditional R^2	0.678		0.635		0.679	

Reference Location: Home (indoors). Reference micro-level sound category: technical. Reference Loudness: low-level. σ^2, residual (within-subject) variance; τ_{00}, random intercept (between-subject) variance; ICC, intraclass-correlation coefficient; N_{ID}, number of persons; N_{obs}, number of observations; CI, confidence intervals. Models CMSFF1/2/3. p-values that are significant at the level of $\alpha < 0.050$ are shown in bold.

effects of the predictor variables on the outcome variable, as opposed to widespread deep learning approaches that obfuscate those relationships. The percentile-Lasso regression approach is assumed to be particularly useful if multiple (psycho-)acoustic parameters—usually highly correlated—are also taken into account in the course of more comprehensive future studies and models. Moreover, the combination of multilevel modeling and the percentile-Lasso approach will also allow time-series analyses and the separate modeling of inter- and intra-individual processes relevant to everyday sound perception.

The more detailed analysis of our data further supported the feasibility of using the three most frequently reported macro-level sound categories (natural, human, and technical; Axelsson et al., 2010; Bones et al., 2018; ISO/TS 12913-2,ISO, 2018), whose mean annoyance ratings differed significantly. In addition, we derived 38 micro-level sound categories from sound and situation descriptions, which shed light on the kinds of sounds people experience in their day-to-day life, including their prevalence and how they were evaluated depending on the loudness level.

Since the bootstrapping for the marginal mean values is based on resampling through the level 2 cluster variable, it cannot reproduce well the different distributions of, e.g., *low-level* sounds and *mid/high-level* sounds. When a second fixed factor was included in the models, the differences between the marginal means of the two loudness levels were equal across all levels of the other fixed factor (see **Figures 2–4**). For study designs with multiple fixed factor and non-homogeneously distributed data, a specific statistical approach must be developed in future research that relates the clustering to several levels of multiple factors.

Regarding the evaluation of low-level sounds, our models revealed the expected significant positive effects of perceived loudness on annoyance perception. However, the effect size of loudness was oftentimes smaller than that of non-auditory variables (Job, 1996; Kim et al., 2017; World Health Organization, 2018), which might be a result of the study design's focusing on low-level sounds. A significant loudness-dependent difference in annoyance mean values was indeed observed for the three macro-level sound categories (natural, human, and technical). However, none of the effects of the Lasso-selected variables in the optimal models for different loudness levels differed significantly. As such, human sound perception strategies seem to be loudness-level independent, as the same predictors for the two sub-models (low-level and mid/high-level sounds) were selected by the Lasso. This could be affected by the fact that no acoustic stimuli were presented. The assessment of sounds that are remembered after days or weeks may be affected by a memory bias. An example could be the Peak-End rule, according to which people usually base their retrospective judgments on the most intense (Peak) and the last event (End) (Steffens et al., 2017). The inclusion of more cognitive processing may further bias the assessment compared to an *in situ* evaluation. Another explanation might be that some participants remembered a low-level sound while answering the questionnaire but—because no sound was provided—connected it with the sound source that they might have experienced in other situations at a shorter distance, i.e., higher loudness. Such a justification could also explain why the other sound characteristics—tonality and timbre—contributed negligibly to explaining the variance of annoyance or were even not significant. Non-significant predictors, like the *General Fade-out*, may be present in the model, especially resulting from the usage of the 1-SE rule. The reason for this lies in the Lasso variable selection method: The Lasso excludes predictors based on regularization but not *p*-values. The model selection is based on cross-validated MSE values in combination with the 1-SE rule to detect the most generalized and parsimonious model with low prediction error. Finally, we performed statistical testing and *p*-value analysis after the model selection process.

Our results of the bivariate analyses of the DIAMONDS psychological situation dimensions (Rauthmann et al., 2014) showed only two strong associations. As expected, positive situations were associated with low annoyance (i.e., pleasant) judgments and negative situations with high annoyance. The other six dimensions showed only small effects; sociality and duty were insignificant. A situation with intellectual, romantic, or social aspects was associated with more pleasant sounds, whereas adversity, deception, and duty were connected with greater annoyance. A reason for the relatively low contribution of the DIAMONDS dimensions to the annoyance perception may be that all participants rated individual sound situations and incorporated high diversity in objective environmental characteristics that were assessed—in the worst case—only once. The situational variables then become individual perceptions, i.e., personal variables (Rauthmann and Sherman, 2020). This represents a particular challenge for online studies and more valid field studies that must capture situations in a way that reduces otherwise enormous diversity.

Another situational variable—perceived control, often interpreted as perceived dominance over the situation—showed a minor effect in our analyses. In contrast, other studies have emphasized control as an important—if not essential—non-auditory factor with a negative effect on annoyance (Kjellberg et al., 1996; Stallen, 1999; Kroesen et al., 2008). An explanation might be found in the retrospective study design, as participants may not have been able to remember every aspect of the situation described, leading to bias. Nevertheless, our results are still consistent with the findings of Graeven (1975), who reported a small but significant effect of control over noise in the neighborhood or at home, and with the more recent results of Hatfield et al. (2002), who found a small negative effect of perceived control on sleep, reading disturbance, and general symptoms.

Although other studies have discussed coping as one of the top three non-acoustical factors of sound perception (Stallen, 1999; Kroesen et al., 2008), active coping was a minor situational factor for predicting annoyance in our study, both in the bivariate and the Lasso regression model. More active coping was associated with greater annoyance. Our result is in line with several studies positing that coping can be seen as a consequence of annoyance (Glass et al., 1972; Botteldooren and Lercher, 2004; Park et al., 2016). At first glance, we observed a direction of the coping effect contrary to that reported by Kroesen et al. (2008). It seems plausible that one might only feel the strong need for coping activity if one feels highly annoyed by a given situation. Kroesen et al., however, defined a different aspect of coping, namely "coping capacity," which diminishes if one's ability to face a threat is limited or reduced. As a consequence of being not able to cope with the situation, stress rises. By extension, perceiving a higher coping capacity leads to less annoyance.

In our study, the self-determined ability to fade out the specific reported sound in the specific situation was a crucial factor in explaining the variance of annoyance after valence, arousal, positivity, and (though not in the Lasso-selected variables) negativity. This is quite interesting, as to our knowledge there is no research on this topic available. As our study was very broad in scope, we could not explore every aspect in depth. It seems worthwhile to further study this construct, its antecedents and consequences, and its person-related (e.g., attention deficit disorder), situational (e.g., fatigue), and sound-related (e.g., saliency) correlates. Here, it would be particularly interesting to investigate the stability and situation dependency of this ability and whether its effect on annoyance evaluation can be reproduced in other contexts, such as the field or the laboratory.

Interestingly, noise sensitivity showed a minimal variance explanation with a small positive effect, meaning that higher noise sensitivity was associated with greater annoyance. Other studies have identified noise sensitivity as the most crucial factor in the annoyance responses caused by noise, considerably stronger than the exposure level (Job, 1988; Öhrström et al., 1988; Ryu and Jeon, 2011). In contrast, Stölzel (2004) and Kroesen et al. (2008) found no added explained variance or only a small correlation between noise sensitivity and annoyance, which is in line with our findings. One explanation for this discrepancy could be that we used a short questionnaire, the LEF-K (Zimmer and Ellermeier, 1998). Another could be that, according to other researchers, noise sensitivity may be seen as a multidimensional construct, which means that it might be different for various aspects of daily life (Schütte et al., 2007) and sound levels (Job, 1999). Therefore, an (additional) measurement such as the NoiSeQ (Schütte et al., 2007) or the short version NoiSeQ-R, which considers three everyday scenarios (Griefahn, 2008), might be advisable for future studies.

Since 1260 (45%) of all 2800 reported sounds were rated as mid/high-level—although we were interested in the low-level sounds perceived by the participants—one could hypothesize that some participants indicated the presumed volume of the sound source rather than what they heard. It could also be the case that when asked to report low-level sounds, the participants intuitively thought of a low-level sound, such as birdsong. Later, when asked to evaluate the loudness, they probably made a more cognitive evaluation, perhaps putting the sound into a context or comparing it with other sounds and situations. For example, birdsong might appear loud in a quiet morning, while it is certainly still a low-level sound compared to an accelerating bus passing by.

Limitations

Some limitations associated with the test design should be addressed. First, we provided no acoustic stimuli to participants. Instead, they recalled sounds, situations, and behavior, potentially introducing a memory bias (Steffens et al., 2017; Greb et al., 2019). Notwithstanding, our results (for example, in terms of valence and arousal) revealed similar values compared to studies that used acoustical stimuli (e.g., Hall et al., 2013). Furthermore, the data we assessed allowed for the interpretation of correlative relationships between variables but revealed neither directions of effects nor moderation and mediation effects. In addition, each participant reported only one to three observations, which makes it inappropriate to calculate personal means of otherwise time-varying measures. All of these drawbacks can be addressed by conducting a field study applying the experience sampling method and by obtaining a high number of repeated measures for each participant. The authors are preparing a large-scale field study, including on-site sound recordings, which aims to replicate and extend the findings of this study.

CONCLUSION

Despite the limitations mentioned above, our study shows how to deal with many influencing variables in the field of sound perception using machine learning for the selection of the most essential variables. Even though no actual acoustical stimuli were used, our recall-based online study revealed some crucial factors associated with annoyance judgments (valence, arousal, sound categories, and mental fade-out ability). The results of this study also have practical implications for manufacturers of technical equipment and domestic installations, as even low-level sounds—such as toilet flushing, which was associated with high annoyance ratings in our study—can be prominent (Kuwano et al., 2003). Manufacturers of heating installations, for example, may offer their customers a sense of perceived control that can lower annoyance perceptions by enabling customers to adjust the flow rate of the heating installation to reduce flow noise, if temporarily desired.

DATA AVAILABILITY STATEMENT

The data sets presented in this article are not readily available as supplementary material, as the researchers plan to conduct further analyses. Requests for access to the data sets can be sent to SV, siegbert.versuemer@hs-duesseldorf.de.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Medical Faculty of the University of Duisburg-Essen, Germany. All participants provided digital written informed consent by confirming the declaration on data collection and processing before participating. They were anonymous to the researchers at all times.

AUTHOR CONTRIBUTIONS

SV and JS designed the study, interpreted data, reviewed, and edited the manuscript. SV collected the data, performed the statistical analysis, programmed the diagrams, and wrote the first draft of the manuscript. PB implemented machine learning, wrote the corresponding passage in the "Materials and Methods" section, and consulted on statistics during editing. JB-S provided initial concept and consulted in the evaluation process. All authors approved the final version of the manuscript.

FUNDING

This study was sponsored by the German Federal Ministry of Education and Research. "FHprofUnt"-Funding Code: 13FH729IX6.

ACKNOWLEDGMENTS

We thank Christian-Maximilian Steier and the Faculty of Social Sciences and Cultural Studies from the University of Applied Sciences Düsseldorf for providing the online survey platform. We further are grateful to Matthias Ladleif for conducting the pre-test of the online survey as part of his bachelor thesis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.570761/full#supplementary-material

REFERENCES

Aletta, F., van Renterghem, T., and Botteldooren, D. (2018). Influence of personal factors on sound perception and overall experience in urban green areas. A case study of a cycling path highly exposed to road traffic noise. *Int. J. Environ. Res. Public Health* 15:1118. doi: 10.3390/ijerph15061118

Axelsson, Ö., Nilsson, M. E., and Berglund, B. (2010). A principal components model of soundscape perception. *J. Acoust. Soc. Am.* 128, 2836–2846. doi: 10.1121/1.3493436

Babisch, W. (2002). The noise/stress concept, risk assessment and research needs. *Noise Health* 4, 1–11. doi: 10.1007/978-1-4684-5625-7_1

Bangjun, Z., Lili, S., and Di Guoqing. (2003). The influence of the visibility of the source on the subjective annoyance due to its noise. *Appl. Acoust.* 64, 1205–1215. doi: 10.1016/S0003-682X(03)00074-4

Bartosova, M., Svetlak, M., Kukletova, M., Borilova Linhartova, P., Dusek, L., and Izakovicova Holla, L. (2019). Emotional stimuli candidates for behavioural intervention in the prevention of early childhood caries: a pilot study. *BMC Oral Health* 19:33. doi: 10.1186/s12903-019-0718-4

Basner, M., Babisch, W., Davis, A., Brink, M., Clark, C., Janssen, S., et al. (2014). Auditory and non-auditory effects of noise on health. *Lancet* 383, 1325–1332. doi: 10.1016/S0140-6736(13)61613-X

Bates, D. M., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Soft.* 67, 1–48. doi: 10.18637/jss.v067.i01

Belojević, G., and Jakovjlevic, B. (2001). Factors influencing subjective noise sensitivity in an urban population. *Noise Health* 4, 17–24.

Bender, R., and Lange, S. (2001). Adjusting for multiple testing—when and how? *J. Clin. Epidemiol.* 54, 343–349. doi: 10.1016/S0895-4356(00)00314-0

Betella, A., and Verschure, P. F. M. J. (2016). The affective slider: a digital self-assessment scale for the measurement of human emotions. *PLoS One* 11:e0148037. doi: 10.1371/journal.pone.0148037

Beutel, M. E., Jünger, C., Klein, E. M., Wild, P., Lackner, K., Blettner, M., et al. (2016). Noise annoyance is associated with depression and anxiety in the general population- the contribution of aircraft noise. *PLoS One* 11:e0155357. doi: 10.1371/journal.pone.0155357

Bones, O., Cox, T. J., and Davies, W. J. (2018). Sound categories: category formation and evidence-based taxonomies. *Front. Psychol.* 9:1277. doi: 10.3389/fpsyg.2018.01277

Botteldooren, D., and Lercher, P. (2004). Soft-computing base analyses of the relationship between annoyance and coping with noise and odor. *J. Acoust. Soc. Am.* 115, 2974–2985. doi: 10.1121/1.1719024

Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9

Bradley, M. M., and Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology* 37, 204–215. doi: 10.1111/1469-8986.3720204

Bynion, T.-M., and Feldner, M. T. (2017). "Self-assessment manikin," in *Encyclopedia of Personality and Individual Differences*, eds V. Zeigler-Hill, and T. K. Shackelford, (Cham: Springer), 1–3. doi: 10.1007/978-3-319-28099-8_77-1

Costa, P. T., and McCrae, R. R. (2008). "The revised NEO Personality Inventory (NEO-PI-R)," in *The SAGE Handbook of Personality Theory and Assessment: Volume 2 — Personality Measurement and Testing (1 Oliver's Yard*, eds G. J. Boyle, G. Matthews, and D. H. Saklofske, (London: SAGE Publications Ltd), 179–198. doi: 10.4135/9781849200479.n9

Crichton, F., Dodd, G., Schmid, G., and Petrie, K. J. (2015). Framing sound: using expectations to reduce environmental noise annoyance. *Environ. Res.* 142, 609–614. doi: 10.1016/j.envres.2015.08.016

Deen, M., and de Rooij, M. (2020). ClusterBootstrap: an R package for the analysis of hierarchical data using generalized linear models with the cluster bootstrap. *Behav. Res. Methods* 52, 572–590. doi: 10.3758/s13428-019-01252-y

Digman, J. M. (1990). Personality structure: emergence of the five-factor model. *Annu. Rev. Psychol.* 41, 417–440. doi: 10.1146/annurev.ps.41.020190.002221

Fields, J. M. (1993). Effect of personal and situational variables on noise annoyance in residential areas. *J. Acoust. Soc. Am.* 93, 2753–2763. doi: 10.1121/1.405851

Fyhri, A., and Klćboe, R. (2006). Direct, indirect influences of income on road traffic noise annoyance. *J. Environ. Psychol.* 26, 27–37. doi: 10.1016/j.jenvp.2006.04.001

German Federal Statistical Office, (2018). *Verteilung der Privathaushalte in Deutschland nach monatlichem Haushaltsnettoeinkommen im Jahr 2017 (in 1.000)*. Available online at: https://de.statista.com/statistik/daten/studie/3048/umfrage/privathaushalte-nach-monatlichem-haushaltsnettoeinkommen/ (accessed June 08, 2020).

Glass, G. V., Peckham, P. D., and Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Rev. Educ. Res.* 42, 237–288. doi: 10.3102/00346543042003237

Graeven, D. B. (1975). Necessity, control, and predictability of noise as determinants of noise annoance. *J. Soc. Psychol.* 95, 85–90. doi: 10.1080/00224545.1975.9923237

Greb, F., Steffens, J., and Schlotz, W. (2018). Understanding music-selection behavior via statistical learning. *Music Sci.* 1:205920431875595. doi: 10.1177/2059204318755950

Greb, F., Steffens, J., and Schlotz, W. (2019). modeling music-selection behavior in everyday life: a multilevel statistical learning approach and mediation analysis of experience sampling data. *Front. Psychol.* 10:411. doi: 10.3389/fpsyg.2019.00390

Griefahn, B. (2008). Determination of noise sensitivity within an internet survey using a reduced version of the Noise Sensitivity Questionnaire. *J. Acoust. Soc. Am.* 123:3449. doi: 10.1121/1.2934269

Groll, A. (2017). *"glmmLasso: Variable Selection for Generalized Linear Mixed Models by L1-Penalized Estimation. R package Version* 1.5.1. Available at: https://CRAN.R-project.org/package=glmmLasso (accessed June 08, 2020).

Groll, A., and Tutz, G. (2014). Variable selection for generalized linear mixed models by L 1-penalized estimation. *Stat. Comput.* 24, 137–154. doi: 10.1007/s11222-012-9359-z

Gunn, W. J., Patterson, H. P., Cornog, J., Klaus, P., and Connor, W. K. (1975). *A Model and Plan for a Longitudinal Study of Community Response to Aircraft Noise. NASA Technical Memorandum*. Washington, DC: NASA.

Guski, R. (1999). Personal and social variables as co-determinants of noise annoyance. *Noise Health* 1, 45–56.

Guski, R., Schreckenberg, D., and Schuemer, R. (2017). WHO environmental noise guidelines for the European region: a systematic review on environmental noise and annoyance. *Int. J. Environ. Res. Public Health* 14:1539. doi: 10.3390/ijerph14121539

Halevy, N., Kreps, T. A., and de Dreu, C. K. W. (2019). Psychological situations illuminate the meaning of human behavior: recent advances and application to social influence processes. *Soc Pers. Psychol. Compass* 13:e12437. doi: 10.1111/spc3.12437

Hall, D. A., Irwin, A., Edmondson-Jones, M., Phillips, S., and Poxon, J. E. W. (2013). An exploratory evaluation of perceptual, psychoacoustic and acoustical properties of urban soundscapes. *Appl. Acoust.* 74, 248–254. doi: 10.1016/j.apacoust.2011.03.006

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., and Olds, C. C. (1992). Summarizing monte carlo results in methodological research: the one- and two-factor fixed effects ANOVA cases. *J. Educ. Stat.* 17, 315–339. doi: 10.3102/10769986017004315

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edn. New York, N.Y: Springer.

Hatfield, J., Job, R. F. S., Hede, A. J., Carter, N. L., Peploe, P., Taylor, R., et al. (2002). Human response to environmental noise: the role of perceived control. *Int. J. Behav. Med.* 9, 341–359. doi: 10.1207/S15327558IJBM0904_04

Hill, E. M., Billington, R., and Krägeloh, C. (2014). Noise sensitivity and diminished health: testing moderators and mediators of the relationship. *Noise Health* 16, 47–56. doi: 10.4103/1463-1741.127855

Irtel, H. (2007). *PXLab: The Psychological Experiments Laboratory: Version 2.1.11*. Available online at: http://irtel.uni-mannheim.de/pxlab/demos/index_SAM.html (accessed May 6, 2019).

ISO (2018). ISO/TS 12913-2: Acoustics - Soundscape: Part 2: Data Collection and Reporting, no. ISO/TS 12913-2:2018(E). Available online at: https://www.iso.org/obp/ui/#iso:std:iso:12913:-2:dis:ed-1:v1:en (accessed May 22, 2002).

Janssen, S. A., Vos, H., Eisses, A. R., and Pedersen, E. (2011). A comparison between exposure-response relationships for wind turbine annoyance and annoyance due to other noise sources. *J. Acoust. Soc. Am.* 130, 3746–3753. doi: 10.1121/1.3653984

Job, R. F. S. (1988). Community response to noise: a review of factors influencing the relationship between noise exposure and reaction. *J. Acoust. Soc. Am.* 83, 991–1001. doi: 10.1121/1.396524

Job, R. F. S. (1996). The influence of subjective reactions to noise on health effects of the noise. *Environ. Int.* 22, 93–104. doi: 10.1016/0160-4120(95)00107-7

Job, R. F. S. (1999). Noise sensitivity as a factor influencing human reaction to noise. *Noise Health* 1, 57–68.

Job, R. F. S., Hatfield, J., Hede, A. J., Peploe, P., Carter, N. L., Taylor, R., et al. (2007). The role of attitudes and sensitivity in reactions to changing noise. *ICA* 19, 1–5.

Kim, A., Sung, J. H., Bang, J.-H., Cho, S. W., Lee, J., and Sim, C. S. (2017). Effects of self-reported sensitivity and road-traffic noise levels on the immune system. *PLoS One* 12:e0187084. doi: 10.1371/journal.pone.0187084

Kjellberg, A., Landström, U., Tesarz, M., Söderberg, L., and Akerlund, E. (1996). The effects of nonphysical noise characteristics, ongoing task and noise sensitivity on annoyance and distraction due to noise at work. *J. Environ. Psychol.* 16, 123–136. doi: 10.1006/jevp.1996.0010

Klatte, M., Spilski, J., Mayerl, J., Möhler, U., Lachmann, T., and Bergström, K. (2017). Effects of aircraft noise on reading and quality of life in primary school children in Germany: results from the NORAH study. *Environ. Behav*. 49, 390–424. doi: 10.1177/0013916516642580

Krippendorff, K. (2004). *Content Analysis: An Introduction to its Methodology*, 2nd Edn. Thousand Oaks, CA: Sage.

Kroesen, M., Molin, E. J. E., and van Wee, B. (2008). Testing a theory of aircraft noise annoyance: a structural equation analysis. *J. Acoust. Soc. Am.* 123, 4250–4260. doi: 10.1121/1.2916589

Krstajic, D., Buturovic, L. J., Leahy, D. E., and Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *J. Cheminform.* 6:10.

Kuwano, S., Seiichiro, N., Kato, T., and Hellbrück, J. (2003). Memorey of the loudness of sounds and its relation to overall impression. *Acoust. Sci. Technol.* 24, 194–196. doi: 10.1250/ast.24.194

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Soft.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lang, P. J. (1980). "Behavioral treatment and bio-behavioral assessment: computer applications," in *Technology in Mental Health Care Delivery*, eds J. B. Sidowski, J. H. Johnson, and T. A. Williams, (Norwood, NJ: Ablex), 119–137.

Lang, P. J., and Bradley, M. M. (1997). *International Affective Picture System (IAPS): Technical Manual and Affective Ratings*. (Gainesville, FL: NIMH Center for the Study of Emotion and Attention), 39–58.

Lazarus, R. S. (1966). *Psychological Stress and the Coping Process*. New York, NY: McGraw Hill.

Lazarus, R. S., and Folkman, S. (1984). *Stress, Appraisal, and Coping*. New York, NY: Springer.

Lercher, P. (1996). Environmental noise and health: an integrated research perspective. *Environ. Int.* 22, 117–129. doi: 10.1016/0160-4120(95)00109-3

Lewin, K. (1936). *Priticiples of Topologicol Psychology*, trans. F. Heider, and G. Heider, (New York, NY: McGraw Hill).

Lindborg, P., and Friberg, A. (2016). Personality traits bias the perceived quality of sonic environments. *Appl. Sci.* 6:405. doi: 10.3390/app6120405

Lix, L. M., Keselman, J. C., and Keselman, H. J. (1996). Consequences of assumption violations revisited: a quantitative review of alternatives to the one-way analysis of variance f test. *Rev. Educ. Res.* 66, 579–619. doi: 10.3102/00346543066004579

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behav. Res. Methods* 49, 1494–1502. doi: 10.3758/s13428-016-0809-y

Miedema, H. M. E., and Vos, H. (1999). Demographic and attitudinal factors that modify annoyance from transportation noise. *J. Acoust. Soc. Am.* 105, 3336–3344. doi: 10.1121/1.424662

Miedema, H. M. E., and Vos, H. (2003). Noise sensitivity and reactions to noise and other environmental conditions. *J. Acoust. Soc. Am.* 113, 1492–1504. doi: 10.1121/1.1547437

Nakagawa, S., and Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* 4, 133–142. doi: 10.1111/j.2041-210x.2012.00261.x

Öhrström, E., Björkman, M., and Rylander, R. (1988). Noise annoyance with regard to neurophysiological sensitivity, subjective noise sensitivity and personality variables. *Psychol. Med.* 18, 605–613. doi: 10.1017/S003329170000828X

Park, S. H., Lee, P. J., Yang, K. S., and Kim, K. W. (2016). Relationships between non-acoustic factors and subjective reactions to floor impact noise in apartment buildings. *J. Acoust. Soc. Am.* 139, 1158–1167. doi: 10.1121/1.4944034

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ* 316, 1236–1238. doi: 10.1136/bmj.316.7139.1236

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Available online at: https://www.R-project.org/ (accessed February 29, 2002).

Rammstedt, B., and John, O. P. (2007). Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German. *J. Res. Pers.* 41, 203–212. doi: 10.1016/j.jrp.2006.02.001

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., and Kovaleva, A. (2012). *Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit. Big-Five-Inventory-10 (BFI-10). GESIS-Working Papers 22*. Köln: GESIS.

Rauthmann, J. F. (2015). Structuring situational information. *Eur. Psychol.* 20, 176–189. doi: 10.1027/1016-9040/a000225

Rauthmann, J. F. (2018). *S8-II - Eine Ultrakurzversion der Situational Eight DIAMONDS*. Available online at: https://www.psycharchives.org/bitstream/20.500.12034/620/1/PT_9007482_S8-II_Fragebogen.docx (accessed April 25, 2019).

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., et al. (2014). The Situational Eight DIAMONDS: a taxonomy of major dimensions of situation characteristics. *J. Pers. Soc. Psychol.* 107, 677–718. doi: 10.1037/a0037250

Rauthmann, J. F., and Sherman, R. A. (2020). The situation of situation research: knowns and unknowns. *Curr. Dir. Psychol. Sci*. doi: 10.1177/0963721420925546

Rauthmann, J. F., Sherman, R. A., and Funder, D. C. (2015). Principles of situation research: towards a better understanding of psychological situations. *Eur. J. Pers.* 29, 363–381. doi: 10.1002/per.1994

Riedel, N., Köckler, H., Scheiner, J., van Kamp, I., Erbel, R., Loerbroks, A., et al. (2019). Urban road traffic noise and noise annoyance-a study on perceived noise control and its value among the elderly. *Eur. J. Public Health* 29, 377–379. doi: 10.1093/eurpub/cky141

Roberts, S., and Nowak, G. (2014). Stabilizing the lasso against cross-validation variability. *Comput. Stat. Data Anal.* 70, 198–211. doi: 10.1016/j.csda.2013.09.008

Rothman, K. (1990). No adjustments are needed for multiple comparisons. *Epidemiology* 1, 43–46. doi: 10.1097/00001648-199001000-00010

RStudio Team (2019). *RStudio: Integrated Development for R*. Available online at: https://rstudio.com/products/rstudio/ (accessed January 19, 2020).

Ryu, J. K., and Jeon, J. Y. (2011). Influence of noise sensitivity on annoyance of indoor and outdoor noises in residential buildings. *Appl. Acoust.* 72, 336–340. doi: 10.1016/j.apacoust.2010.12.005

Schneider, S. L. (2008). "Applying the ISCED-97 to the German educational qualifications," *The International Standard Classification of Education* (Mannheim: MZES), 77–102.

Schreckenberg, D., Belke, C., Faulbaum, F., Guski, R., Möhler, U., and Spilski, J. (2016). *Effects of Aircraft Noise on Annoyance and Sleep Disturbances Before and After Expansion of Frankfurt Airport - Results of the NORAH study, WP 1 'Annoyance and Quality of Life'*. (Hamburg: Inter.noise), 997–1006.

Schreckenberg, D., Griefahn, B., and Meis, M. (2010). The associations between noise sensitivity, reported physical and mental health, perceived environmental quality, and noise annoyance. *Noise Health* 12, 7–16. doi: 10.4103/1463-1741.59995

Schütte, M., Marks, A., Wenning, E., and Griefahn, B. (2007). The development of the noise sensitivity questionnaire. *Noise Health* 9, 15–24. doi: 10.4103/1463-1741.34700

Serrou, M. B. (1995). *La protection des Riverains Contre le Bruit des Transports Terrestres*. Available at: http://temis.documentation.developpement-durable.gouv.fr/docs/Temis/0026/Temis-0026438/11191.pdf (accessed September 6, 2002).

Shepherd, D., Heinonen-Guzejev, M., Hautus, M. J., and Heikkilä, K. (2015). Elucidating the relationship between noise sensitivity and personality. *Noise Health* 17, 165–171. doi: 10.4103/1463-1741.155850

Smith, C. A., and Lazarus, R. S. (1990). "Emotion and adaptation," in *Handbook of Personality: Theory and Research*, ed. L. A. Pervin, (New York, NY: Guilford), 609–637.

Stallen, P. J. M. (1999). A theoretical framework for environmental noise annoyance. *Noise Health* 1, 69–80.

Steffens, J., Steele, D., and Guastavino, C. (2017). Situational and person-related factors influencing momentary and retrospective soundscape evaluations in day-to-day life. *J. Acoust. Soc. Am.* 141, 1414–1425. doi: 10.1121/1.4976627

Stölzel, K. (2004). *Zusammenhang zwischen Umweltlärmbelästigung und Lärmempfindlichkeit – epidemiologische Untersuchung im Rahmen der Berliner Lärmstudie*. Ph.D. thesis, Humboldt-Universität zu Berlin, Berlin.

Suk, H.-J. (2006). *Color and Emotion – A Study on the Affective Judgment Across Media and in Relation to Visual Stimuli*. Dissertation, School of Social Sciences, Universität Mannheim. Available online at: https://madoc.bib.uni-mannheim.de/1336

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.* 58, 267–288.

UNESCO Institute for Statistics, (2015). *ISCED 2011 Operational Manual: Guidelines for Classifying National Education Programmes and Related Qualifications*. Paris: OECD.

van Kamp, I., Job, R. F. S., Hatfield, J., Haines, M., Stellato, R. K., and Stansfeld, S. A. (2004). The role of noise sensitivity in the noise-response relation: a comparison of three international airport studies. *J. Acoust. Soc. Am.* 116, 3471–3479. doi: 10.1121/1.1810291

van Renterghem, T. (2019). Towards explaining the positive effect of vegetation on the perception of environmental noise. *Urban For. Urban Green.* 40, 133–144. doi: 10.1016/j.ufug.2018.03.007

Võ, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., and Jacobs, A. M. (2009). The berlin affective word list reloaded (BAWL-R). *Behav. Res. Methods* 41, 534–538. doi: 10.3758/BRM.41.2.534

Wolsink, M., Sprengers, M., Keuper, A., Pedersen, T. H., and Westra, C. A. (1993). "Annoyance from wind turbine noise on sixteen sites in three countries," in *Proceedings of the 1993 European Community Wind Energy Conference*, Lübeck, 273–276.

World Health Organization, (2011). *Burden of Disease from Environmental Noise: Quantification of Healthy Life Years Lost in Europe*. Geneva: World Health Organization.

World Health Organization, (2018). *Environmental noise guidelines for the European Region*. Copenhagen: World Health Organization.

Yu, L., and Kang, J. (2008). Effects of social, demographical and behavioral factors on the sound level evaluation in urban open spaces. *J. Acoust. Soc. Am.* 123, 772–783. doi: 10.1121/1.2821955

Zapf, A., Castell, S., Morawietz, L., and Karch, A. (2016). Measuring inter-rater reliability for nominal data - which coefficients and confidence intervals are appropriate? *BMC Med. Res. Methodol.* 16:93. doi: 10.1186/s12874-016-0200-9

Zimmer, K., and Ellermeier, W. (1998). Ein Kurzfragebogen zur Erfassung der Lärmempfindlichkeit. *Umweltpsychologie* 2, 54–63.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

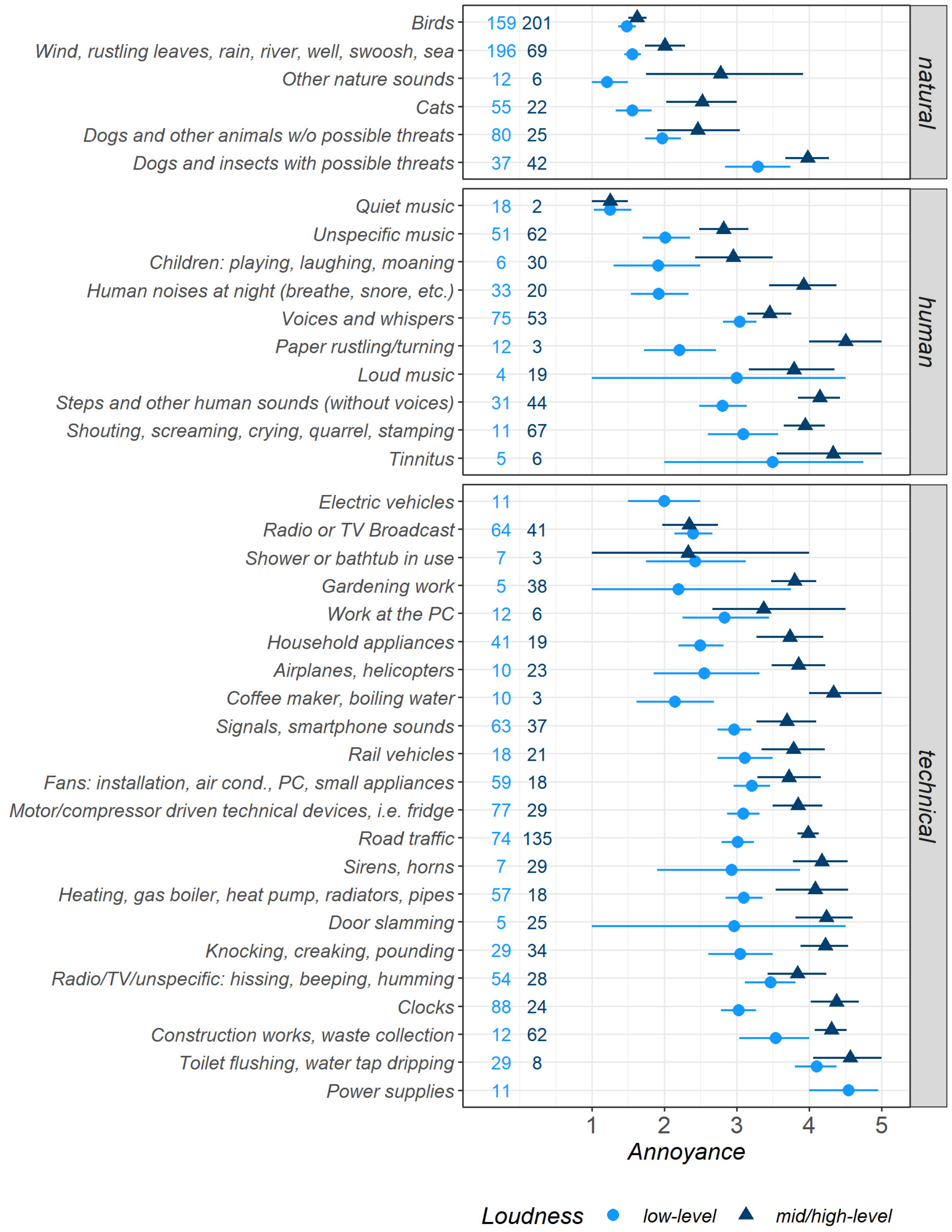


FIGURE A1 | Estimated marginal means for *Annoyance* for sounds from 38 *micro-level sound categories*, separated by binary *Loudness* levels, displayed with 95% confidence intervals (both determined by bootstrapping) and the numbers of observations per category and *Loudness* level subsets. *Very pleasant* = 1, *very annoying* = 5. Models *Micro3a* and *Micro3b*.

frontiers in Psychology

ORIGINAL RESEARCH
published: 29 October 2020
doi: 10.3389/fpsyg.2020.554285

A Conceptual Model of the Healthy Acoustic Environment: Elements, Framework, and Definition

*Jing Chen and Hui Ma**

School of Architecture, Tianjin University, Tianjin, China

Noise has been proved to be a risk factor of physiological and psychological health. Therefore, creating a high-quality acoustic environment for people is particularly important. The aims of this study are to explore the basic elements, propose a conceptual framework, and identify the definition of a healthy acoustic environment. Through the method of grounded theory, 75 respondents participated in interviews. The results revealed that (1) "sound sources and acoustic environment," "people's demands," "criteria and standards of a healthy acoustic environment," "matching process," "secondary fitting process," "context," and "acoustic environment quality" are the basic elements of a healthy acoustic environment; (2) "matching process" and "secondary fitting process" connect all the other categories and reflect the processes by which a healthy acoustic environment is judged; (3) based on the associations revealed in the framework, a healthy acoustic environment is defined as a supportive acoustic environment that can match people's physiological, psychological, and behavioral demands in context, and that also fits the criteria and standards. The proposal of a conceptual model for a healthy acoustic environment can provide a new perspective on designing and establishing a high-quality acoustic environment required by people in the near future.

Keywords: healthy acoustic environment, conceptual framework, definition, physiological demands, psychological demands, behavioral demands, criteria and standards, grounded theory

OPEN ACCESS

Edited by:
Francesco Aletta, University College London, United Kingdom

Reviewed by:
Andreas Liebl, HSD Hochschule Döpfer University of Applied Sciences, Germany
Woon-Seng Gan, Nanyang Technological University, Singapore
Pyoung Jik Lee, University of Liverpool, United Kingdom

Correspondence:
Hui Ma
mahui@tju.edu.cn

Specialty section:
This article was submitted to Environmental Psychology, a section of the journal Frontiers in Psychology

Received: *21 April 2020*
Accepted: *25 September 2020*
Published: *29 October 2020*

Citation:
Chen J and Ma H (2020) A Conceptual Model of the Healthy Acoustic Environment: Elements, Framework, and Definition. Front. Psychol. 11:554285. doi: 10.3389/fpsyg.2020.554285

INTRODUCTION

Noise is an important public health issue and is attracting a growing concern since it has negative impacts on human health and well-being (Basner et al., 2015; Dorota et al., 2018). With rapid urbanization, new noise sources (e.g., wind turbines and leisure noise) continue to appear in cities (World Health Organization [WHO], 2018), and the risk of exposure to noise is gradually increasing (Dorota et al., 2018). Long-term noise exposure data have shown that 65% of Europeans living in major urban areas were exposed to daytime noise levels greater than 55 dB and more than 20% of them were exposed to night-time noise levels greater than 50 dB (European Environment Agency [EEA], 2018), which would induce adverse effects, such as ischemic heart disease, cognitive impairment, obesity, and metabolic effects (Clark and Paunovic, 2018; Kempen et al., 2018). Therefore, how to build a healthy acoustic environment against such a background has become a pressing issue for all countries around the world.

To establish a high-quality acoustic environment, countries have chiefly focused on developing laws and regulations related to noise mitigation. An early important attempt on law enactments can be observed in the Noise Control Act of 1972, United States, which aimed to establish an acoustic environment for all Americans, free from noise that jeopardized their health and welfare. Since then, other countries and regions have also enacted laws and regulations (e.g., Ministry of environmental protection of China, 2008), among which the regulations formulated by the European Union have had the greatest impact worldwide. Such efforts were mainly reflected in the Green Paper on Future Noise Policy (European Commission, 1996) and the Environmental Noise Directive (2002). The laws and regulations served to prevent more residents from being exposed to high levels of noise to a certain degree (King and Murphy, 2016). However, the noise regulations are characterized by "passive control," with the purpose of protecting people from adverse effects (Environmental Noise Directive, 2002). With people's increasing requirements for health and a healthy environment (World Health Organization [WHO], 1986, 1991, 2006), whether the current acoustic environment, established under the guidance of "protecting people from being negatively affected," can satisfy people's demands is worthy of further discussion.

Moreover, in order to integrate the associations between environmental noise and health, numerous conceptual models (e.g., Lazarus and Folkman, 1984; Stokols, 1987; Van Kamp, 1990; Rashid and Zimring, 2008) were proposed based on psychological stress theory (Lazarus, 1966). However, these models focused on revealing the impact mechanism of environmental noise on non-auditory health. The specific health dimensions and acoustical indicators that should be considered are still not clear. Therefore, from a holistic perspective, illustrating the specific dimensions and acoustical parameters of a healthy acoustic environment are necessary in order to achieve an overall health. It is worth mentioning that these conceptual frameworks, together with other health-related researches (e.g., Baum et al., 2001; Schabracq, 2003), have laid a theoretical foundation for further study to establish a holistic and practical framework on acoustic environment and health.

Furthermore, the emergence of "soundscape" shifted the concern of acoustic research from the objective acoustic environment to subjective perceptions, and it also extended the research scope from regarding sounds as psychophysical stressors to regarding them as resources (Kang et al., 2016, 2020). Indeed, environmental sounds also have perceptible positive effects rather than negative impacts (Krzywicka and Byrka, 2017; Torresin et al., 2019). For instance, Terhardt and Stoll (1981) developed a descriptor for determining the pleasantness of noise as early as 1981. Axelsson et al. (2010) clearly identified that "pleasantness" was one of the dimensions in the model of perceived affective quality of soundscape. Botteldooren et al. (2006) proposed the embodiment of likeness to music of a soundscape. The aim of the exploring the positive dimensions of soundscapes was to build a high-quality acoustic environment to promote people's health and well-being. To achieve this goal, the association between positive soundscapes and health-related effects was explored. The short-term health effects were reported to be related to physiological indicators, such as skin conductance level, heart rate, respiration rate, electromyogram, cardiovascular response, and saliva cortisol (Annerstedt et al., 2013; Hume and Ahtamad, 2013; Medvedev et al., 2015), while long-term psychophysical effects involved self-reported physical and mental health (Booi and van den Berg, 2012; Shepherd et al., 2013). The results showed that positive soundscapes were associated with a faster stress-recovery process and better self-reported health condition (Alvarsson et al., 2010; Aletta et al., 2018a,b; Park et al., 2020). The findings of soundscape research indicated the possibility to create a healthy acoustic environment. Although former studies have tentatively explored the relationship between positive soundscape and health, and the goal of establishing a healthy acoustic environment has been proposed (Aletta and Kang, 2019), it remains unclear what elements should be considered when we want to build a healthy acoustic environment and what people care about most when mentioning a healthy acoustic environment. The elements of most concern may be core factors in creating a healthy acoustic environment. At the same time, it is also necessary to thoroughly explore the framework and definition of a healthy acoustic environment in order to guide practical work such as policymaking and noise control for the future.

Above all, establishing a healthy acoustic environment is of great importance for ensuring the health of the population as well as promoting sustainable development of the environment. Therefore, through a grounded theory approach, this study aims to (1) explore people's demands for a healthy acoustic environment and present the basic elements thereof, (2) propose a conceptual framework of a healthy acoustic environment, and (3) define the concept of a healthy acoustic environment.

MATERIALS AND METHODS

Grounded Theory (GT) is a sociological approach to discovering theory from data (Glaser and Strauss, 1968). With systematic procedures of data collection and analysis, the GT approach allows for substantial data, in-depth insights, and multidisciplinary participants, and it is useful for elucidating the underlying defined pattern of a certain phenomenon, which is well suited for the establishment of theoretical frameworks in initial research. Although it is a sociological method, GT has been employed, adapted, and refined in a diverse array of fields such as education, social work, and nursing (Strauss and Corbin, 1990; Charmaz, 2014), and emerging studies have proved that it is also an effective way to explore people's understanding of the acoustic environment (Liu and Kang, 2016; Park et al., 2016; Acun and Yilmazer, 2018). Therefore, GT was used to perform this study.

Participants

The principle of data sampling in GT is to select respondents who can provide the most informative insights on the research questions. In order to collect comprehensive and extensive opinions on a healthy acoustic environment, two types of respondents were considered: ordinary residents and professionals. Before the formal interviews, 5 ordinary Chinese residents and 3 acoustic professionals were selected as targets to conduct the interview. The pre-interview mainly involved the semistructured questions of the cognition of a healthy acoustic environment. The preliminary findings showed ordinary Chinese residents seemed to provide more personal feelings based on their daily experience, while the acoustic professionals seemed to be more capable of providing expertise-based opinions, which were all helpful for enriching categories. Therefore, in the formal interview, two types of respondents were all selected. Ordinary Chinese residents were selected and interviewed face to face on streets and in parks, offices, factories, and residential areas in the Beijing–Tianjin–Hebei region. In order to obtain more diverse viewpoints, in addition to acoustic professionals, professionals with the research or education background in medicine science, environment science, sociology, psychology, and architecture were also invited to participate in investigations. Finally, the first type of respondents comprised 44 ordinary Chinese residents (labeled as P01–P44), and the other comprised 31 worldwide professionals (labeled as P45–P75). There were 27 professionals with a research background in acoustics, among whom 3, 3, and 2 professionals had an interdisciplinary research background in sociology, psychology, and environment science, respectively. 3 professionals had a research background in medicine science, and 1 professional had a research background in healthy building. Among the 75 respondents, there were 37 males and 38 females, ranging in age from 23 to 76 years old (average age = 41).

Interview Procedure

To start the investigation, an interview outline was created. As shown in **Table 1**, the interview outline mainly focused on three parts. First, the basic information of each respondent was obtained. Thereafter, the characteristics of a healthy acoustic environment and people's expectation of a healthy acoustic environment were investigated, mainly to determine people's understanding of a healthy acoustic environment. The final part focused on people's opinions on the current noise policy and future policy, to understand people's attitudes toward a healthy acoustic environment. Since ordinary residents have limited knowledge of an acoustic environment, in order to make it easier to start the investigation, two approachable questions were provided before starting the second part of the interview. It should be noted that questions were given as guides only, and additional questions would be added if respondents mentioned significant information. The respondents were encouraged to freely express their opinions relating to a healthy acoustic environment. The investigation was carried out from March to August 2019. The face-to-face interviews lasted from 8 to 30 min. Respondents voluntarily signed informed consent for their involvement in the interview and allowance of audio recording during the face-to-face interview. All respondents were informed of their right to confidentiality, anonymity, and withdrawal from the study at any time. Finally, interviews were organized into transcripts comprising a total of 79,009 words.

TABLE 1 | Interview outline.

Category	Questions
Basic information	Name; age; gender; mainly research fields (for professionals only);
Characteristics and people's demands	How is the acoustic environment in your daily life? (for ordinary residents only) What kind of acoustic environment do you like? (for ordinary residents only) As a researcher/resident, in your opinion, what features should a healthy acoustic environment possess? Please give more words to describe a healthy acoustic environment. What do you hope the healthy acoustic environment can bring for you and your family?
Attitudes	Are you satisfied with the current noise policy in your country or area? Why? Most of the noise standards and noise policies are made to avoid the adverse effects brought by noise. Do you think it is necessary to make those criteria and policies based on the standard of a "healthy acoustic environment" rather than "no harm?"

Data Analysis

The interview transcripts were coded using multistep analysis techniques (Strauss and Corbin, 1990). Data were coded with qualitative analysis software.

Firstly, in the open coding, the verbal transcript data were broken down into labels by searching for key phrases, significant factors, and relations. Labels were then gradually conceptualized and grouped together by comparing their associations and similarities. It was worth noting that data conceptualization was not obtained immediately but developed by repeatedly comparing the labels with each other and with the newly emerging codes. Finally, categories emerged.

In axial coding, the data related to categories were constantly compared, on the one hand, to rationalize the classification of the categories and to develop their subcategories, and on the other hand, to determine how the categories were linked and crosscut. The category was compared with each other to discover any existing associations. By constant comparison, initial relationships among categories were developed, and the embryonic form of the conceptual framework was created. During the final stage of this procedure, based on the relationships identified, the *coding paradigm* (Strauss and Corbin, 1990) was used to further develop the linkages among categories.

*The coding paradigm focuses on specifying a category (**phenomenon**), that is a central idea, an event, or a happening, in terms of the **causal conditions** that give rise to it; the **context** (its specific set of properties) in which it is embedded; the **intervening conditions** that are similar to the context; the **action/interactional strategies** by which it is handled, managed, carried out; and the **consequences** of those strategies (Smyrnova and Kang, 2010).*

Finally, in selective coding, the category that was central to the phenomenon was selected as the core category. All categories related to the core category were integrated to develop a conceptual framework and to refine the theory.

RESULTS

During open coding, 3133 labels were identified from the two different types of respondents. An example of the open coding process is shown in **Table 2**. Verbal data were broken down into labels of a1 to a39, and they were then conceptualized by

TABLE 2 | An example of the open coding process.

Memos		Labels	Conceptualizing data	Conceptualizing data	Categories
(In your opinion, what features should a healthy acoustic environment possess?) P (01): A healthy acoustic environment should cover more natural sounds than human voices and traffic noise.	→	a1: Key phrases: natural sounds a2: Key phrases: less human voices a3: Key phrases: less traffic noise	aa1: The sound sources of a healthy acoustic environment can be natural sounds, music, and chirping sounds from children running or playing (a1, a6, a21, a22). aa2: The sound sources of a healthy acoustic environment should be less human voices, less traffic noise, and no conversation (a2, a3, a16). aa3: The characteristics of a healthy acoustic environment are soothing, steady, quiet, not absolutely quiet, not regular, lively, and harmonious (a4, a5, a8, a14, a15, a20, a26, a27, a34). aa4: A healthy acoustic environment should make people relaxed, comfortable, and pleasant (a7, a23, a32, a33, a38, a39). aa5: A healthy acoustic environment should not make people scared (a9). aa6: A healthy acoustic environment is the matching result between the acoustic environment and people's demands. If the characteristics of the acoustic environment (e.g., soothing, quiet, and steady) can have positive or promotive effects on people's demands (e.g., relaxation and sleep well), it is a healthy acoustic environment (a10, a24, a29, a37). aa7: A healthy acoustic environment is the matching result between the acoustic environment and people's demands. If the acoustic environment (too quiet) have negative effects on people's demands (security), it is an unhealthy acoustic environment (a11). aa8: With different activities (work, go out for a walk, sleep at night), people have different kinds of demands (ensure work efficiency; relax; sleep well). So they desire different characteristics (quiet, no regular voices, no conversation; lively, music of square dancing, chirping sounds from children running or playing; quiet, not completely silent; a13, a18, a19, a24, a25, a29). aa9: The healthy acoustic environment should meet people's implicit behavioral demands: work and sleep (a17, a28). aa10: A healthy acoustic environment is the matching result between the acoustic environment and people's demands. If the characteristics of the acoustic environment (e.g., quiet, no regular voices, and no conversation) will not cause negative effects on people's demands (e.g., work efficiency), it is a healthy acoustic environment (a18, a36). aa11: A healthy acoustic environment should meet the limits values of the noise guidelines (a30). aa12: Both layers (limits in noise guidelines and people's psychological demands) should be used to measure a healthy acoustic environment, and people's psychological demands are more important than noise guidelines (a30, a31, a35).	Aa1: Sound sources (aa1, aa2) Aa2: Perceived characteristics of the acoustic environment (aa3) Aa3: Psychological demands (aa4, aa5, aa12) Aa4: Positive/promotive effect—matching—healthy (aa6) Aa5: Negative effect—mismatching—unhealthy (aa7) Aa6: Behavioral states (aa8) Aa7: Behavioral demands (aa9) Aa8: No negative effect—matching—healthy (aa10) Aa9: Standards (aa11) Aa10: Fit standards and meet people's psychological demands (aa12)	A1: Sound sources and acoustic environment (Aa1 Aa2) A2: People's demands (Aa3, Aa7, Aa10) A3: Matching process (Aa4, Aa5, Aa8) A4: Context (Aa6) A5: Criteria and standards of a healthy acoustic environment (Aa9) A6: Secondary fitting process (Aa10) A7: Acoustic environment quality (Aa4, Aa5, Aa8)
P (02): It should be soothing or quiet sounds, such as music with a steady rhythm, which can make me relaxed. But it is not absolutely quiet. If the environment is too quiet, it makes me feel scared, so it is unhealthy.	→	a4: Key phrases: soothing a5: Key phrases: quiet a6: Key phrases: music with steady rhythm a7: Key phrases: make me relaxed a8: Key phrases: not absolutely quiet a9: Key phrases: It shouldn't make me feel scared. a10: Relation: soothing, quiet sounds, music with steady rhythm - relaxed - healthy a11: Relation: too quiet - scared - unhealthy			
P (32): In my opinion, a healthy sound environment is diverse. When I work, it is quiet and it is necessary to have no regular voices or conversation to ensure my work efficiency; When I take my child out for a walk in the evening, the healthy sound environment is lively and it is better with music of square dancing and chirping sounds from children running or playing, so I can relax completely; A healthy acoustic environment at night is quiet, but not completely silent, which makes me sleep well.	→	a12: Key phrases: diverse a13: Significant factors: work a14: Key phrases: quiet a15: Key phrases: no regular sounds a16: Key phrases: no conversation a17: Key phrases: ensure work efficiency a18: Relation: work - quiet, no regular voices, no conversation - ensure work efficiency - healthy a19: Significant factors: take child out for a walk a20: Key phrases: lively a21: Key phrases: music of square dancing a22: Key phrases: chirping sounds from children running or playing a23: Key phrases: relax a24: Relation: go out for a walk - lively, music of square dancing, chirping sounds from children running or playing - relax - healthy a25: Significant factors: sleep at night a26: Key phrases: quiet a27: Key phrases: not completely silent a28: Key phrases: make me sleep well a29: Relation: sleep - quiet, not completely silent - sleep well - healthy			
P (66): Personally, I think a healthy acoustic environment should first meet the upper limits of the noise guidelines set by the World Health Organization. In addition, it should meet people's subjective psychological demands, such as comfortable, pleasant, and relaxed. I think the latter is more important than the former. Finally, a healthy sound environment should not only protect people's demands from being negatively affected, but also take on a catalytic or promotive role, such as keeping one's pleasure and evoking positive emotions.	→	a30: Key phrases: meet upper limits of the noise guidelines a31: Key phrases: meet people's subjective psychological demands a32: Key phrases: comfortable a33: Key phrases: pleasant a34: Key phrases: harmonious a35: Key phrases: people's psychological demands are more important than noise guidelines. a36: Relation: acoustic environment - people's demands (not being negatively affected) - healthy a37: Relation: acoustic environment - people's demands (take on a catalytic or promotive role) - healthy a38: Key phrases: keep one's pleasure a39: Key phrases: evoke positive emotions			

comparing their associations and similarities. The labels a1, a6, a21, and a22 described the sound sources in a healthy acoustic environment, and were thus integrated into the concept aa1. This concept aa1 was then further conceptualized into "sound sources" (Aa1) together with a similar concept aa2. The labels a4, a5, a8, a14, etc. described the characteristics of a healthy acoustic environment, so they were grouped together as aa3 and then integrated into "perceived characteristics of the acoustic environment" (Aa2). Then, Aa1 and Aa2 were grouped into the category "sound sources and acoustic environment" (A1). Similarly, relation labels were also developed with the same procedures. For example, a10, a11, a18, a24, etc. described the relation between the acoustic environment and people's demands. Therefore, the relation was defined as a matching relation (aa6, aa7, aa10) during the original conceptualization process. Then, the relation was further refined to "positive/promotive effect–matching–healthy" (Aa4), "negative effect–mismatching–unhealthy" (Aa5), and "no negative effect–matching–healthy" (Aa8) respectively in the second stage of the conceptualization process, and eventually they evolved into the category "matching process" (A3). Finally, with a similar coding process, seven categories were identified, as follows: "sound sources and acoustic environment" (A1), "people's demands" (A2), "matching process" (A3), "context" (A4), "criteria and standards of a healthy acoustic environment" (A5), "secondary fitting process" (A6), and "acoustic environment quality" (A7).

In axial coding, on the one hand, subcategories of each category and dimensions of each subcategory were developed. For example, "sound sources" and "perceived characteristics of the acoustic environment" were developed as two subcategories of "sound sources and acoustic environment." Based on the conceptual data (aa3), dimensions of "perceived characteristics of the acoustic environment" were also developed; they were "characteristics of auditory sensation" and "characteristics of auditory perception." All codes are shown in **Figure 1**. The categories are presented in the gray-filled boxes, while their subcategories, dimensions, and key points are shown in other boxes below, where the subcategories are presented in bold, and dimensions are presented in bold and italics, combined with their key points listed only by bullet points. Key points were directly integrated by labels of key phrases, significant factors, and relations, while dimensions were integrated by key points. It was necessary to note that some key points were further dimensionalized. For example, the key point "quiet" fell under the dimension of "sound exposure level" and "characteristics of auditory sensation," while "sound exposure level" and "characteristics of auditory sensation" were under the subcategory of "characteristics of the acoustic environment." On the other hand, the *coding paradigm* (Strauss and Corbin, 1990) was used to develop relationships among categories. To illustrate coding results clearly, respondents' direct quotations, which were all listed in **Table 3** except for special notes, were included. A detailed explanation of categories and the causal inference among them are presented in the following section "Elements of the Healthy Acoustic Environment."

In selective coding, a conceptual framework (**Figure 2**) was finally created, which reflected the defined pattern of a healthy acoustic environment. The conceptual framework and

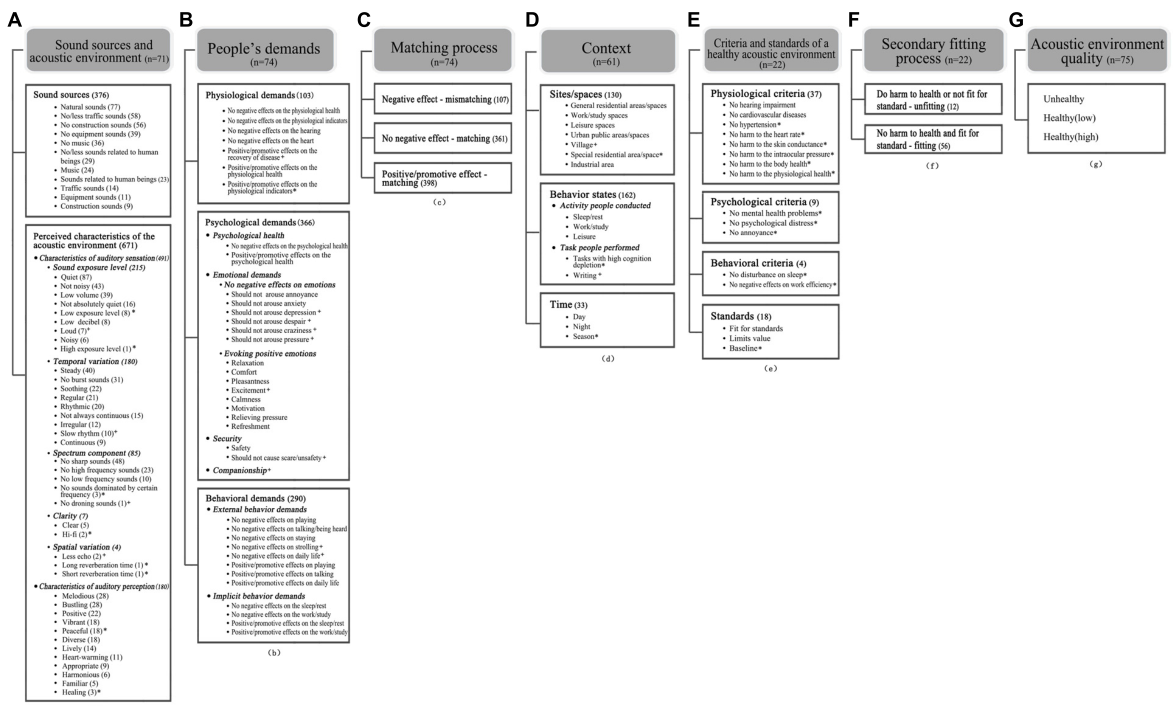


FIGURE 1 | Categories, subcategories, corresponding dimensions, and key points created in the open coding process.

TABLE 3 | Respondents' quotations.

P01	People have different tastes. So it is better to have no music in a healthy acoustic environment.
P02	When I am sleeping, it is quiet, and there are no loud or regular sounds. If I can sleep well, it is a healthy acoustic environment.
P03	I used to live near the railway. I could not stand the rumbling noise. When the train passed, it startled me. I was annoyed and despairing. A healthy acoustic environment should not be annoying and anxiety-provoking.
P06	When I am doing high cognitive tasks, the droning sounds from the air-conditioner and chatting sounds from colleagues always distract me. These sounds should not appear in a healthy acoustic environment. It has a serious negative impact on our work efficiency.
P08:	The acoustic environment in my office is unhealthy. When I want to concentrate on my work, there is always chatting from colleagues and traffic sounds from outside. Therefore, I cannot focus my attention entirely on my work.
P11	When I am shopping in the mall, I cannot feel the sounds from the equipment at all. Even if I could hear them, the acoustic environment with equipment sounds and human voices should be lively and exciting. This is also a healthy acoustic environment, in my opinion.
P13	I really don't want to hear traffic sounds, but I can't make the road disappear, and I can't change my place of work either. In my opinion, at an appropriate volume, the acoustic environment with traffic noise can also be a healthy acoustic environment.
P15	Once at the Convention Center, I talked to my client about cooperation. It was noisy. He couldn't hear me clearly and I needed to raise my voice to make myself heard. I was anxious. In my opinion, a healthy acoustic environment should have positive effects on our behavior, and it should not affect our talking and emotions.
P16	Noise has become a disaster that threatens our health because some negative effects have been proved. Maybe our bodies have been damaged before we know it, which is terrible. I do not want such things to happen to us. A healthy acoustic environment must protect our bodies from being negatively affected.
P24	I am often at home alone, so I would like to listen to some music. Even if I don't listen, this sound is always with me. I needed to be accompanied, and it did that. In my opinion, it is a healthy acoustic environment.
P25	In parks or other public places, if the environment is too quiet, it is frightening, and it makes people feel insecure. A healthy acoustic environment should provide a sense of security for us, especially in an empty and silent environment.
P33	A healthy acoustic environment should be different from the general acoustic environment. It should promote people's health. I think I'm in a healthy acoustic environment now, with birds' singing, people chatting, and laugher. I take my wife here with the hope that she can recover soon. I think a healthy acoustic environment should be able to help disease recovery.
P36	In my opinion, it is necessary to follow the current policy when establishing a healthy acoustic environment, and no disturbance in our daily life is also necessary. In particular, cars should not be allowed to whistle in the community.
P54	No effect of noise is impossible. Even if noise is not heard, people claim to be annoyed. So, some baseline for noise effects, for example, the WHO guidelines, is necessary. In order to create a healthy acoustic environment, it is necessary not only to avoid an unhealthy environment but also to preserve a pleasant environment.
P59	First of all, a healthy acoustic environment should not cause hearing impairment, or physical and mental health problems. Additionally, it should be a positive acoustic environment, and have a positive effect on people's health, like positive soundscapes.
P62	A healthy acoustic environment depends on the site. I used to study the acoustic environment in hospitals. Quiet is necessary in a hospital because the patient needs to rest for recovery. If the acoustic environment interferes with patients' recovery, it's unhealthy. While a healthy acoustic environment in an urban park should be lively, pleasant, relaxing, and stress-relieving. I think the latter acoustic environment can be defined as a healthy acoustic environment in a broad sense. As x has studied, the acoustic environment in the classroom with a restorative effect on children's attention can also be treated as a healthy acoustic environment.
P67	As a first step, healthy acoustic environments are those that "do not cause harm to health." This could be either physiological (e.g., cardiovascular, etc.) or simply psychological distress. When adverse health effects have been addressed and excluded, the second layer comes into play, which is about a "supportive" acoustic environment, i.e., those that do not only "permit," but basically "promote" well-being and quality of life.

the definition of a healthy acoustic environment are presented in section "Conceptual Framework and Definition of the Healthy Acoustic Environment."

Elements of the Healthy Acoustic Environment

Sound Sources and Acoustic Environment

The central idea (*phenomenon*) of this research could be labeled as the judgment of the acoustic environment quality. The category of "sound sources and acoustic environment" could be divided into two subcategories: "sound sources" and "perceived characteristics of the acoustic environment." The sound sources were the basis of people's perception, while gradual perception and interpretation of the acoustic environment were necessary conditions to motivate the judgment of the acoustic environment quality. Therefore, "sound sources and acoustic environment" was the first *causal condition* that gave rise to the *phenomenon*.

The "sound sources and acoustic environment" consisted of key points related to the characteristics of a healthy acoustic environment. To illustrate the characteristics of a healthy acoustic environment clearly, the frequency of labels was taken into consideration.

The "sound sources" were supported by six types of sound sources, which were natural sounds, traffic sounds, construction sounds, equipment sounds, music sounds, and sounds related to human beings. As listed in **Figure 1A**, the sound sources of a healthy acoustic environment can be diversified and have many manifestations. The frequencies of labels showed that people preferred natural sounds to traffic sounds, construction sounds, and equipment sounds in a healthy acoustic environment. Conflicting views were observed on music sounds and sounds related to human beings. Some respondents mentioned that music should be noted in a healthy acoustic environment while others showed a negative attitude toward music because it was difficult to find a music genre that everybody would appreciate (e.g., P01). It was also observed that sound sources in a healthy acoustic environment were closely related to people's behavior states. For example, when people were at work with high cognition depletion, human voices in the environment

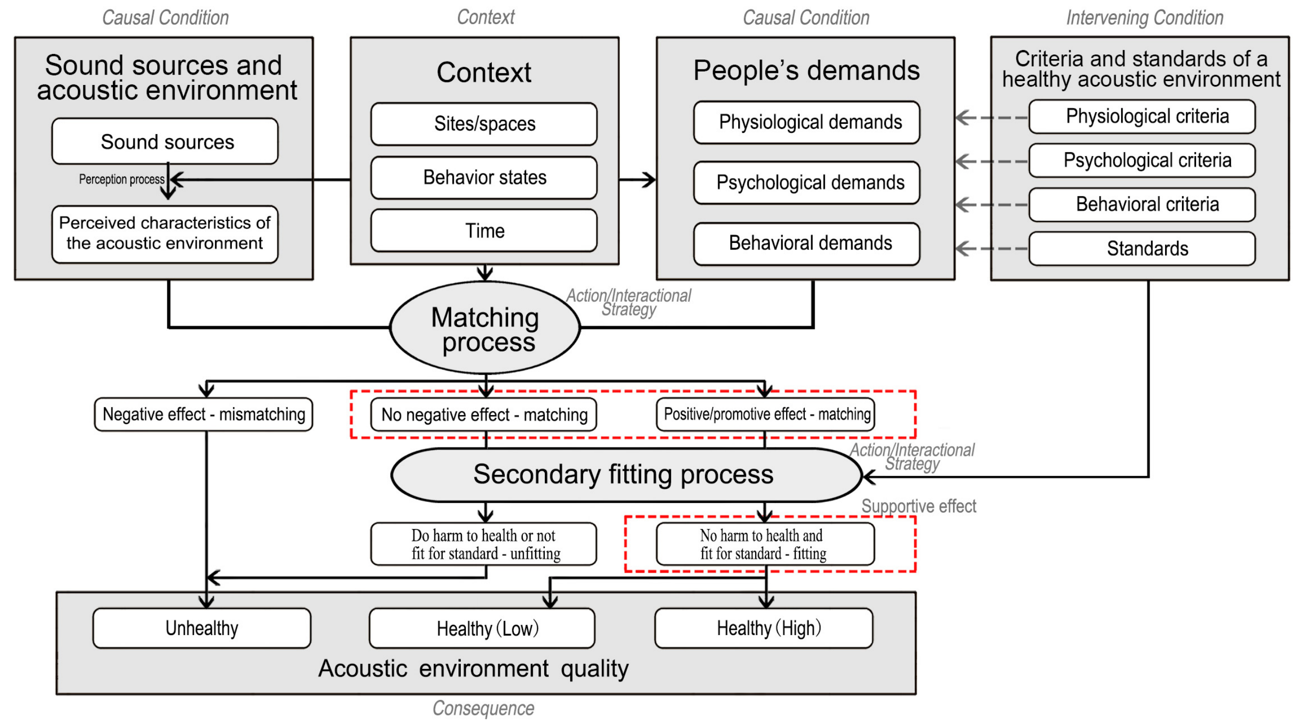


FIGURE 2 | The conceptual framework of a healthy acoustic environment.

could not be accepted (e.g., P06). However, when people were at leisure, human voices in the environment were considered lively and positive (e.g., P11). Interestingly, different opinions on traffic sounds, equipment sounds, and construction sounds were also found in this study, although those sounds were generally perceived as negative sound sources. As P13 mentioned, it was unrealistic to establish an acoustic environment that did not contain any traffic sounds; thus, some sound sources that people did not like could also be tolerated in a healthy acoustic environment if the sound volume was controlled at an appropriate level. Therefore, although natural sounds were preferred most, other sound categories could also be acceptable in a healthy acoustic environment, and this depended on specific context, such as people's preferences, people's behavior states, and some realistic conditions.

The "perceived characteristics of the acoustic environment" could be divided into two dimensions according to people's degree of interpretation of the acoustic signal (ISO 12913-1, 2014). These were "characteristics of auditory sensation" and "characteristics of auditory perception." "Characteristics of auditory sensation" consisted of the direct and preliminary descriptive words mentioned by people (e.g., *quiet* and *not noisy*). According to respondents' descriptive words, five parameters of the acoustic environment were identified: sound exposure level, temporal variation, spectrum component, clarity, and spatial variation. As shown in **Figure 1A**, the first prominent characteristic of a healthy acoustic environment was low exposure level, because labels of *quiet*, *not noisy*, *low volume*, *low exposure level*, and *low decibel* were mentioned much more than the others labels related to sound exposure level. In addition, mild temporal variation was another characteristic of a healthy acoustic environment, which were mainly supported by labels of *steady*, *no burst sounds*, and *soothing*. Other characteristics, such as *regular*, *rhythmic*, and *not always continuous*, also seemed to be related to temporal variation of sounds. Thus, temporal variation of sounds should be considered in a healthy acoustic environment. Moreover, participants frequently mentioned that the sounds in a healthy acoustic environment should *not be sharp*. Labels of *no high frequency sounds*, *no low frequency sounds*, *no sounds dominated by certain frequency*, and *no droning sound* were also mentioned by participants. Therefore, sounds obviously dominated by certain spectrum components should be avoided in a healthy acoustic environment. Lastly, *clear*, *hi-fi*, *less echo*, *long reverberation time*, and *short reverberation time* were also mentioned by a few respondents, suggesting that clarity and spatial variation of sounds might also be parameters that should be considered in a healthy acoustic environment. "Characteristics of auditory perception" consisted of descriptive words that focused on further interpretation of the "characteristics of auditory sensation" in context, such as *melodious*, *positive*, and *bustling*. As shown, the descriptive words contained more personal positive emotions, which reflected people's positive expectation on a healthy acoustic environment.

People's Demands

By analyzing the verbal data of the interview, it was observed that when respondents referred to a healthy acoustic environment, their demands were unconsciously mentioned, such as *sleep well* (P02) and *ensure work efficiency* (P32, in **Table 2**, the first column). As described by respondents, the expected characteristics of a healthy acoustic environment depended on people's demands. If their demands were met by the acoustic environment, the acoustic environment was judged as healthy. Therefore, "people's demands" was regarded as the other *causal condition* that gave rise to the judgment of the acoustic environment quality *(phenomenon)*. In this study, "people's demands" consisted of three subcategories: physiological, psychological, and behavioral demands.

Physiological demands were composed of descriptive words related to physiological health. During the interview, because respondents did worry that the acoustic environment would negatively impact their physiological health, they expected that a healthy acoustic environment could protect them from being negatively affected (e.g., P16); thus, the key points of *no negative effects on the physiological health, no negative effects on the hearing etc.* were mentioned. In addition, some respondents also mentioned that a healthy acoustic environment should not only protect people's physiological health from being negatively affected but also have a positive or promotive effect on people's physiological health. For example, the key point of *positive effects on the recovery of disease* was mentioned by P33. All the key points mentioned by respondents are listed in **Figure 1B**.

Psychological demands consisted of demands for psychological health, emotion, companionship, and security. As listed in **Figure 1B**, demands for psychological health were supported by key points of *no negative effects on the psychological health* and *positive/promotive effects on the psychological health*. In addition to the general demands on psychological health, people were used to adopting a specific negative or positive emotional change to evaluate whether the acoustic environment was healthy (e.g., P03). On the one hand, some respondents (e.g., P03, P15) mentioned past experiences of negative emotions caused by the acoustic environment, so they held the opinion that a healthy acoustic environment should not arouse negative emotions (e.g., *annoyance*, *anxiety*, and *depression*). On the other hand, respondents also mentioned (e.g., P62; P66, in **Table 2**, the first column) that a healthy acoustic environment should evoke positive emotions (e.g., *relaxation*, *comfort*, and *pleasantness*). Moreover, some respondents mentioned demands for security (e.g., P25) and companionship (e.g., P24) when referring to a healthy acoustic environment. Demand for security was composed of descriptive phrases of *safety* and *should not cause scare/unsafety*, while demand for companionship was composed of descriptions of people's expectations to be accompanied by sounds.

Behavioral demands were also mentioned by respondents, and they could be divided into external behavioral demands and implicit behavioral demands according to Watson's (1913) behavioral psychology. External behaviors could be directly observed, such as *playing* and *talking*. Compared to external behaviors, implicit behaviors could only be observed with the help of equipment or experiments. In this study, the implicit behavior demands included demands on *sleep*, *rest*, *work*, and *study*. The former two demands should be a concern in situations where people need to rest, and the latter two in cases where people need to focus attention and thinking. Consistent with physiological and psychological demands, behavioral demands also had two layers: no negative effects and positive/promotive effects. All the key points mentioned by respondents are listed in **Figure 1B**.

Matching Process

According to the respondents' description logic, the health of an acoustic environment was depended on whether the "perceived characteristics of the acoustic environment" could meet "people's demands." For example, as P02 described, if the perceived acoustic environment (*quiet*, *no loud or regular sounds*) could meet the demands (*sleep well*), that acoustic environment would be evaluated as healthy. In contrast, as P08 described, if the perceived acoustic environment (*colleague chatting sounds and traffic sounds*) could not meet the demand (*focus attention on work*), it would be judged as unhealthy. Therefore, the relation between "sound sources and acoustic environment" and "people's demands" was gradually conceptualized as a matching relation (aa6, aa7, and aa10 in **Table 2**, the third column). Matching process reflected the process by which the judgment of the acoustic environment quality (*phenomenon*) was handled and carried out. Therefore, the matching process was considered as *action/interactional strategy* in terms of the *coding paradigm*.

The "matching process" was supported by 74 samples in this paper, and it contained three subcategories with a causal relationship. The subcategories were the cognitive outputs of "matching process," namely "negative effect—mismatching," "no negative effect—matching," and "positive/promotive effect—matching." "Negative effect—mismatching" was supported by relation labels related to the "negative effects" of the acoustic environment (e.g., a11 in **Table 2**, the second column), while "no negative effect—matching" and "positive/promotive effect—matching" were, respectively supported by relation labels that were related to "no negative effects" (e.g., a18 and a36 in **Table 2**, the second column) and "positive/promotive effects" of the acoustic environment (e.g., a10, a24, a29, and a37 in **Table 2**, the second column).

Context

In the interviews, people used to link their demands on a healthy acoustic environment and the characteristics of a healthy acoustic environment to the context. For example, as described, during *working*, *leisure*, and *sleeping* (different behavior states), the demands of P32 were to *ensure working efficiency*, *relax*, and *sleep well*, respectively, and expected characteristics of the acoustic environment correspondingly were *quiet*, *lively*, and *quiet, but not completely silent*, which showed in different "behavior states," people had different types of demands and the characteristics that could match their demands were also different. If the characteristics of the acoustic environment could meet people's demands in a specific behavior state, the acoustic environment was evaluated as healthy. The example suggested that "people's demands," "sound sources and acoustic environment," and the "matching process" were all embedded in the "behavior states." Therefore, "behavior states" could be regarded as the *context* to the judgment of the acoustic environment quality (*phenomenon*). Similarly, the judgment was also embedded in "site/space" (e.g., P62) and "time" (e.g., P32).

The "context" was supported by 61 samples in this study, and it contained three subcategories: "sites/spaces," "behavior states," and "time." "Sites/spaces" was composed of the specific site or space that people mentioned when referring to a healthy acoustic environment. All the sites and spaces mentioned by people (e.g., *residential areas*, *office*, *urban park*, and *karaoke bar*) were integrated into *general residential areas/spaces*, *work/study spaces*, *leisure spaces*, etc. "Behavior states" consisted of key points related to "the activity people conducted" (e.g., *sleep*, *rest*, and *work*) and "the task people performed" (e.g., *tasks with high cognition depletion* and *writing*). "Time" was supported by key points of *day*, *night*, and *season*. All the key points are shown in **Figure 1D**.

Criteria and Standards of a Healthy Acoustic Environment and Secondary Fitting Process

"Criteria and standards of a healthy acoustic environment" could be regarded as an *intervening condition* of the judgment of the acoustic environment quality (*phenomenon*) since it regulated cognitive outputs of the "matching process." Under the supplements of "criteria and standards of a healthy acoustic environment," the outputs of the "matching process," namely, "no negative effect—matching" and "positive/promotive effect—matching," were judged once more in order to exclude acoustic environments that did harm to people's health or that were not fit for standards. Therefore, "criteria and standards for a healthy acoustic environment" can also be considered supplements to "people's demands."

The process of supplemental measurement was called the "secondary fitting process," and it contained two subcategories with a causal relationship: "do harm to health or not fit for standard—unfitting" and "no harm to health and fit for standard—fitting." "Secondary fitting process" also reflected the process by which the judgment of the acoustic environment quality (*phenomenon*) was handled and carried out, which had a similar role to the "matching process." Thus, it was also considered as *action/interactional strategy*.

In this paper, the "criteria and standards for a healthy acoustic environment" was supported by four subcategories: "physiological criteria," "psychological criteria," "behavioral criteria," and "standards." "Physiological criteria," "psychological criteria," and "behavioral criteria" consisted of evidence-based descriptive phrases related to the effects of the acoustic environment on physiology, psychology, and behavior, such as *no hearing impairment* (e.g., P59), and *do not cause harm to health* (e.g., P67). "Standards" consisted of descriptive phrases related to standards, policies, or guidelines, such as *follow the policy* (e.g., P36), and *limits value of the noise guideline* (e.g., P66 in **Table 2**, the first column). All the key points mentioned by respondents are shown in **Figure 1E**.

Acoustic Environment Quality

According to respondents' description, the acoustic environment quality was divided into three levels in this paper, namely "unhealthy," "healthy (low)," and "healthy (high)." They were the final *consequences* of the judgment of the acoustic environment quality.

Conceptual Framework and Definition of the Healthy Acoustic Environment

Based on the interpretation of each element, a conceptual framework of a healthy acoustic environment was developed to illustrate the relationships among the seven elements. As shown in **Figure 2**, "sound sources and acoustic environment," "context," "people's demands," "criteria and standards of a healthy acoustic environment," and "acoustic environmental quality" are shown in gray-filled square boxes and their subcategories in rounded boxes inside, while "matching process" and "secondary fitting process" are shown in gray-filled elliptical boxes, and their subcategories in rounded boxes below. As interpreted in section "Elements of the Healthy Acoustic Environment," the central idea (*phenomenon*) of this research could be labeled as the judgment of the acoustic environment quality, while the "matching process" and "secondary fitting process" reflected the processes by which the judgment of the acoustic environment quality was handled and carried out. Therefore, these two processes could be used to connect all the other categories. Based on the associations among these categories, the defined pattern of a healthy acoustic environment was identified.

In a specific "site/space," "time," and "behavior state," people had specific "demands" on physiology, psychology, and behavior. If the "sound sources and acoustic environment" had a negative effect on "people's demands," the acoustic environment mismatched "people's demands" and the output of "matching process" was "negative effect—mismatching." Thus, the acoustic environment was directly judged as "unhealthy." If the acoustic environment did not have a negative effect on "people's demands" or had a positive/promotive effect on "people's demands," the acoustic environment matched "people's demands" successfully. Thus, the outputs of "matching process" were "no negative effect—matching" or "positive/promotive effect—matching." Then, the "criteria and standards of a healthy acoustic environment" came into play to measure the outputs complementally in order to exclude acoustic environments that did harm to people's health or that were not fit for standards. Finally, if the acoustic environment has "negative effect" on people's demands or "does harm to people's health or not fits for standard," it will be judged as "unhealthy." If the acoustic environment has "no negative effect" on people's demands and "does no harm to people's health and fits for standard," it will be judged as "healthy (low)." If the acoustic environment has "positive/promotive effect" on people's demands and "does no harm to people's health and fits for standard," it will be judged as "healthy (high)." It should be mentioned that the acoustic environment with the characteristics of "no harm to health and fit for standards—fitting" and "no negative effect—matching" or "positive/promotive effect—matching," as shown in **Figure 2** with red dotted boxes, are consistent with the goals of a "supportive environment," which suggests that the resources in the physical or social environment should meaningfully impact on people's body, feelings, behaviors, and health (World Health Organization [WHO], 1991; Wagemakers et al., 2010; Jiang and Shen, 2018). Therefore, a healthy acoustic environment can be defined as a supportive acoustic environment that can match people's physiological, psychological, and behavioral demands in context, and that also fits the criteria and standards.

DISCUSSION

Regarding the Elements and Conceptual Framework of a Healthy Acoustic Environment

Associations between environmental sounds and negative health outcomes (Basner et al., 2015; Dorota et al., 2018) or positive effects (Alvarsson et al., 2010; Aletta et al., 2018a,b) have been investigated by researchers and institutions worldwide over the past decades. These works have made a great contribution to revealing the negative or positive effects of environmental sounds on people's health. However, it remains unclear what elements should be considered and what people care about most when we want to build a healthy acoustic environment. With a grounded theory approach, this study explored the elements of a healthy acoustic environment. The proposal of these elements, together with their subcategories and dimensions, provided an opportunity for subsequent research on a healthy acoustic environment in a specific context.

Based on the associations among these elements, a conceptual framework of a healthy acoustic environment was developed. Compared with previous studies, the conceptual framework of a healthy acoustic environment is a framework with comprehensive considerations of acoustical parameters and people's demands, and with wide applicability in context. Previously, studies either focused on examining the associations between acoustical environmental factors and a specific health outcome, such as stress (Rashid and Zimring, 2008), adverse birth outcomes (Nieuwenhuijsen et al., 2017), hearing loss and tinnitus (Sliwinska-Kowalska and Zaborowski, 2017), annoyance (Guski et al., 2017), sleep (Basner and McGuire, 2018; Meng et al., 2020b), and the cardio-metabolic system (Van Kempen et al., 2018), or focused on exploring the associations between health outcome and the acoustic environment with specific sound sources or specific characteristics, such as transport noise (Brown and van Kamp, 2017; Kempen et al., 2018; Van Kempen et al., 2018), wind turbine noise (Seltenrich, 2019), occupational noise (Themann and Masterson, 2019), low-frequency sounds (Abbasi et al., 2018), and high-frequency sounds (Fletcher et al., 2018). In addition, some integrated frameworks (e.g., Lazarus and Folkman, 1984; Stokols, 1987; Van Kamp, 1990; Rashid and Zimring, 2008) had been proposed and compared (e.g., Lercher, 1996). They mainly focused on the impact mechanism of environmental noise on health, and specific health dimensions were still not clear when we wanted to build a healthy acoustic environment. It still seems a challenge to clearly identify many-to-many relationships between specific acoustical environmental factors and specific health outcomes in practice and to construct an appropriate framework to guide such research. Despite the many challenges in identifying such complex relationships, it is extremely important for research to unravel such complexities if overall health is to be obtained (Zhang et al., 2019). The holistic conceptual framework of a healthy acoustic environment proposed in this research aims to support a movement in this direction.

The Definition and Significance of a Healthy Acoustic Environment

In our study, a healthy acoustic environment is defined as a supportive acoustic environment that can match people's physiological, psychological, and behavioral demands in context and that also fits criteria and standards. It can be seen that there are two key elements in assessing a healthy acoustic environment: "people's demands" and "criteria and standards of a healthy acoustic environment." Further revelation on these two elements will contribute to understanding the connotation of a healthy acoustic environment.

In terms of "people's demands," although the physiological, psychological, and behavioral demands determined in this study were not new and most of their dimensions have been considered and studied in former research (e.g., Andringa and Lanser, 2013; Darvishi et al., 2019; Fredriksson et al., 2019; Waye et al., 2019), some key points of psychological demands (e.g., demands on emotion, security, and companionship) and behavioral demands (e.g., demands on cognition) were first defined as terms related to health, suggesting that people's demands for a healthy acoustic environment were not only confined to their physiological health but extended to a wider scope. Therefore, a healthy acoustic environment could be considered as a demand-oriented definition rather than being a narrow-health-oriented concept. The results support the definition of "health" from the perspective of the "acoustic environment" provided by WHO that "health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (World Health Organization [WHO], 2006). Moreover, the results encourage and support future research related to health outcomes provided by soundscapes.

Moreover, interestingly, as shown in **Figure 1B**, frequency of labels showed that psychological demands were mentioned most, closely followed by behavioral demands, while physiological demands were mentioned least. It seems that in a healthy acoustic environment, people are more concerned about their psychological feelings and behavioral demands than physiological demands. Quantitative research with large samples is needed for further verification, but it is significant for policymakers and researchers to pay sufficient attention to people's psychological and behavioral demands in a healthy acoustic environment.

This research has also revealed that a healthy acoustic environment should provide supportive effects on people's physiology, psychology, and behavior rather than only protect people from being negatively affected. In fact, the positive effects of the acoustic environment on people's physiological indices, psychological feelings, and behavioral responses have been observed in previous research. Clear-cut evidence suggests that interacting with nature sounds could evoke a reduced skin conductance level (Alvarsson et al., 2010), with a similar tendency observed for heart rate (Hume and Ahtamad, 2013; Medvedev et al., 2015). The restorative effects of positive soundscapes on people's psychological experience (e.g., Herranz-Pascual et al., 2019; Meng et al., 2020a; Shu and Ma, 2020) and cognition aspects (e.g., Zhang et al., 2017; Gill et al., 2018; Shu and Ma, 2019) have been reported by many researchers, and the increased possibility of positive behavior triggered by the acoustic environment was also observed (Chen and Ma, 2019). Although part of the promotive effects make sense after stress induction, the evidence observed in studies also supports the restorative benefits and potential promotive effects of the acoustic environment on people's physiological indices, psychological feelings, and behavioral responses. Whether the acoustic environment has a broader catalytic effect needs to be verified in further empirical research, to which sufficient attention should be paid in the future.

The "criteria of a healthy acoustic environment" consisted of four subcategories in this paper: "physiological criteria," "psychological criteria," "behavioral criteria," and "standards." It is worth noting that this study only defined the "criteria and standards of a healthy acoustic environment" within a limited scope, because, respondents held limited evidence on specific criteria and standards despite some of them being professionals. Furthermore, this study indicated that the "criteria and standards of a healthy acoustic environment" need to be systematically combed with the specific context because the judgment of a healthy acoustic environment is embedded in "sites/spaces," "behavior states," and "day-night." Therefore, the saturation of "criteria and standards of a healthy acoustic environment" requires a systematic review based on the specific context. The aim of this study is to explore the overall framework of a healthy acoustic environment. Thus, detailed contents of the category in the framework need to be supplemented by follow-up research.

Comparison of the Codes Between Ordinary Residents and Professionals

To collect comprehensive and extensive opinions on a healthy acoustic environment, two types of respondents were selected in this study: ordinary residents and professionals. The verbal data from these two types of respondents were coded together because their understandings of the healthy acoustic environment were all necessary to develop the saturated categories and an integrated framework. Based on their diverse opinions, the elements and conceptual framework of the healthy acoustic environment were proposed, and the connotation was also identified. It was also meaningful to highlight the different opinions on a healthy acoustic environment between ordinary residents and professionals to further understand the connotation of a healthy acoustic environment. Therefore, their verbal data were later coded separately. As shown in **Figure 1**, the codes only mentioned by professionals were marked with "*," while the codes only mentioned by ordinary residents were marked with " + " and the codes mentioned by both types of respondents were not marked.

The results showed that professionals provided more opinions based on their expertise. This is mainly reflected in three points. Firstly, more terminology was mentioned by professionals, such as *low/high exposure level*, *hi-fi*, and *long/short reverberation time* (**Figure 1A**). Secondly, it seemed it was easier for professionals to give a relatively complete and systematic evaluation system (e.g., P54, P59, P62, and P67) than for ordinary residents. Thirdly, professionals contributed more diverse and evidence-based key points to the category "criteria and standards of a healthy acoustic environment," such as *no hypertension*, *no harm to heart rate*, and *no harm to skin conductance*, as observed in **Figure 1E**, which enriched the empirical evidence for a healthy acoustic environment. Compared with professionals, ordinary residents were more likely to provide key points from their experiences (e.g., P03, P11, and P24); thus, many key points related to their feelings were mentioned, such as *should not arouse depression*, *excitement*, and *should not cause scare/unsafety*, as observed in **Figure 1B**. These differences are likely to lead to different priorities in the framework of a healthy acoustic environment.

However, the aims of this study are to explore the elements, conceptual framework, and definition of a healthy acoustic environment for all people. Although some differences could be observed between the two groups of respondents, seven elements of the healthy acoustic environment were all mentioned by both groups of respondents and all the key points were essential parts to make up the framework. A complete framework covering all respondents' opinions, whether they are professionals or ordinary residents, seems to be more meaningful for all people than two separate conceptual models. Therefore, an integrated framework of a healthy acoustic environment was proposed in this study.

Practical Value to Acoustic Environmental Policy

Although countries have previously established noise guidelines and policies regarding different areas and different human activities (e.g., Environmental Noise Directive, 2002; Ministry of environmental protection of China, 2008), most are based on some specific health dimensions (e.g., annoyance and sleep) while the holistic perspective is lacking in policymaking. The results of this study show that systematic consideration of people's demands is necessary for a healthy acoustic environment, which supports and promotes the rationalization of noise policy and lays the foundation for establishing the standards of a healthy acoustic environment in future.

In addition, current noise policy is established under the guidance of "no negative effect." The results revealed that with people's increasing requirements in relation to health and healthy environment, people hope the acoustic environment can play a promotive role on their physiological indices, psychological feelings, and behavioral demands, which may provide some hints on the parameters and limits values for future policymaking. It is worth mentioning that the aims of this study were to determine the elements, framework, and definition of a healthy acoustic environment. There is no meticulous exploration of the parameters of a healthy acoustic environment, which needs to be studied in a specific context in future. Parameters and their limits values can provide more practical value for acoustic environmental policy.

Limitations

Many researchers focused on sound-related health outcomes, but there was not a consistent understanding on the connotation and framework of a healthy acoustic environment yet. This study proposed the definition and the framework with a grounded theory approach. However, there were some limitations.

Compared to other studies with grounded theory approach (Liu and Kang, 2016; Zhu et al., 2020), the duration of the face-to-face interview was short. The reason might be that the healthy acoustic environment is a new concept and it is a challenge even for professionals to respond to this topic, so ordinary residents have even fewer knowledge or opinion on this topic. To minimize this limitation, some encouraging or substitutive questions were also prepared and provided in the formal interview, in order to make it easier for ordinary residents to respond and to achieve an interview as deep as possible.

In addition, seven categories were saturated because there were not any new subcategories emerging after the 33th respondent. In order to make the dimensions and key points more saturated, additional respondents were interviewed. Around 70 samples, all the codes tended to be stable. It was worth mentioning that the saturation of the dimensions and key points in this study seemed not able to be achieved by interview approach because even experts in the acoustic environment and health could not provide comprehensive codes to the seven elements of a healthy acoustic environment without a systematic review. Therefore, the detailed contents of the subcategories require targeted research in a specific context under the guidance of the holistic framework. This needs further study through combining the qualitative research and systematic literature review of empirical researches.

Lastly, data collection and analysis process were all handled by the researchers, which made them part of the process and it may influence the integration of the codes. It is a limitation of GT and similar qualitative methods (Glaser and Strauss, 1968). In order to overcome the limitation, two researchers conducted the coding process separately, and the coding results were checked with a group of people with the background of acoustics. Moreover, this study followed the standardized procedure and analysis of GT and the researchers displayed all the key points, dimensions, subcategories, categories, and all the coding processes as detailed as possible.

CONCLUSION

This paper presented a pilot study on a healthy acoustic environment. Semistructured interviews were conducted to explore the basic elements, a conceptual framework, and definition of a healthy acoustic environment. Overall, the main conclusions are as follows:

1. The elements of a healthy acoustic environment are “sound sources and acoustic environment,” “people’s demands,” “matching process,” “context,” “criteria and standards of a healthy acoustic environment,” “secondary fitting process,” and “acoustic environment quality.”
2. A conceptual framework was established based on the associations among these categories. The central idea (*phenomenon*) of this research can be labeled as the judgment of the acoustic environment quality. “Sound sources and acoustic environment” and “people’s demands” are the *causal conditions* that give rise to this *phenomenon*. “Context” is the *context* in which the *phenomenon* is embedded. “Matching process” and “secondary fitting process” are the *action/interactional strategies* whereby the *phenomenon* is handled and carried out. “Criteria for a healthy acoustic environment” can be regarded as *intervening condition* of the *phenomenon*. “Acoustic environment quality” (i.e., “unhealthy,” “healthy (low),” and “healthy (high)”) is the *consequence* of the *phenomenon*.
3. Based on the associations revealed in the framework, a healthy acoustic environment is defined as a supportive acoustic environment that can match people’s physiological, psychological, and behavioral demands in context, and that also fits the criteria and standards.

DATA AVAILABILITY STATEMENT

The interview data generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Academic Committee of School of Architecture, Tianjin University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HM and JC: research idea and study design, data collection, data analysis, and manuscript writing. Both authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the National Natural Science Foundation of China (Grant Nos. 51978454 and 51678401).

REFERENCES

Abbasi, A. M., Motamedzade, M., Aliabadi, M., Golmohammadi, R., and Tapak, L. (2018). Study of the physiological and mental health effects caused by exposure to low-frequency noise in a simulated control room. *Build. Acoust.* 25, 233–248. doi: 10.1177/1351010x18779518

Acun, V., and Yilmazer, S. (2018). A grounded theory approach to investigate the perceived soundscape of open-plan offices. *Appl. Acoust.* 131, 28–37. doi: 10.1016/j.apacoust.2017.09.018

Aletta, F., and Kang, J. (2019). Promoting Healthy and Supportive Acoustic Environments: Going beyond the Quietness. *Int. J. Environ. Res. Public Health.* 16:4. doi: 10.3390/ijerph16244988

Aletta, F., Oberman, T., and Kang, J. (2018a). Associations between Positive Health-Related Effects and Soundscapes Perceptual Constructs: A Systematic Review. *Int. J. Environ. Res. Public Health.* 15:2392. doi: 10.3390/ijerph15112392

Aletta, F., Oberman, T., and Kang, J. (2018b). Positive health-related effects of perceiving urban soundscapes: a systematic review. *Lancet* 392:S3. doi: 10.1016/s0140-6736(18)3204432040

Alvarsson, J. J., Wiens, S., and Nilsson, M. E. (2010). Stress recovery during exposure to nature sound and environmental noise. *Int. J. Environ. Res. Public Health* 15, 1036–1046. doi: 10.3390/ijerph7031036

Andringa, T. C., and Lanser, J. J. L. (2013). How pleasant sounds promote and annoying sounds impede health: A cognitive approach. *Int. J. Environ. Res. Public Health.* 10, 1439–1461. doi: 10.3390/ijerph10041439

Annerstedt, M., Jonsson, P., Wallergard, M., Johansson, G., Karlson, B., Grahn, P., et al. (2013). Inducing physiological stress recovery with sounds of nature in a virtual reality forest - results from a pilot study. *Physiol. Behav.* 118, 240–250. doi: 10.1016/j.physbeh.2013.05.023

Axelsson, O., Nilsson, M. E., and Berglund, B. (2010). A principal components model of soundscape perception. *J. Acoust. Soc. Am.* 128, 2836–2846. doi: 10.1121/1.3493436

Basner, M., Brink, M., Bristow, A., de Kluizenaar, Y., Finegold, L., Hong, J., et al. (2015). Icben review of research on the biological effects of noise 2011-2014. *Noise Health* 17, 57–82. doi: 10.4103/1463-1741.153373

Basner, M., and McGuire, S. (2018). WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Effects on Sleep. *Int. J. Environ. Res. Public Health.* 15:45. doi: 10.3390/ijerph150 30519

Baum, A., Revenson, T., and Singer, J. E. (eds.) (2001). *Handbook of Health Psychology*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers. doi: 10. 5860/choice.38-5852

Booi, H., and van den Berg, F. (2012). Quiet areas and the need for quietness in Amsterdam. *Int. J. Environ. Res. Public Health.* 9, 1030–1050. doi: 10.3390/ ijerph9041030

Botteldooren, D., De Coensel, B., and De Muer, T. (2006). The temporal structure of urban soundscapes. *J. Sound Vibr.* 292, 105–123. doi: 10.1016/j.jsv.2005. 07.026

Brown, A. L., and van Kamp, I. (2017). WHO Environmental Noise Guidelines for the European Region: A Systematic Review of Transport Noise Interventions and Their Impacts on Health. *Int. J. Environ. Res. Public Health.* 14:44. doi: 10.3390/ijerph14080873

Charmaz, K. (2014). *Constructing grounded theory: A practical guide through qualitative analysis*, 2nd Edn. Los Angeles, CA: Sage Publications Limited.

Chen, J., and Ma, H. (2019). An impact study of acoustic environment on users in large interior spaces. *Build. Acoust.* 26, 139–153. doi: 10.1177/ 1351010x19848119

Clark, C., and Paunovic, K. (2018). WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Cognition. *Int. J. Environ. Res. Public Health.* 15:285. doi: 10.3390/ ijerph15020285

Darvishi, E., Golmohammadi, R., Faradmal, J., Poorolajal, J., and Aliabadi, M. (2019). Psychophysiological Responses to Medium Levels of Occupational Noise: An Exposure-Response Relationships. *Acoust. Aust.* 47, 217–228. doi: 10.1007/s40857-019-00159-0

Dorota, J., Marie-ève, H., Poonum, W., James, C., Jos, V., Jordis, W., et al. (2018). Development of the WHO Environmental Noise Guidelines for the European Region: An Introduction. *Int. J. Environ. Res. Public Health.* 15:813. doi: 10. 3390/ijerph15040813

Environmental Noise Directive (2002). Directive 2002/49/EC of the European Parliament and Council of 25 june 2002 relating to the assessment and management of environmental noise. *Off. J. Eur. Commun*. 189, 0012–0026.

European Commission (1996). *Green paper of the European Commission: future noise policy*. Belgium: European Commission*com*. .

European Environment Agency [EEA] (2018). *The Noise Observation and Information Service for Europe*. Denmark: European Environment Agency.

Fletcher, M. D., Jones, S. L., White, P. R., Dolder, C. N., Leighton, T. G., and Lineton, B. (2018). Effects of very high-frequency sound and ultrasound on humans. Part I: Adverse symptoms after exposure to audible very-high frequency sound. *J. Acoust. Soc. Am.* 144, 2511–2520. doi: 10.1121/1.5063819

Fredriksson, S., Kim, J. L., Toren, K., Magnusson, L., Kahari, K., Soderberg, M., et al. (2019). Working in preschool increases the risk of hearing-related symptoms: a cohort study among Swedish women. *Int. Arch. Occup. Environ. Health.* 92, 1179–1190. doi: 10.1007/s00420-019-01453-0

Gill, C., Packer, J., and Ballantyne, R. (2018). Applying attention restoration theory to understand and address clergy's need to restore cognitive capacity. *J. Relig. Health.* 57, 1779–1792. doi: 10.1007/s10943-018-0571-9

Glaser, B. G., and Strauss, A. L. (1968). The Discovery of Grounded Theory: Strategies for Qualitative Research. *Soc. Forces.* 46, 377–380. doi: 10.2307/ 2575405

Guski, R., Schreckenberg, D., and Schuemer, R. (2017). WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Annoyance. *Int. J. Environ. Res. Public Health.* 14:39. doi: 10.3390/ijerph14121539

Herranz-Pascual, K., Aspuru, I., Iraurgi, I., Santander, A., Eguiguren, J. L., and Garcia, I. (2019). Going beyond Quietness: Determining the Emotionally Restorative Effect of Acoustic Environments in Urban Open Public Spaces. *Int. J. Environ. Res. Public Health* 16:1284. doi: 10.3390/ijerph16071284

Hume, K., and Ahtamad, M. (2013). Physiological responses to and subjective estimates of soundscape elements. *Appl. Acoust.* 74, 275–281. doi: 10.1016/j. apacoust.2011.10.009

ISO (12913)-1 (2014). *Acoustics-Soundscape-Part 1: Definition and Conceptual Framework*. Geneva: International Organisation for Standardization.

Jiang, H., and Shen, H. M. (2018). Supportive organizational environment, work-life enrichment, trust and turnover intention: A national survey of PRSA membership. *Publ. Relat. Rev.* 44, 681–689. doi: 10.1016/j.pubrev.2018.08.007

Kang, J., Aletta, F., Gjestland, T. T., Brown, L. A., Botteldooren, D., Schulte-Fortkamp, B., et al. (2016). Ten questions on the soundscapes of the built environment. *Build. Environ.* 108, 284–294. doi: 10.1016/j.buildenv.2016. 08.011

Kang, J., Ma, H., Xie, H., Zhang, Y., and Li, Z. Z. (2020). Research progress on the acoustic environments of healthy buildings (in Chinese). *Chin. Sci. Bull.* 65, 288–299. doi: 10.1360/TB-2019-0465

Kempen, E. V., Casas, M., Pershagen, G., and Foraster, M. (2018). WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Cardiovascular and Metabolic Effects: A Summary. *Int. J. Environ. Res. Public Health.* 15:379. doi: 10.3390/ ijerph15020379

King, E. A., and Murphy, E. (2016). Environmental noise - 'Forgotten' or 'Ignored' pollutant? *Appl. Acoust.* 112, 211–215. doi: 10.1016/j.apacoust.2016.05.023

Krzywicka, P., and Byrka, K. (2017). Restorative Qualities of and Preference for Natural and Urban Soundscapes. *Front. Psychol.* 8:1705. doi: 10.3389/fpsyg. 2017.01705

Lazarus, R. S. (1966). *Psychological stress and the coping process*. New York, NY: McGraw-Hill Press.

Lazarus, R. S., and Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer Press.

Lercher, P. (1996). Environmental noise and health: an integrated research perspective. *Environ. Int.* 22, 117–129. doi: 10.1016/0160-4120(95)00109-3

Liu, F. F., and Kang, J. (2016). A grounded theory approach to the subjective understanding of urban soundscape in Sheffield. *Cities.* 50, 28–39. doi: 10.1016/ j.cities.2015.08.002

Medvedev, O., Shepherd, D., and Hautus, M. J. (2015). The restorative potential of soundscapes: A physiological investigation. *Appl. Acoust.* 96, 20–26. doi: 10.1016/j.apacoust.2015.03.004

Meng, Q., Jiang, J., Liu, F., and Xu, X. (2020a). Effects of the Musical Sound Environment on Communicating Emotion. *Int. J. Environ. Res. Public Health* 17:2499. doi: 10.3390/ijerph17072499

Meng, Q., Zhang, J. W., Kang, J., and Wu, Y. (2020b). Effects of sound environment on the sleep of college students in China. *Sci. Total Environ.* 705:135794. doi: 10.1016/j.scitotenv.2019.135794

Ministry of environmental protection of China (2008). *Environmental quality standard for noise, GB3096-2008*. China: Ministry of Environmental Protection.

Nieuwenhuijsen, M. J., Ristovska, G., and Dadvand, P. (2017). WHO Environmental Noise Guidelines or the European Region: A Systematic Review on Environmental Noise and Adverse Birth Outcomes. *Int. J. Environ. Res. Public Health.* 14:1252. doi: 10.3390/ijerph14101252

Park, S. H., Lee, P. J., Jung, T., and Swenson, A. (2020). Effects of the aural and visual experience on psycho-physiological recovery in urban and rural environments. *Appl. Acoust.* 169:107486. doi: 10.1016/j.apacoust.2020.107486

Park, S. H., Lee, P. J., and Yang, K. S. (2016). Perception and Reaction to Floor Impact Noise in Apartment Buildings: A Qualitative Approach. *Acta Acust. Acust.* 102, 902–911. doi: 10.3813/aaa.919004

Rashid, M., and Zimring, C. (2008). A review of the empirical literature on the relationships between indoor environment and stress in health care and office settings. *Environ. Behav.* 40, 151–190. doi: 10.1177/0013916507311550

Schabracq, M. J. (2003). The handbook of work and health psychology. *Int. J. Confl. Manag.* 15, 6–26. doi: 10.1108/eb022905

Seltenrich, N. (2019). Assessing Potential Health Impacts of Wind Turbine Noise: A Longitudinal Look at Multiple End Points. *Environ. Health Perspect.* 127:94003. doi: 10.1289/ehp5374

Shepherd, D., Welch, D., Dirks, K. N., and McBride, D. (2013). Do quiet areas afford greater health-related quality of life than noisy areas? *Int. J. Environ. Res. Public Health.* 10, 1284–1303. doi: 10.3390/ijerph10041284

Shu, S., and Ma, H. (2019). Restorative Effects of Classroom Soundscapes on Children's Cognitive Performance. *Int. J. Environ. Res. Public Health.* 16:15. doi: 10.3390/ijerph16020293

Shu, S., and Ma, H. (2020). Restorative effects of urban park soundscapes on children's psychophysiological stress. *Appl. Acoust.* 164:107293. doi: 10.1016/j.apacoust.2020.107293

Sliwinska-Kowalska, M., and Zaborowski, K. (2017). WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Permanent Hearing Loss and Tinnitus. *Int. J. Environ. Res. Public Health.* 14:19. doi: 10.3390/ijerph14101139

Smyrnova, Y., and Kang, J. (2010). Determination of perceptual auditory attributes for the auralization of urban soundscapes. *Noise Control Eng. J.* 58, 508–523. doi: 10.3397/1.3484177

Stokols, D. (1987). "Conceptual strategies of environmental psychology," in *Handbook of environmental psychology*, eds D. Stokols., I. Altman (New York, NY: John Wiley Sons).

Strauss, A., and Corbin, J. (1990). *Basics on Qualitative Research: Grounded Theory Procedures And Techniques*. New York, NY: Sage Publications, Inc.

Terhardt, E., and Stoll, G. (1981). Skalierung des Wohlklangs (der sensorischen Konsonanz) von 17 Umweltschallen und Untersuchung der beteiligten Hörparameter. *Acta Acust. Acust.* 48, 247–253.

Themann, C. L., and Masterson, E. A. (2019). Occupational noise exposure: A review of its effects, epidemiology, and impact with recommendations for reducing its burden. *J. Acoust. Soc. Am.* 146, 3879–3905. doi: 10.1121/1.5134465

Torresin, S., Albatici, R., Aletta, F., Babich, F., and Kang, J. (2019). Assessment methods and factors determining positive indoor soundscapes in residential buildings: A systematic review. *Sustainability* 11:5290. doi: 10.3390/su11195290

Van Kamp, I. (1990). *Coping with noise and its health consequences*. Ph.D. Dissertation, University of Groningen: Groningen.

Van Kempen, E., Casas, M., Pershagen, G., and Foraster, M. (2018). WHO environmental noise guidelines for the european region: a systematic review on environmental noise and cardiovascular and metabolic effects: a summary. *Int. J. Environ. Res. Public Health* 15:379. doi: 10.3390/ijerph15020379

Wagemakers, A., Vaandrager, L., Koelen, M. A., Saan, H., and Leeuwis, C. (2010). Community health promotion: A framework to facilitate and evaluate supportive social environments for health. *Eval. Program Plan.* 33, 428–435. doi: 10.1016/j.evalprogplan.2009.12.008

Watson, J. B. (1913). Psychology as the Behaviorist Views It. *Psychol. Rev.* 20, 158–177. doi: 10.1037/h0074428

Waye, K. P., Fredriksson, S., Hussain-Alkhateeb, L., Gustafsson, J., and van Kamp, I. (2019). Preschool teachers' perspective on how high noise levels at preschool affect children's behavior. *PLoS One* 14:13. doi: 10.1371/journal.pone.0214464

World Health Organization [WHO] (1986). *Ottawa charter for health Promotion*. Ottawa: Health and Welfare Canada/canadian Public Health Association.

World Health Organization [WHO] (1991). *Sundsvall statement on supportive environments for health*. Sundsvall: World Health Organization.

World Health Organization [WHO] (2006). *Constitution*. Sundsvall: World Health Organization.

World Health Organization [WHO] (2018). *Environmental Noise Guidelines for the European Region*. Copenhagen: WHO Regional Office for Europe.

Zhang, Y., Kang, J., and Kang, J. (2017). Effects of soundscape on the environmental restoration in urban natural environments. *Noise Health* 19, 65–72. doi: 10.4103/nah.NAH_73_16

Zhang, Y. F., Tzortzopoulos, P., and Kagioglou, M. (2019). Healing built-environment effects on health outcomes: environment-occupant-health framework. *Build. Res. Informat.* 47, 747–766. doi: 10.1080/09613218.2017.1411130

Zhu, J. X., Kang, J., Ma, H., and Wang, C. (2020). Grounded theory-based subjective evaluation of traditional Chinese performance buildings. *Appl. Acoust.* 168:107417. doi: 10.1016/j.apacoust.2020.107417

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 05 November 2020
doi: 10.3389/fpsyg.2020.01850

Check for updates

EEG Correlates of Learning From Speech Presented in Environmental Noise

Ehsan Eqlimi[1*], *Annelies Bockstael*[1,2,3], *Bert De Coensel*[1,4], *Marc Schönwiesner*[5,6], *Durk Talsma*[7] *and Dick Botteldooren*[1]

[1] WAVES Research Group, Department of Information Technology, Ghent University, Ghent, Belgium, [2] École d'Orthophonie et d'Audiologie, Université de Montréal, Montreal, QC, Canada, [3] Erasmushogeschool Brussel, Brussels, Belgium, [4] ASAsense, Bruges, Belgium, [5] Faculty of Biosciences, Pharmacy and Psychology, Institute of Biology, University of Leipzig, Leipzig, Germany, [6] International Laboratory for Brain, Music and Sound Research (BRAMS), Université de Montréal, Montreal, QC, Canada, [7] Department of Experimental Psychology, Ghent University, Ghent, Belgium

How the human brain retains relevant vocal information while suppressing irrelevant sounds is one of the ongoing challenges in cognitive neuroscience. Knowledge of the underlying mechanisms of this ability can be used to identify whether a person is distracted during listening to a target speech, especially in a learning context. This paper investigates the neural correlates of learning from the speech presented in a noisy environment using an ecologically valid learning context and electroencephalography (EEG). To this end, the following listening tasks were performed while 64-channel EEG signals were recorded: (1) attentive listening to the lectures in background sound, (2) attentive listening to the background sound presented alone, and (3) inattentive listening to the background sound. For the first task, 13 lectures of 5 min in length embedded in different types of realistic background noise were presented to participants who were asked to focus on the lectures. As background noise, multi-talker babble, continuous highway, and fluctuating traffic sounds were used. After the second task, a written exam was taken to quantify the amount of information that participants have acquired and retained from the lectures. In addition to various power spectrum-based EEG features in different frequency bands, the peak frequency and long-range temporal correlations (LRTC) of alpha-band activity were estimated. To reduce these dimensions, a principal component analysis (PCA) was applied to the different listening conditions resulting in the feature combinations that discriminate most between listening conditions and persons. Linear mixed-effect modeling was used to explain the origin of extracted principal components, showing their dependence on listening condition and type of background sound. Following this unsupervised step, a supervised analysis was performed to explain the link between the exam results and the EEG principal component scores using both linear fixed and mixed-effect modeling. Results suggest that the ability to learn from the speech presented in environmental noise can be predicted by the several components over the specific brain regions better than by knowing the background noise type. These components were linked to deterioration in attention, speech envelope following, decreased focusing during listening, cognitive prediction error, and specific inhibition mechanisms.

Keywords: auditory attention, auditory perception, EEG, inhibition, learning context, long-range temporal correlations, speech in noise, speech processing

OPEN ACCESS

Edited by:
Andrey R. Nikolaev,
Lund University, Sweden

Reviewed by:
Carine Signoret,
Linköping University, Sweden
Giovanni Di Liberto,
École Normale Supérieure, France

****Correspondence:***
Ehsan Eqlimi
ehsan.eqlimi@ugent.be

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

Received: *13 March 2020*
Accepted: *06 July 2020*
Published: *05 November 2020*

Citation:
Eqlimi E, Bockstael A, De Coensel B, Schönwiesner M, Talsma D and Botteldooren D (2020) EEG Correlates of Learning From Speech Presented in Environmental Noise. Front. Psychol. 11:1850. doi: 10.3389/fpsyg.2020.01850

1. INTRODUCTION

The human brain is remarkably capable of focusing on one specific sound while suppressing all others (Alain, 2007). Nevertheless, processing of relevant information largely depends on the specific interaction of the acoustic features of speech and noise signals, their informative content, attention, state, and the prior knowledge (familiarity with the presented topic) of the listener (Szalma and Hancock, 2011). To understand the underlying mechanisms of this diverse phenomenology in human sound interaction, short-term features of distracting events, state of the listener, information flow, and loss of efficiency need to be studied. One key aspect of the study design is ecological validity (Chaytor and Schmitter-Edgecombe, 2003), meaning that realistically complex stimuli and conditions are included possibly in addition to artificially designed stimuli.

In a learning context, the ability to acquire and retain vocal information strongly affects the overall learning performance. This is even more challenging when this occurs in the presence of environmental noise. One of the effects involved in this ability is known as the cocktail party effect (Cherry, 1978), and this refers to the ability of the brain to direct attention to a target sound despite the presence of distracting sounds. Although the underlying mechanisms are indispensable to learn from information presented in an acoustically rich environment (Lehmann and Schönwiesner, 2014), they are far from fully understood.

Attention directs both cognitive and sensory resources to the target sounds (Schneider and Shiffrin, 1977). In general, such resources are limited in capacity based on the bottleneck (Pashler, 1984) and capacity sharing (Kahneman, 1973) theories. Most of the observed effects of noise on learning (Alain, 2007) can be explained by attention, including unlocking undesired attention focus as well as an increased cognitive load when listening to speech in noise (Rudner, 2016). Moreover, listening performance and speech intelligibility in background noise can be impaired by distracting attention away from the narrative and hampering relevant sounds (Ljung et al., 2009; Clark and Sörqvist, 2012). However, attention focusing and appropriate gating of (ir)relevant stimuli are not only the matter of cortical processing but also peripheral neurophysiological stages of auditory analysis are involved. Attention can be modulated by bottom-up factors (referring to external stimulus-driven responses that guide the attention due to inherent properties of salient events relative to the background) as well as top-down task-specific functions (referring to internal modulation of attention that is driven by cognition based on prior knowledge, expectations, and learned schemas) (Katsuki and Constantinidis, 2014; Kaya and Elhilali, 2017).

Auditory attention-related research (especially bottom-up attention) mostly adopts an event-related potential (ERP) design (Alain, 2007). However, a classical ERP design with repeated stimuli conflicts with the idea of ecologically valid stimuli and studying top-down attention. In the current paper, the single-trial EEG experiment was used to study how auditory-related neural responses vary depending on acoustical stimulus and listening condition. The power spectrum of EEG signal exhibits peaks in different frequency ranges reflecting different underlying mechanisms (Buzsáki et al., 2013; He, 2014). Therefore, one of the most common methods to process the single-trial EEG signals is spectral analysis, which relies on partitioning the signal into the different frequency sub-bands (Clayton et al., 2015).

Previous studies using spectral analysis have shown different frequency bands contribute to the various underlying mechanisms during listening to speech in noise, such as top-down attention (Gazzaley and Nobre, 2012), cortical inhibition (Uusberg et al., 2013), language processing (Pulvermüller et al., 1997), neural entrainment to speech (Riecke et al., 2018), and excitation-inhibition balance (Poil et al., 2012). The roles of the different frequency bands in these mechanisms are discussed separately below.

Low-frequency EEG signals ($1-8$ Hz) can be modulated by attention (Kerlin et al., 2010; Braboszcz and Delorme, 2011). Two important mechanisms may be associated with the low-frequency EEG. The first one is the mismatch between current and desired levels of attention (Clayton et al., 2015) and the transition of the fatigue state (Borghini et al., 2014), which is observed as a continuous increase of low-frequency power with time on task [unlike the alpha-band activity ($8-13$ Hz) (Mierau et al., 2017)]. Frontomedial theta-band ($4-8$ Hz) activity has been linked to both enhanced attention over short time-scale cognitive tasks and reduced attention (increased attentional fatigue) following long time-scale cognitive tasks (Wascher et al., 2014; Clayton et al., 2015). Moreover, it has been shown that the delta-band ($1-4$ Hz) absolute power is higher in the mind wandering compared to the focused state over the fronto-central region (Braboszcz and Delorme, 2011).

The second mechanism is the information and attention selection (Schroeder and Lakatos, 2009; Herrmann et al., 2016). This means that the attention can use a mechanism of selection leading to oscillatory entrainment to a task-relevant stimulus (Schroeder and Lakatos, 2009). However, neural entrainment is a broader concept and refers to the temporal alignment of neural signals with regularities in an exogenously occurring stimulus, such as speech (Obleser and Kayser, 2019) and even aperiodic (speech) signals (Obleser et al., 2012; Goswami and Leong, 2013).

Speech following (and speech envelope following/tracking) as one the manifestation of the neural entrainment refers to the relation between the neural and sound signals (Obleser and Kayser, 2019). Although it has been measured in various frequency bands (Obleser and Kayser, 2019), its impact on low-frequency EEG (delta and theta bands) has been shown in several electrophysiological experiments (Luo and Poeppel, 2007; Doelling et al., 2014; O'Sullivan et al., 2014; Kayser et al., 2015). The basic hypotheses of these studies are the following: (1) entrainment occurs also at other frequencies, but this effect is obscured by stronger signals; (2) the low-frequency speech envelope entrainment of brain activity could be robust against different background noises (Ding et al., 2014); and (3) the speech envelope is constituted by slow temporal modulations, which contribute to speech recognition despite different background sounds (Houtgast and Steeneken, 1985; Rosen, 1992; Kerlin et al., 2010; Ding and Simon, 2013; Ríos-López et al., 2017).

It has also been shown that attended and unattended stimuli could be decoded by low-frequency single-trial EEG in a cocktail party scenario based on a stimulus-reconstruction algorithm (O'Sullivan et al., 2014). This stimulus-reconstruction method indicated the slow amplitude envelope of attended speech ($\leq$ 8 Hz) is tracked more strongly by the low-frequency EEG (2−8 Hz) compared to the unattended speech. Furthermore, it has been shown that in the multi-talker speech perception, the attended speaker is represented over the non-primary auditory cortex (AC) while the individual speakers are represented over the primary AC (O'Sullivan et al., 2019).

Alpha-band activity ($\sim 8 - 13$ Hz) is also often modulated by auditory attention, especially by the inhibition function (Strauß et al., 2014). The term "alpha-as-inhibition" is used to highlight that alpha-band activity, beyond resting state, could reflect inhibition of the distracting sound (Clark, 1996; Uusberg et al., 2013). Increased alpha-band activity over the task-irrelevant brain regions reflects less involvement of those regions. Hence, comparison of alpha power between task-relevant and task-irrelevant cortical regions can be an indicator for inhibition (Pfurtscheller and Da Silva, 1999; Chang et al., 2010). In fact, alpha event-related synchronization (ERS) reflects inhibition and alpha event-related desynchronization (ERD) releases from inhibition (Foxe et al., 1998; Snyder and Foxe, 2010; Klimesch, 2012).

Not only absolute alpha power over a fixed frequency band but also alpha peak frequency (APF) and its corresponding power can be associated with attention, inhibition, memory, and cognitive demand (Klimesch, 1997; Clark et al., 2004; Haegens et al., 2014; Gulbinaite et al., 2017). APF (Doppelmayr et al., 1998) and individual alpha frequency (IAF) (Klimesch, 1999) indicate the actual frequency limits of alpha activity, which exhibit variability within and between subjects (Haegens et al., 2014). APF is also linked to the number of spiking neurons or the input level (Mierau et al., 2017). If the input level increases with respect to the baseline level, APF increases until the oscillation becomes unstable and then it is replaced by a lower frequency (Mierau et al., 2017). Although APF increases with a higher allocation of attentional resources, it decreases with lower attentional demand and cognitive load due to unstable state and overloaded attention capacity (Hutt et al., 2016; Mierau et al., 2017). Higher APF can be accompanied by lower alpha power resulting in task-relevant regions that exhibit increased APF during task performance (Hutt et al., 2016). Studies focusing on power-related frequency shifts have suggested a rather complex relationship between alpha frequency and power (Kawabata, 1972). Other studies have shown that APF decreases with increasing attentional demand and task difficulty (Angelakis et al., 2004; Haegens et al., 2014), which could be explained by unstable state and overloaded attention capacity. Enhanced APF might reflect a state of cognitive preparedness and the attentional switch between wandering and focused states of mind (Braboszcz and Delorme, 2011).

In addition to the peaks at the frequency ranges, a predominant "$\frac{1}{f}$" component in the EEG power spectra leads a power-law function, i.e., $p \propto \frac{1}{f}^{a}$, where p is power, f is frequency, and a is the scaling exponent (He, 2014). Therefore, the EEG time series exhibit scale-free dynamics and do not have a characteristic scale (He et al., 2010; He, 2014). Furthermore, the ongoing EEG signals hold a memory of their own dynamics on time-scales, which could be linked to the scale-free dynamics and the self-similarity concept in fractal geometry (Palva et al., 2013). Long-range temporal correlations (LRTC) are the most common measures with which to quantify how slowly the autocorrelations of the signal decay in power-law function (Linkenkaer-Hansen et al., 2001; Nikulin and Brismar, 2005; Palva et al., 2013). Alpha-band LRTC could reflect an optimal balance between excitation and inhibition states (Poil et al., 2012). Decreased alpha-band LRTC compared to the resting state correlates with better attentional performance (Colosio et al., 2017). Higher alpha-band LRTC during resting-state could predict high performance in decision making (Colosio et al., 2017), working memory (Mahjoory et al., 2019) and attention tasks (Irrmischer et al., 2018).

Increased beta-band ($\sim 13 - 30$ Hz) power over the fronto-lateral region has been observed in the mind wandering compared to focused state (Braboszcz and Delorme, 2011). Furthermore, the beta-band activity can be related to the maintenance of current sensorimotor or cognitive task (Engel and Fries, 2010; Weiss and Mueller, 2012; Zhao et al., 2012). A quasi-harmonic relationship has been suggested between the beta and alpha peaks or central frequencies only during rest (Van Albada and Robinson, 2013; Haegens et al., 2014). The lack of a strict relationship between the beta and alpha peak frequencies during task-based conditions may reflect independent networks being activated (Jones et al., 2009; Haegens et al., 2014).

Localized gamma-band activity ($\sim 30 - 45$ Hz) has been found in task-relevant cortical regions (MacDonald and Barth, 1995; Cervenka et al., 2011; Siegel et al., 2011). Gamma-band activity plays a central role in attention, perception and language processing (Pulvermüller et al., 1997). Furthermore, gamma-band activity in sensory cortices has often been linked with enhanced attention to these particular sensory inputs (Ahveninen et al., 2013). It has also been shown that gamma-band power in auditory areas increases during extended auditory attention tasks (Kaiser and Lutzenberger, 2005; Ahveninen et al., 2013). According to popular theory, gamma waves may be implicated in creating the unity of conscious perception and semantic processing (Buzsaki, 2006).

In this study, we aimed to investigate the different mechanisms involved in learning from the speech presented in noise using single-trial EEG and mimicking an ecologically valid context. To this end, 23 participants were exposed to the following listening tasks while 64-channel EEG signals were recorded: (1) attending to a lecture in the background noise (LA), (2) attending to the background noise alone (BA), and (3) not attending to the sound while still being exposed to the background noise (BUA). For the background noise, realistic environmental sound fragments from continuous highway noise (HW), fluctuating traffic (FT), and multi-talker babble (MT) were used. A written exam on the lecture was taken after 13 sets of 5-min lectures and the BA task for assessing the amount of information that participants have actually acquired and retained from the lectures.

We hypothesized several neural mechanisms, such as cortical inhibition, auditory attention, neural entertainment, and predictive coding, can be affected by the listening conditions we have designed. Therefore, five qualitative hypotheses were considered: (1) alpha-as-inhibition, (2) excitation-inhibition balance reflected in the alpha band, (3) low-frequency envelope following, (4) maintenance of current cognitive task, and (5) semantic processing and cognitive prediction violation or error.

The alpha-as-inhibition hypothesis (Uusberg et al., 2013) implies that alpha-band activity mediates inhibition of task-irrelevant cortical areas. The excitation-inhibition balance hypothesis (Poil et al., 2012) relates the task performance and optimal information processing to the long-range temporal correlations of alpha-band activity. The low-frequency envelope following hypothesis (Luo and Poeppel, 2007; Kerlin et al., 2010; Obleser and Kayser, 2019) implies the neural entrainment and tracking of speech (and background sound) envelope can be reflected in the low-frequency bands, i.e., delta and theta frequency bands. However, here, the relation between the EEG and sound signals has not been analyzed (which is the main tool to measure the envelope following) due to our unsupervised approach. In fact, we have assumed that the strong representation of low-frequency EEG signals (i.e., changes in spectral characteristics) may be related to the envelope following. Although, the neural entrainment and envelope following occurs also at higher frequencies but this effect is obscured by stronger signals (note that no source reconstruction was used in this paper). The hypothesis of maintenance of current cognitive task (Spitzer and Haegens, 2017) implies that the preservation of the current brain state and the long-range communication can be associated with the beta-band activity. Finally, the last hypothesis suggests that semantic or higher-level processes (specifically semantic violations) due to speech processing induce power changes in the gamma-band activity (Braeutigam et al., 2001; Buzsaki, 2006; Hald et al., 2006; Penolazzi et al., 2009). Moreover, the generative models for the perception, such as the predictive coding (Sedley et al., 2016) assume the precision of prediction, changes to predictions, and violations (errors) in predictions are encoded with the alpha, beta, and gamma frequency bands, respectively. These assumptions can be in accordance with our hypotheses.

Since a few EEG indicators, such as alpha peak frequency and power, alpha long-range temporal correlations, and delta absolute power were evaluated in a recent work by our group (Eqlimi et al., 2019), a wider range of EEG features (see below) was estimated for investigating our hypotheses. More precisely, the following features were estimated: spectral features and peak frequency of the alpha-band activity (hypothesis 1), the alpha-band LRTC (hypothesis 2), the spectral features of the delta and theta (hypothesis 3), the beta (hypothesis 4), and the gamma (hypothesis 5) frequency bands. To group these features, the hypothesis that different listening tasks (LA, BA, and BAU) create a variance in the EEG features that will also be responsible for at least part of the observed differences in learning from speech in noise, was used. Variance in the EEG features between participants is likewise expected to be informative for the observed differences in learning from speech. Different techniques are available for data-driven aggregation of the broad collection of EEG features. Principal Component Analysis (PCA) of the z-score for each feature is the lowest order approach. It could be extended to higher-order statistical methods and machine learning (e.g., using deep learning auto-encoders). Because of the amount of data available and the advantage of explainable results, we decided to use PCA based on z-score normalized data. To explain the meaning of the EEG-PC scores (the representation of EEG features in the PC domain), they were compared between the listening tasks (LA, BA, and BUA) and background noises (MT, HW, and FT) using linear mixed-effect modeling (Bates et al., 2015). Assuming that the EEG PCs grasp the main variance between listening conditions observed through the different listening tasks, a supervised analysis was performed to relate them to the information acquiring and retaining z-scores (the exam results) in the lecture attended (LA) task using linear fixed and mixed-effect modeling. Also, for this predictive model, higher-order statistical approaches and machine learning techniques could have been used, yet we again opted for reducing the degrees of freedom in the model in view of the available data.

2. MATERIALS AND METHODS

2.1. Participants

Twenty-three young healthy adults (mean age: 27 years, SD: 3.18, 13 females, 20 right-handed), all English speakers, participated in the experiment. Participants had normal hearing measured by pure-tone audiometry. All participants signed the informed consent and received modest financial compensation for their participation. Based on self-reports, none of them had a history of psychiatric or neurological disorders. A full battery of audiological tests was conducted including tonal audiometry, tympanometry, stapedial reflex measurement, speech in noise, and otoacoustic emissions (OAE) with contralateral suppression. No participants were excluded on the basis of this extensive testing of the auditory periphery. Our test population was young adults and therefore their hearing capabilities were fully developed (Klatte et al., 2013).

2.2. Tasks and Stimuli

The main stimulus was about 1 h of English lectures mixed with realistic background noise and presented through a loudspeaker while 64-channel EEG signals were recorded. Participants were instructed to pay attention to the lectures and were informed that there would be a written exam afterward. This task is hereafter referred to lecture attended (LA). The lectures were read by a male speaker and recorded in an anechoic room. To level out participants' particular interests, 13 different 5-min topics were presented over one long lecture. The lectures were related to topics for which prior knowledge is expected to be minimal in order to facilitate the focusing of attention during the presentation.

For the background noise, three 5-min realistic environmental sound fragments from continuous highway noise (HW), fluctuating traffic (FT), and multi-talker babble (MT) sounds were used. Within these fragments, a few discrete instances of very salient sounds were added. In addition, four lecture fragments were presented in silence with a low level pink noise (PK) (a.k.a $\frac{1}{f}$ noise) at a level of 35 dB(A). The signal-to-noise ratio (SNR) of lectures and background noise was set to 5 dB, with lectures at a level of 68 dB(A) and overall background noise level at 63 dB(A). This assured that the background noise did not mask the lecture energetically. The sound levels reported here refer to the A-weighted equivalent continuous sound levels in decibels (LAeq) which were measured over about 360 s.

For the multi-talker babble sound, recordings were made at a cocktail party where about twenty people were having conversations. The recorded speech was not intelligible. A few 3-s phone ringing sounds were added to the multi-talker sound at certain times. For the highway sound, the noise of dense traffic was recorded, for which no individual car passages could be recognized. A few 5-s emergency vehicle sounds were added to the highway sound at certain times. For the fluctuating traffic sound, recordings were made at the corner of a one-way car lane with a bicycle lane next to it, close to a park. Car passages were added to the quietest periods of the fluctuating traffic noise. In addition, at certain times, a few 1-s sounds of honking car were added. The level of the salient sounds (emergency siren, phone ringing, and car's horn) was not high enough to mask the lectures energetically. The order of presentation was completely random in both lecture and background noise while assuring the two lectures in silence were not presented in succession.

The written exam was presented after the BA condition (see below), which ensured that there was a time span of 45 min between the lectures and the exam. The purpose of presenting the exam is to quantify the amount of information that participants have actually acquired and retained from the lectures. The type of questions and evaluation of the exam is explained in section 2.9. A sufficiently long time interval between the learning phase and the memory retrieval during the exam was chosen for two reasons: (1) to avoid that the last lecture would be more prominently in short term memory; (2) to avoid sequential recall as much as possible. Testing the memory and learning in a timescale of minutes and hours was discussed in Tetzlaff et al. (2012) and Kelley and Whatson (2013). For example, memory retention was tested after 30 min (Menzel et al., 2001). The choice of 45 min was a compromise between the duration of the experiment and assuring the above.

To increase the range of monitored listening conditions and to allow to implicitly calibrate for inter-person differences, the participants were exposed to two additional tasks. Firstly, as a reference for top-down attention-driven listening, 12 different 3-min fragments of background noise were presented with equivalent levels of 63 dB(A) and participants were asked to pay attention to the background noise by focusing on the number of salient events, such as phone-ringing, emergency vehicle, and honking car sounds. However, this was only to make them focus on the background sound and no questions were asked about this afterward. This task hereafter is referred to as background attended (BA). Finally, 12 different 3-min background noise fragments were presented and the participants were instructed not to pay attention to any sound, which hereafter is referred to background unattended (BUA). The BUA task is definitely different from the resting state because not paying attention to the low-level characteristics of the sounds is inevitable. Unlike the BA task, the participants during BUA were instructed not to focus on the information related to the salient events. BUA task was presented after the exam which made the participants very aware that no further attention was needed at this point, and they could relax.

By listening task (or simply task) along with this paper, we mean the tasks that the participant had to perform during the experiment, i.e., LA, BA, and BUA. By listening condition (or simply condition), we mean the conditions that the subjects were flooded with the listening tasks and the stimuli. In total, all subjects were exposed to ten different listening conditions depending on the task and noise: LA-PK, LA-MT, LA-HW, LA-FT, BA-MT, BA-HW, BA-FT, BUA -MT, BUA-HW, BUA-FT. For instance, LA-MT refers to the condition that the task is LA and the background noise is MT. The experimental protocol is schematized in **Figure 1**.

Figure 2 depicts the sound level fluctuations as a function of time (line plots) and standard spectrograms (heatmaps) for one of the sound fragments presented during the LA and BUA listening tasks. From **Figure 2**, the FT background noise stands out in terms of sound level fluctuation. For the HW background noise, the sound level is quite stable. Finally, the MT background noise exhibits somewhat more fluctuations in the sound level than the HW noise, but the differences between the loudest and the quietest sounds levels are higher in the FT.

Note that the background sounds used in the LA and BUA tasks were the same (except the time duration). Furthermore, the type and order of background sounds presented in BA and BUA were identical for all participants. The only difference between the stimuli presented during BA and BUA is that additional salient sounds were added in the last three fragments of BA due to the increased chance of focusing on the background noise sound in the BA task.

2.3. EEG Recording

EEG signals during the different listening conditions were acquired continuously using a BioSemi System (Amsterdam, NL) from 64 active electrodes placed according to the standard $10-20$ layout (Oostenveld and Praamstra, 2001) at a sampling frequency of $2,048$ Hz. Subjects were asked to keep their eyes open and focus on a dot located in the center of the monitor to minimize eye movement. Signals from seven external electrodes were also recorded which were applied to the nose, neck, two left & right mastoids (M1 and M2), left (HEOGL) & right (HEOGR) outer canthi, and below the left eye (VEOGD). In addition, two external channels were used for capturing the sound signals (SoundL and SoundR) together with EEG signals.

2.4. EEG Pre-processing

The EEG data were offline re-referenced to the nose electrode (channel 65th) and re-sampled to 512 Hz using an anti-aliasing finite impulse response (FIR) low-pass filter. The EEG data were then filtered using an FIR bandpass filter (Hamming windowed sinc) of order 3,380 from 0.5 to 134 Hz to remove extremely slow drifts and sharp oscillations.

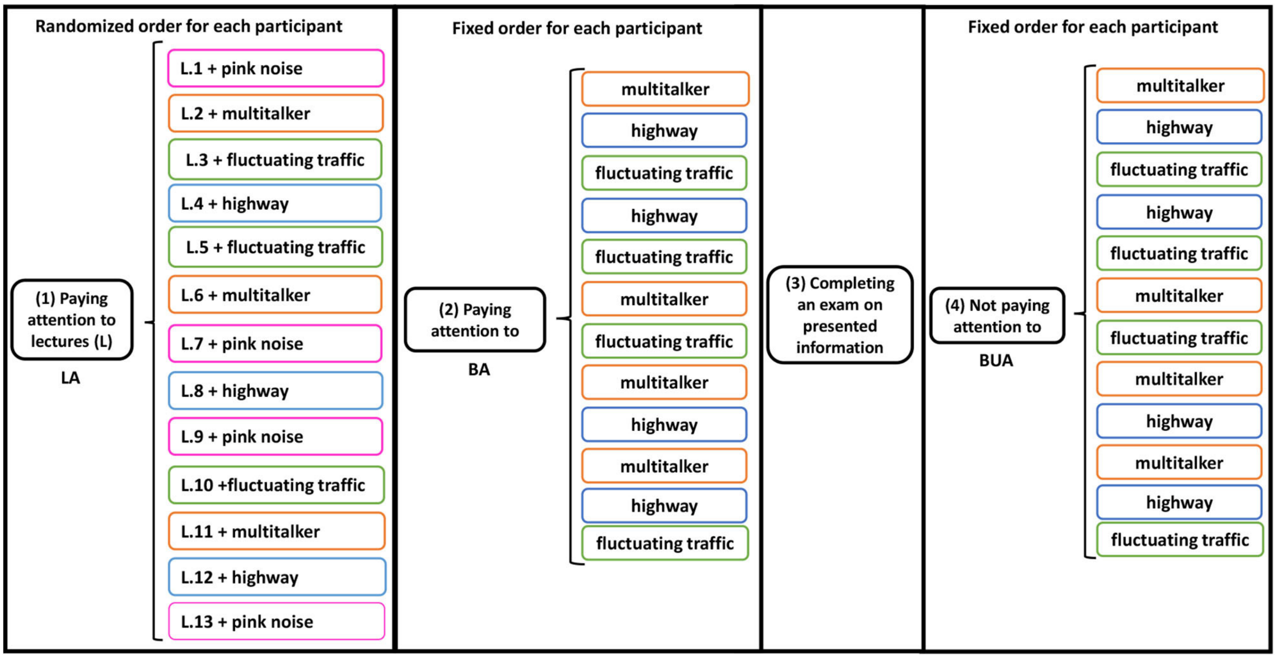


FIGURE 1 | Schematic of experimental protocol and auditory stimuli presentation. Three sequential listening tasks were performed: (1) Lecture attended (LA), (2) Background attended (BA), and (3) Background unattended (BUA). In first task, in addition to multi-talker, highway and fluctuating traffic sounds as the background noises, the lectures were also presented in pink noise and without any background noise. After second task, a written exam was asked to complete about vocal information in the first task. Equivalent levels of background noise and lecture were ∼63 and 68 dB(A), respectively. The lectures were are shown by L.i, $i = 1, \ldots, 13$, and the noises are distinguished by different colors in the figure.

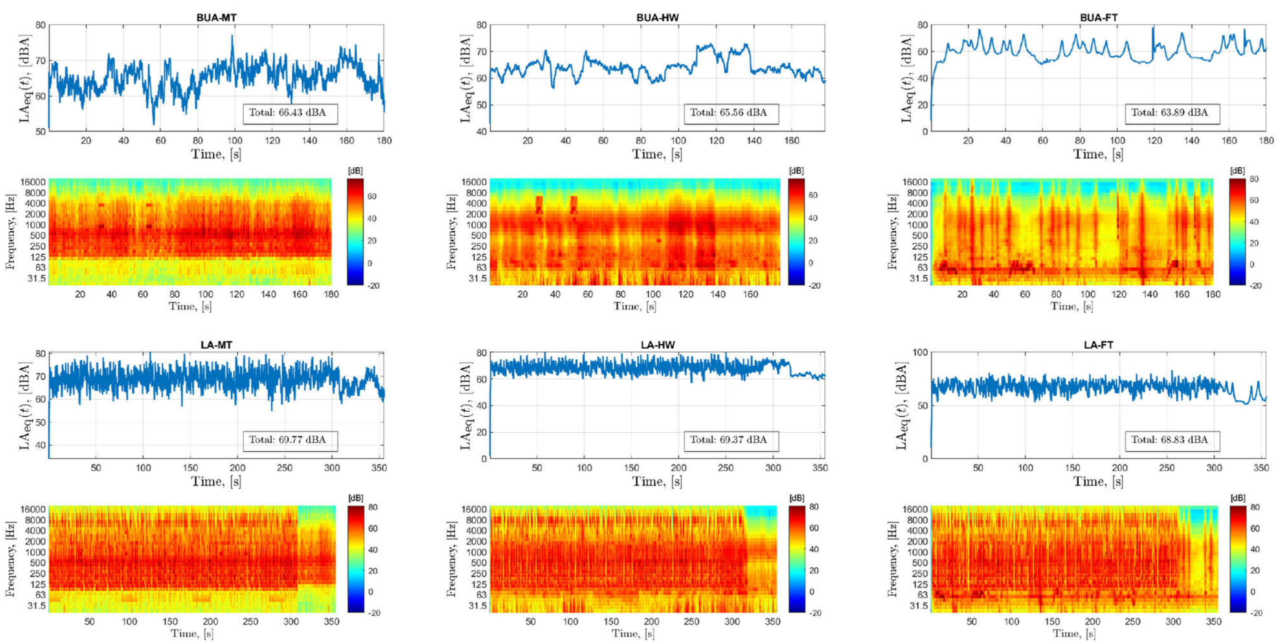


FIGURE 2 | Acoustic characteristics of stimuli presented in different listening conditions. Each panel corresponds to one of the sound fragments presented during the conditions labeled at the top of each panel. The line plots show the sound level fluctuation with time (A-weighted, equivalent continuous sound level in decibels, LAeq). The averaged LAeqs over the whole duration of the fragment have been annotated in the line plots. The heatmaps show the standard spectrogram of the sound fragments (with time resolution of 0.01 s and one-third octave frequency bands). LA, lecture attended; BUA, background unattended tasks; MT, multi-talker; HW, highway; FT, fluctuating traffic background noises.

EEG signals were cleaned up in two steps. At first, non-repeating big artifacts were removed based on visual inspection. In a second step, infomax independent components analysis (ICA) (Bell and Sejnowski, 1995) with EEGLAB version 13.1.1b (Delorme and Makeig, 2004) using default settings was applied to identify and remove eye blink and movement artifacts. To identify the ICA components related to eye artifacts, some rules of thumbs were applied: (1) no more than three ICA components were removed; (2) both temporal and spatial plots should confirm the diagnosis of eye artifact, meaning frontally located components and a typical blink or nystagmus pattern; (3) in case of doubt, the temporal pattern of the supposed ICA component was compared with the temporal pattern of the Electrooculography (EOG) channels to make sure that the incidence of potential eye artifacts coincide; (4) only eye artifacts were removed.

Since playing audio files typically has a latency of a few milliseconds, the sound was recorded together with the EEG on a free channel which could be used to synchronize with the presented audio signal. For this purpose, at first, the presented audio files were re-sampled to 512 Hz (using an anti-aliasing FIR low-pass filter) and then the cross-correlation between re-sampled audio signals and recorded sound signals together with the EEG was calculated. The lag corresponding to maximum cross-correlation is the delay in audio files with respect to EEG measurement. To compensate for this delay, all 64-channel EEG signals were shifted with estimated delays. For the analysis in this manuscript, this synchronization is less important.

Finally, the power spectrum plots of all EEG channels were visually inspected, and the fragments whose all channels were extremely noisy were excluded. In addition, using the power spectrum and a combination of visual inspection and automatic method (median-based criteria), the channels that were extremely noisy were excluded.

2.5. EEG Signal Processing

First, the continuous EEG signals were split into separate fragments corresponding to the 3 or 5 min exposures, based on sound signal recorded as extra EEG channel. Each EEG fragment was then analyzed per channel. Three types of EEG feature were estimated: (1) low-frequency-based features, such as absolute and relative powers, bandwidth, central frequency and spectral edge frequency for delta and theta activities, (2) alpha-band based features, such as alpha peak frequency and power, individual alpha frequency, absolute and relative powers, and alpha-band scaling exponent value as a dynamic measure to quantify LRTC, and (3) high-frequency-based features, such as bandwidth, central frequency, and spectral edge frequency for the beta and gamma signals. Moreover, wide-band absolute power, theta/alpha ratio power, and absolute power for lower and upper alpha were estimated. The subsequent sections describe how a broad range of EEG features was estimated.

2.5.1. Power Spectra-Based Features

To estimate the power spectrum density, Welch algorithm was applied. We used 1 s hamming window with 0.5 s overlap, 2^{14} frequency bins, and frequency sampling of 512 Hz. The power spectrum density, $\mathbf{p}$, is estimated for the frequency range $\mathbf{f} \leftarrow [0 - 256\ Hz]$ with frequency resolution of $\frac{2^6}{2^{14}}$ Hz. In addition to six fixed frequency bands including δ (1 − 4 Hz), θ (4 − 8 Hz), α (8 − 13 and 7 − 13 Hz), β (13 − 30 Hz), and γ (30 − 45 Hz), the lower (8 − 10 Hz), upper (10 − 13 Hz) α band, and the wide-band (1 − 45 Hz) were separately analyzed.

Absolute and relative powers (AP and RP) were calculated from the 64 scalp locations in the mentioned frequency bands. Relative power was computed as the ratio of power in a given band to sum of power from 1 to 45 Hz. Moreover, the $\frac{\theta}{\alpha}$ power ratio (RPTA) was computed. For the frequency band 1 to 45 Hz, only the absolute power was computed. In addition to these power-based features, the following frequency-based features (Szeto, 1990; Drummond et al., 1991; Estrada et al., 2004; Vural and Yildiz, 2010) were computed for the different frequency bands using the definitions in Vural and Yildiz (2010): (1) central frequency, (2) bandwidth, and (3) spectral edge frequency 95%. The central frequency (CF) is defined as the center of gravity for frequency between the lower and upper cutoff frequencies of the power spectrum. The bandwidth (B) quantifies the width of the power spectrum over a specific central frequency. The spectral edge (SE) frequency 95% is defined the frequency below which 95% of the total power (in a specific frequency band) are located (Szeto, 1990).

2.5.2. Alpha Peak Frequency Based on Root-MUSIC

To estimate the alpha peak frequency and power, we used the root-multiple signal classification (root-MUSIC) algorithm (Barabell, 1983). The root-MUSIC as a subspace-based method estimates the frequency content of a signal using an eigenspace method. The root-MUSIC algorithm has been described in recent work from our group (Eqlimi et al., 2019). In this paper, the preprocessed EEG signals were band-pass filtered at 7 − 13 and 8 − 13 Hz (using Butterworth band-pass filter of order 2) for two reasons: (1) there is no consensus on the alpha range (like other frequency bands) and both lower cutoff frequencies (7 and 8 Hz) have been used in literature (Freeman and Quiroga, 2012; Clayton et al., 2015); (2) it has been shown that there is a 2.8 Hz between-subject variability (mean = 10.3 Hz) for the alpha peak frequency (Haegens et al., 2014). The root-MUSIC algorithm was performed on each filtered EEG channel with $P = 2$ as the dimension of the signal subspace. The maximum powers in μV^2 and corresponding frequency in Hz were found. MP2713 and MP2813 terms (which are used in the following sections) stand for MUSIC-based alpha peak power which are estimated in alpha frequency ranges of 7 − 13 and 8 − 13 Hz, respectively with $P = 2$ components. The corresponding alpha peak frequencies are denoted by MF2713 and MF2813.

2.5.3. Individual Alpha Frequency Based on Fitting Process

Individual alpha frequency (IAF) could also be estimated based on the Gaussian fit approach (Nikulin and Brismar, 2006; Van Albada and Robinson, 2013; Haegens et al., 2014). We employed the algorithm which has been suggested in Neurophysiological Biomarker Toolbox (NBT) version

0.5.5 (Hardstone et al., 2012) to quantify IAF. Firstly, PSD (**p**) and its corresponding frequencies (**f**) of each EEG signal with a 0.1 Hz resolution were estimated. The peak amplitudes and locations of **p** in the range of 8–13 Hz were found (using Matlab function "findpeaks"). A polynomial ($y_0 = p_1x + p_2$) function was fitted to ln(**p**) for considering a $\frac{1}{f}$ baseline. Then, $\mathbf{z} \leftarrow e^{p_2} + \mathbf{f}^{p_1}$ and $\mathbf{s} \leftarrow \mathbf{p} - \mathbf{z}$ were calculated to remove the $\frac{1}{f}$ component of the spectrum (Nikulin and Brismar, 2006).

A Gaussian function, $y_1 = a_1 e^{-(\frac{x-b_1}{c_1})^2}$ was fitted to the detrended power spectrum, **s**, to consider one peak. 95% prediction bounds, i.e., confidence interval, $[cl_l, cl_u]$ for a_1 and b_1 were calculated. If $a_1 + y_0(b_1) > cl_u$, then $f_\alpha \leftarrow b_1$ and $p_\alpha \leftarrow a_1$. To determine IAF, center of gravity within the alpha band could be estimated. At first, the individual frequency interval, namely $[f_1, f_2] \leftarrow [\text{TF}, |5 - (f_\alpha - 1)| + f_\alpha]$ is calculated. TF stands for transition frequency and defined as the EEG frequency lower than the alpha peak frequency showing the minimum power (Klimesch, 1999). Then, the center of gravity was calculated using $\text{IAF} \leftarrow \frac{\sum_{k=f_1}^{f_2} \mathbf{f}(k)\mathbf{p}(k)}{\sum_{k=f_1}^{f_2} \mathbf{p}(k)}$. Finally, f_2 was updated by $f_2 = |5 - (\text{IAF} - 1)| + \text{IAF}$ and IAF was re-calculated based on same definition. Compared to root-MUSIC based alpha peak frequency (MF2813), the IAF is expected to be less sensitive to bandwidth around the observed frequency, yet both parameters are highly correlated.

2.5.4. Long-Range Temporal Correlations of Alpha Activity

Processes that do not have a characteristic scale (i.e., scale-free processes) cannot be described completely in terms of spectral concepts (e.g., peak frequency). There is convincing evidence that EEG time series exhibit scale-free dynamics (He, 2014). One of the successful methods to analyze these scale-free signals is long-range temporal correlations (LRTC). LRTC has been developed to quantify how much future dynamics of a signal are influenced by past temporal events (Linkenkaer-Hansen et al., 2007).

In fractal geometry, LRTC could be interpreted by a self-similarity behavior, which suggests the signal dynamics are similar in different time scales. One of the most common techniques to quantify LRTC is detrended fluctuation analysis (DFA) (Peng et al., 1994). The presence of a trend in the signal can cause an overestimation of LRTC, hence DFA tries to eliminate the trend. Indeed, DFA is employed to quantify how slowly the autocorrelations of signals decay in power law, which is called the scaling exponent value, a. The power or scaling law states that a relative change in one quantity results in a proportional relative change in another, namely one quantity varies as a power of another. Distributions of the form $p(x) = Cx^{-a}$ are said to follow a power law. The constant a is called the exponent of the power or scaling exponent value (SEV) (Newman, 2005). If $0.5 < a < 1$, the signal likely exhibits strong LRTC (Hardstone et al., 2012).

We employed the DFA algorithm to quantify LRTC for each EEG channel signal in the alpha band using the NBT version 0.5.5 as suggested in Hardstone et al. (2012). First, the EEG signals were band-pass filtered from 8 to 13 Hz (alpha range used in Hardstone et al., 2012) using the Hamming windowed FIR filter of order 0.25 s (2 cycles of the lowest frequency, 8 Hz). Second, the amplitude envelope of the band-pass filtered signal was estimated based on the Hilbert transform. Third, the cumulative sum of the amplitude envelope was calculated as follows:

$$\mathbf{c}(t) = \sum_{k=1}^{t} \mathbf{e}(k) - \bar{\mathbf{e}}, \tag{1}$$

where $\mathbf{e}(k)$ is the amplitude envelope at time instant k, $\bar{\mathbf{e}}$ is mean of the amplitude envelope, and $\mathbf{c}(t)$ is the cumulative sum of amplitude envelope at time instant t (a.k.a signal profile). We defined a set of window size, $\mathbf{s} = \{s_1, ...s_N\}$, which are equally spaced on a logarithmic scale in a predefined calculation range. The cumulative sum of amplitude envelope ($\mathbf{c}(t)$) was then split into a set of n separated time windows of length $\forall\ l \in \mathbf{s}$, which have 50% overlap. For each time window, the linear trend was removed using a least squares method and obtained the detrended version. After calculating the standard deviation of the detrended time windows, the fluctuation function as the mean standard deviation of all windows was computed as follows:

$$\bar{\mathbf{f}}(l) = \frac{1}{n} \sum_{i=1}^{n} \sigma_{\hat{\mathbf{w}}_i^l}, \quad l \in \mathbf{s}, \tag{2}$$

where $\sigma_{\hat{\mathbf{w}}_i^l}$ is the standard deviation of ith time window of length $l \in \mathbf{s}$, n is the number of time windows. Finally, we plotted the fluctuation function, $\bar{\mathbf{f}}(l)$, along l on logarithmic axes. The slope of the trend line was computed in a predefined fitting interval using the linear regression as a measure for LRTC which is called scaling exponent value (SEV). Two different calculation ranges of 2.5–180 s and 0.1–180 s were evaluated (SEV1 and SEV2, respectively). A fitting interval of 5–18 s was considered such that the filter effect is negligible (Hardstone et al., 2012). The signal length in the LA task was about 360 s, whereas the signal length in the BA and BUA tasks was about 180 s. To minimize the effect of signal length, 180 s was selected as the upper bound of calculation range for the three listening tasks.

2.6. Unsupervised Analysis Using Principal Component Analysis

Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ contains n observations of p EEG features, where could be obtained by concatenating the EEG features per participant, channel, stimulus, and condition in rows. In order to emphasize variation and identify strong patterns in EEG features, a principal component analysis (PCA) was applied on **X** which is a broad dataset including explicit listening conditions and persons. All power-based EEG features (i.e., absolute and peak powers) were mapped to logarithmic scale (log-transforming) before applying PCA. Since the EEG features do not have the same scales, the data was normalized using z-score transformation such that each column of **X** re-centered to have zero mean and scaled to have a unit standard deviation.

PCA seeks a linear combination of features such that the maximum variance is extracted from the feature. One of the methods of performing PCA is the singular value decomposition (SVD) method. The SVD decomposes $\mathbf{X}$ into three matrices, i.e., $\mathbf{X} = \mathbf{USV}^{\mathrm{T}}$. The PCA results are expressed by two matrices: (1) the PC loadings (coefficients), $\mathbf{V} \in \mathbb{R}^{p \times p}$, can be understood as the weights for each original variable when calculating the principal component; (2) the PC scores (PCSs), $\mathbf{Us}^{\mathrm{T}} \in \mathbb{R}^{n \times p}$ referring to the representation of $\mathbf{X}$ in the PC space, where $\mathbf{s} \in \mathbb{R}^{p \times 1}$ is the vector containing the main diagonal elements of $\mathbf{S}$ (i.e., the singular values). In other words, each observation in the original space may be projected onto a given PC in order to get a coordinate value along the PC-line. This new coordinate value is known as the PC score. The PC scores are the representation of $\mathbf{X}$ in the PC space. In fact, the PC scores can be calculated with $\mathbf{X}/\mathbf{V}^{\mathrm{T}}$.

2.7. Grouping the Channels in Subregions

The 64 EEG channels were labeled with six fixed subregions: frontal, central, left and right temporal, parietal, and occipital. This division, while allowing four main lobes of cerebrum (Graimann et al., 2010), also considers the central region and left & right hemispheres for the temporal lobe. A similar grouping of channels has been used in previous studies, e.g., for the short-term memory task (Schack et al., 2002). Although subsequent analyses are presented in section 2.8 was performed per channel, the subregion was used a categorical fixed factor in the mixed modeling of EEG-PC scores. However, EEG-PC scores averaged across subregions were used to model the exam result (section 2.9).

2.8. Statistical Analysis of EEG-PC Scores

Linear mixed-effect modeling (LMEM) was used to model the EEG-PC scores as a linear combination of the predictors using the LME4 package (Bates et al., 2015) of the statistical software R (R Core Team, 2019) to explain to origin of EEG-PC scores. The LMEM extends the general linear models (GLMs) to allow both fixed and random effects. A fixed effect is a constant variable across individuals while a random effect varies across individuals. Different LMEMs have been built separately for the nine response variables (EEG-PC scores) as a function of the fixed and mixed (random) effects of interest. Since the person-dependent effects may not be captured in the response variables, the participant variable has been considered as a random effect in all the LMEMs.

On the one hand, the EEG-PC scores were modeled as a function of task type and channel subregion for each specific background noise type based on formula (3), which is hereafter referred to within-background modeling:

$$\mathrm{LMEM}_{\text{within-background}} \leftarrow \mathbf{PCS}_i^j \sim (\mathbf{1}|\mathbf{participant}) + \mathbf{1} + \mathbf{task} + \mathbf{subregion}. \quad (3)$$

In formula (3), $\mathbf{PCS}_i^j \in \mathbb{R}^{n_j \times 1}$ is a vector including *ith* EEG-PC scores for *jth* background noise and all 64 EEG channels, where $i = \{1, ..., 9\}$, $j = \{1, ..., 4\}$ and n_j is the number of observations belonging to all listening tasks in jth background noise. The symbol "$\sim$" implies that left term is modeled as a function of right terms. The fixed effects include **task** and **subregion**. The constant and random terms are expressed in **1** and (**1|participant**), respectively, where **participant** is a categorical variable that has 23 possible outcomes. The term **task** includes the listening task types and has three possible values: lecture attended (LA), background attended (BA), and background unattended (BUA). The last term, **subregion**, is another categorical variable and has six possible outcomes: frontal, parietal, occipital, central, left, and right temporal. Since for each type of background noise, one model is separately defined, no interaction between task and background noise type can be considered.

On the other hand, the EEG-PC scores were modeled as a function of background noise type and channel subregions for each specific listening task based on formula (4), which hereafter referred to within-task modeling:

$$\mathrm{LMEM}_{\text{within-task}} \leftarrow \mathbf{PCS}_i^k \sim (\mathbf{1}|\mathbf{participant}) + \mathbf{1} + \mathbf{background} + \mathbf{subregion}. \quad (4)$$

In formula (4), $\mathbf{PCS}_i^k \in \mathbb{R}^{n_k \times 1}$ is a vector including *i*th EEG-PC scores for *k*th listening task noise and all 64 EEG channels, where $i = \{1, ..., 9\}$, $k = \{1, 2, 3\}$, and n_k is the number of observations belonging to all background noises tasks for *k*th listening task. The term **background** includes the background noise types and takes four possible values: pink (PK), multi-talker (MT), highway (HW), and fluctuating traffic (FT).

After estimating the coefficients (intercept and slope) for each fitted model, general linear hypotheses and Tukey *post-hoc* multiple comparisons were then performed to test for the significance of EEG-PC scores changes across the task and background types. For example, we may consider the six pairwise comparisons between the background noises for the fitted model of the first EEG-PC score in the LA task. The question is which specific background's means (compared with each other) are different. A pairwise Tukey's test examines more than one pair of means the same time and corrects for family-wise error rate.

2.9. Statistical Analysis of Exam Results

As mentioned in the section 2.2, we performed a written exam to check the participant's learning during the lecture attended (LA) task. The exam was carried out after all lectures and the attentive listening to background sounds (see **Figure 1**). Open and closed questions were asked per topic. Open questions were either factual or insight questions. Closed questions consisted of sentences that had to be completed with a specific word or number. The questions were carefully designed so that the answers could be found well-spread over the whole lecture. Over the different topics, the order of question types was randomized. For the open questions, the answers could always be found in three or four connected sentences.

The total number of keywords vary per topic. This was deliberately done to capture as closely as possible anything the participants might have recalled, which is important for the EEG analyses (distinguishing between attention and no attention with remembered keywords as ground truth). The topics of the lectures were chosen to avoid prior knowledge by the participants, yet some topics may be more difficult to grasp than others for the average participant. Moreover, there could be small differences in difficulty between the questions. Prior knowledge and logical reasoning of listeners about the answers are not reflected in listening conditions (background sound) nor in the EEG during listening. Therefore, the number of correctly retained keywords was normalized per participant, background noise and topic and the exam z-scores were calculated as follows:

$$\text{Exam z-score} = \frac{\#\text{Correctly Retained Keywords} - \mu_{\text{pink}}}{\sigma_{\text{pink}}}, \quad (5)$$

where μ_{pink} and σ_{pink} are the mean and standard deviation of the number of correctly retained keywords across all subjects for each topic presented in pink noise (LA in silence), respectively. This a fair reference, as all topics are sufficiently represented in silence.

To validate the predictability of exam results by a linear combination of the EEG-PC scores, we used the linear fixed and mixed-effect modeling as explained in the previous section. In fact, the response variable here is the exam z-scores and the EEG-PC scores are considered as the predictors. Moreover, to show that the EEG contains more information than the listening condition, the exam results was also modeled as a function of background noise type and performance was compared to the models based on EEG.

Person-dependent differences in the exam results may include the following: mental state, traits, physiological differences, prior knowledge, etc. Some of these differences may reflect in EEG, others may not. Hence it is useful to use both linear fixed and mixed-effect modeling. Linear fixed-effect model regresses the exam results as a function of desired fixed factors without considering participant as a random factor, whereas linear mixed-effect model includes participant as a random effect to capture between-subject variability. The latter implies that a fixed offset in exam results per participant is included in the model. Linear fixed-effect models (LFEM) are expressed in the following formulas:

$$\text{LFEM}_{\text{constant}} \leftarrow \textbf{exam z–scores} \sim \mathbf{1}, \quad (6)$$

$$\text{LFEM}_{\textbf{background}} \leftarrow \textbf{exam z–scores} \sim \mathbf{1} + \textbf{background type}, \quad (7)$$

$$\text{LFEM}_{\text{EEG}} \leftarrow \textbf{exam z–scores} \sim \mathbf{1} + \sum_{i=1}^{i=9}\sum_{j=1}^{j=6}\mathbf{PCS}_{ij}^{\text{Avg}}, \quad (8)$$

where exam z-scores (as the response variable) were defined by a vector including all exam z-scores computed by Equation (5). $\mathbf{PCS}_{ij}^{\text{Avg}}$ includes ith EEG-PC scores for jth channel subregion for lecture attended task in all background noises, which were obtained by averaging the PC scores across the channels corresponding to the given subregion (see the section 2.7). The background type term is a categorical variable that has four possible outcomes: pink, multi-talker, highway, and fluctuating traffic.

Similarly, linear mixed-effect models (LMEM) could be expressed in

$$\text{LMEM}_{\text{constant}} \leftarrow \textbf{exam z–scores} \sim (\mathbf{1}|\textbf{participant}) + \mathbf{1}, \quad (9)$$

$$\text{LMEM}_{\textbf{background}} \leftarrow \textbf{exam z–scores} \sim (\mathbf{1}|\textbf{participant}) + \mathbf{1} + \textbf{background type}, \quad (10)$$

$$\text{LMEM}_{\text{EEG}} \leftarrow \textbf{exam z–scores} \sim (\mathbf{1}|\textbf{participant}) + \mathbf{1} + \sum_{i=1}^{i=9}\sum_{j=1}^{j=6}\mathbf{PCS}_{ij}^{Avg}, \quad (11)$$

where exam z-scores and $\mathbf{PCS}_{ij}^{\text{Avg}}$ are defined same as the linear fixed effect models. The constant and random terms are shown by **1** and (**1**|**participant**), respectively.

Since 54 EEG-PC scores (the 9 components for each of the 6 subregions) are available to regress exam z-scores, a stepwise regression method can be used to choose the most contributing predictive variables. The backward-elimination approach was applied on both full models (LFEM_{EEG} and LMEM_{EEG}). To this end, we used "step" function in "STATS" v3.6.2 package of the statistical software R (R Core Team, 2019). This function starts from 54 candidate variables, tests the effect of the deletion of each variable using the Akaike information criterion (AIC) (Akaike, 1974), deletes the variable whose loss gives the least statistically insignificant deterioration of the model fit, and repeats this process until no further variables can be deleted without a statistically significant loss of fit.

3. RESULTS

The results consist of two parts: (1) unsupervised analysis of the EEG features observed under different listening conditions (sections 3.1–3.3) and (2) supervised analysis to predict the exam results in lecture attended task (section 3.4). Section 3.1 presents the loading of principal components (PC) on underlying features; section 3.2 demonstrates the scalp topographies of the PC scores; and section 3.3 explains the relationship between PC scores, listening conditions and backgrounds. In the last section, a supervised training of models was used to investigate the predictability of acquiring and retaining performance scores (exam results) by EEG-PC scores.

3.1. Principal Component Analysis

The explained variances by the ten most important principal components in percent are displayed in **Figure 3A**. Together these ten components explain about 94% of the variability in the dataset. The coordinates of individual EEG feature in principal component (PC) domain are visualized in **Figure 3B**. The correlation between a feature (variable) and a PC is used as the coordinates of the variable on the PC. The size and darkness of circles in **Figure 3B** is proportional to the correlation value between an EEG-feature and a given PC. The positive and negative correlation values are visualized by cool and warm colors, respectively.

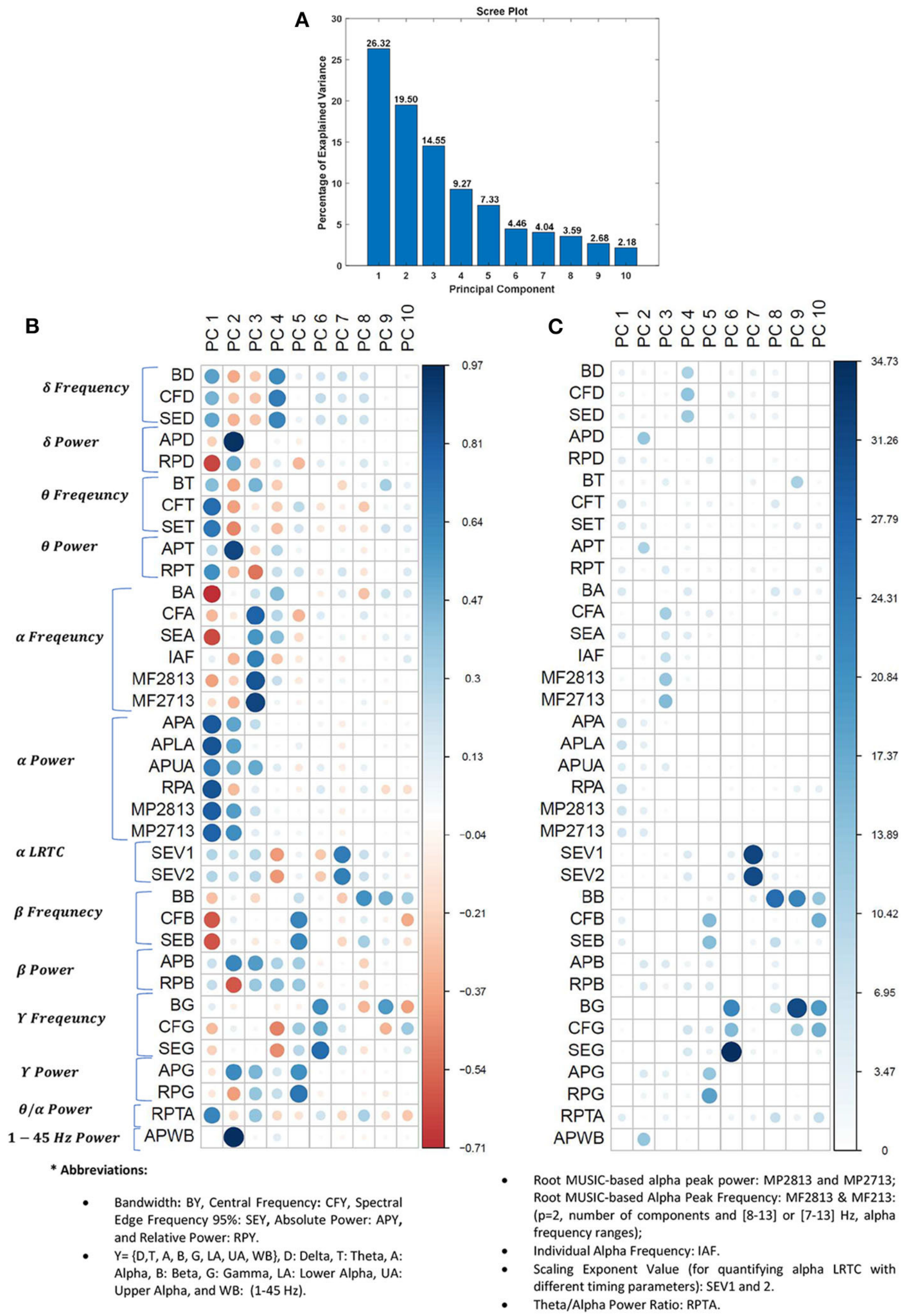


FIGURE 3 | PCA on EEG features. **(A)** Scree plot displays the percentage of explained variance in a downward curve, ordering the eigenvalues from largest to smallest. **(B)** The coordinates of EEG features in PC domain in the rows. The positive and negative correlation values between features and PCs are visualized by cool and warm colors, respectively. **(C)** The contribution of EEG feature to the PCs in percentage, i.e., the squared coordinates were normalized to total sum of squared coordinates on the PCs. The larger and darker circles indicate the EEG features contributes more to the given component. The difference between **(B,C)** is that the **(B)** shows the correlation between features and PCs, while **(C)** shows the representation quality of features on the PCs (i.e., normalized squared correlation values in percentage).

Figure 3C visualizes the contribution of EEG features to the PCs in percentage. The contribution of *ith* EEG feature to *jth* PC is expressed in $\frac{(y_{ij})^2}{\sum_{j=1}^{n}(y_{ij})^2} \times 100$, where y_{ij} is the coordinate of *ith* EEG feature on *jth* PC and $n = 10$ is the number of PCs. In fact, in **Figure 3C**, the squared coordinates were normalized to total sum of squared coordinates on the PCs. The squared coordinates can be a quantity to measure the quality of representation of the features on PC domain.

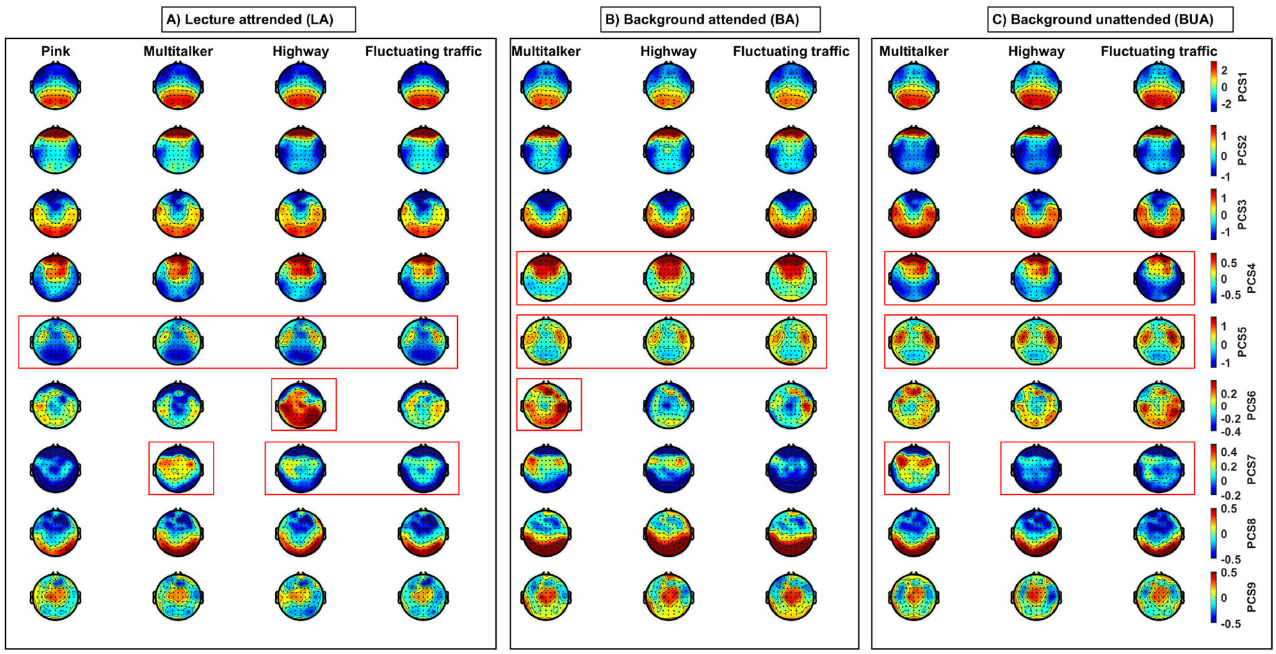


FIGURE 4 | Topographic maps of nine first EEG-PC scores (**PCS**i, i = 1, . . . , 9) of EEG-features (in rows) for three listening tasks (in three panels): **(A)** lecture attended (LA); **(B)** background attended (BA); and **(C)** background unattended (BUA) in different noises. Heads are in vertex view, nose is above, and left ear is on the left side. Each topographic map has been obtained by averaging EEG-PC scores across participants and fragments per listening condition and EEG channel. Each column belongs to one specific listening condition. The type of background noises is shown above the corresponding columns. The warm and cool color-coded areas represent the positively and negatively correlated cortical areas with the extracted components, respectively. Red frames show some spatial and activation differences suggesting the EEG-PC scores might contribute to statistical significance in the discrimination between the three listening tasks and the three background noises. Specifically, (i) PCS4 is higher in BA compared to BUA, (ii) PCS5 is the maximal and minimal in BUA and LA, respectively, (iii) PCS 6 is the maximal for highway in LA and multi-talker in BA, and (vi) PCS7 is the maximal for multi-talker in LA and BUA.

As can be seen from **Figures 3B,C**, the different features contribute to each component. Accurate grouping of these PCs is not possible due to presence of different positively and negatively correlating features with the PC scores (**Figure 3B**). It is worth noting that normalized version of squared coordinates (**Figure 3C**) shows that the last five PCs have more specific loading (representation quality) than those of first five PCs. Specifically, the long-range temporal correlations of alpha band and frequency information of gamma and beta bands are most contributing features to represent PC domain.

3.2. Scalp Topographic Maps

For visualization across the scalp, 2D topographic maps of the component scores are shown in **Figure 4**. The topographies of the nine first PC scores ($\mathbf{PCS}_i, i = 1, ..., 9$) were obtained by averaging across all subjects and the specific fragments for each listening conditions. In fact, for cth EEG channel, jth listening task, and kth noise, the average value of ith PC scores was calculated using $\overline{\text{PCS}}_i^{jk}(c) = \sum_{p=1}^{N} \sum_{f=1}^{l} \mathbf{S}_p^f(c)$, where $\mathbf{S} = \mathbf{PCS}_i^{jk}$, $N = 23$, and l are the number of participants and stimulus fragments, respectively.

Note that here we do not aim at reporting the statistical differences between the listening conditions in terms the PC scores. However, some spatial and activation differences can be observed between different listening conditions (shown by red frames in **Figure 4**) suggesting the EEG-PC scores might contribute to statistical significance (refer to section 3.3) in the discrimination between the three listening tasks and the three background noises. Each component is a linear combination of different positively and negatively correlated features with the components (refer to **Figure 3**). Therefore, in **Figure 4**, both of the warm and cool color coded areas are important, which represent the positively and negatively correlated cortical areas with the extracted components, respectively.

The qualitative differences of some components between different conditions have been shown by red frames in **Figure 4.** Specifically, the PCS 5 is the lowest in the LA compared to other tasks, the PCS 6 is the highest in the highway during the LA, and the PCS 7 is the highest in the multi-taker for the three tasks. In addition, these topographies indicate that different PCs contribute to different cortical areas. For example, the third PC score is positively dominant over temporal and occipital regions.

3.3. Explainable Origin of EEG-PC Scores

The unsupervised extraction of PCs from our dataset implicitly attempts to discriminate between participant, listening task (LA, BA, BUA), and background (MT, HW, FT). One way to analyse the origin of a PC is to construct a regression model for its score based on the above-mentioned factors as explained in section 2.8.

A constant mixed-effect model for predicting EEG-PC scores is expressed in $[\mathbf{PCS}_i \sim (\mathbf{1}|\mathbf{participant})]$, where $\mathbf{PCS}_i$ is ith PC score for all listening conditions and channel subregions. If channel subregion is added as a fixed factor to the constant model, the new model could better predict all PC scores ($p < 10^{-15}$) compared to the constant model. By adding listening task type to the current model, all PC scores except the sixth PC score are better predicted ($p < 10^{-8}$). Background noise type as an additional fixed effect could improve the current model for all PC scores except the ninth PC score ($p < 0.05$). By adding interaction between background noise and task types, the improvement of current model is significant for all PC scores ($p < 10^{-4}$) except the ninth PC score. Since the interaction between task and background noise type significantly improves modeling EEG-PC scores, its effect was separately investigated using two distinct models, within-background and within-task modeling based on formulas (3) and (4), respectively.

Tukey *post-hoc* multiple comparison for within-background and within-task models were reported in **Tables 1**, **2**, respectively (refer to section 2.8). Each test in the sub-matrices was run independently. For example, for a particular background type and PC score, the listening conditions are compared. In each 4×4 and 3×3 sub-matrices, upper triangular elements denote p-values of significant differences for corresponding comparisons, lower triangular elements denote which noise or task results in higher values of the given EEG-PC score and main diagonal elements denote which background noise or task results in the maximum/minimum values of the given EEG-PC score. For example, in **Table 2**, MT $\mathbf{PCS}_1$ is significantly higher than that of PK during LA because $e_{1,4}$ and $e_{4,1}$ elements of matrix corresponding to LA and $\mathbf{PCS}_1$ are $<$ 0.001 and an arrow directed toward MT, respectively. Note $e_{i,j}$ represents the element at the ith row and jth column of the sub-matrix.

In within-background modeling (see **Table 1**), the first PC score is (significantly) the highest and lowest in the BUA and BA tasks for all the background noises, respectively. Moreover, the BUA task has the highest $\mathbf{PCS}_3$ values compared to other tasks for all background noises. For the MT background noise, the LA has the highest $\mathbf{PCS}_2$ compared to other tasks. The fourth PC score exhibits significant contrast between background attended and other tasks for all background noises. The lecture attended can be discriminated from other tasks for all background noises by the fifth PC score. The sixth PC score has a significant contrast between LA and BA tasks in the MT background noise. The seventh PC score is the highest for the BUA compared to other tasks in the MT and HW noises. For all the background sounds $\mathbf{PCS}_8$ is consistently minimal in the BA task. Finally, the ninth PC score is the maximal and minimal for the BA task in the MT and FT sounds, respectively.

In within-task modeling (see **Table 2**), the MT has the highest $\mathbf{PCS}_1$ compared to other background noises in the LA task, whereas in the BUA task, the MT has the lowest $\mathbf{PCS}_1$. The second PC score in the HW is significantly lowest compared to other noises in the LA task. The third PC score exhibits only significant differences in the BA and BUA tasks. The background sounds can be discriminated by the fourth PC score in the LA and BA tasks. The fifth PC score has the highest values in the HW noise during the LA task compared to other background noises. The sixth and seventh PC score are significantly able to distinguish the background sounds for all the listening tasks. The eighth PC score exhibits the highest value for the MT and HW in the LA and BA tasks, respectively. The ninth PC score is not very capable of distinguishing the background sounds.

Remark 1: The statistical results reported in **Tables 1**, **2** have been obtained by eliminating the person-dependent effects, while in the previous section, the topographic maps (**Figure 4**) were obtained by averaging across all subjects without eliminating the person-dependent effects. As a result, the differences are seen in **Figure 4** are not only due to differences between tasks and between noises (like **Tables 1**, **2**) but also due to differences between participants. This means that some of the differences seen in the tables and the topographies are not comparable due to the presence of the effect of the changes between individuals. For example, in the highway noise, although the second PC scores of the BUA task are qualitatively lower than other tasks based on **Figure 4**, **Table 1** shows only the dominance of the LA over the BUA. To explain this difference, we performed Tukey's *post-hoc* testing of linear fixed-effect modeling (without **participant** as a random factor). The *post-hoc* test revealed that BUA<BA ($p < 10^{-5}$) and BUA<LA ($p < 10^{-5}$) meaning that the second PC score can be affected by individual differences likely due to the wideband power ($1 - 45$ Hz) contributing to this component.

Remark 2: Referring to section 2.8, in **Tables 1**, **2**, the results were shown for a model also including the subregions. This implies that a statistically significant difference in one subregion is sufficient for obtaining significant differences. In the maps of **Figure 4**, the reader is expected to interpret the differences in this way. However, the effect of different subregions were separately investigated to model the exam results in the LA task (refer to section 2.9 and 3.4).

3.4. Predictability of Exam Results in Lecture-Attended Task

As noted in section 2.9, the exam z-score defined in Equation (5) is a fairer measure compared to the exam scores ($\frac{\#\text{Correctly Retained Keywords}}{\#\text{Total Keywords}}$) to quantify the amount of information that participants have actually acquired and retained from the lectures. To normalize the exam results (the number of correctly retained keywords) and find the exam z-scores, the exam results of a lecture-attended task in pink noise (lecture in silence) were used. **Figure 5** visualizes the number of correctly retained keywords for lecture attended task in pink noise across thirteen topics. Mean and standard deviation values (μ_{pink} and σ_{pink} in Equation 5) are shown by circles and triangles, respectively. The boxplots display the median marked as a bold line. The lower and upper whiskers represent another 50% data distributed outside the interquartile box. As can be seen from **Figure 5**, the number

TABLE 1 | Tukey *post-hoc* multiple comparison testing for within-background model.

		MT			HW			FT		
		LA	**BA**	**BUA**	**LA**	**BA**	**BUA**	**LA**	**BA**	**BUA**
PCS_1	LA	–	$<10^{-5}$	$<10^{-5}$	–	$<10^{-5}$	$<10^{-5}$	–	<0.05	<0.001
	BA	↑	Min	$<10^{-5}$	↑	Min	$<10^{-5}$	↑	Min	<0.001
(α, $\frac{\theta}{\alpha}$ *pow.*)	BUA	←	←	Max	←	←	Max	←	←	Max
PCS_2	LA	Max	$<10^{-4}$	$<10^{-4}$	–	–	<0.05	–	$<10^{-4}$	–
	BA	↑	–	–	–	–	–	↑	Min	$<10^{-4}$
(δ, θ, *WB pow.*)	BUA	↑	–	–	↑	–	–	–	←	–
PCS_3	LA	–	$<10^{-4}$	<0.001	–	<0.05	<0.05	–	–	$<10^{-5}$
	BA	↑	Min	$<10^{-4}$	↑	Min	<0.001	–	–	$<10^{-5}$
(α *freq.*)	BUA	←	←	Max	←	←	Max	←	←	Max
PCS_4	LA	–	<0.001	<0.05	–	$<10^{-4}$	<0.01	–	$<10^{-4}$	–
	BA	←	Max	<0.001	←	Max	$<10^{-4}$	←	Max	$<10^{-4}$
(δ *freq.*)	BUA	↑	↑	Min	↑	↑	Min	–	↑	–
PCS_5	LA	Min	$<10^{-6}$	$<10^{-6}$	Min	$<10^{-4}$	$<10^{-4}$	Min	$<10^{-4}$	$<10^{-4}$
	BA	←	–	$<10^{-6}$	←	–	<0.01	←	Max	<0.01
(β *freq.*,γ *pow.*)	BUA	←	←	Max	←	←	Max	←	↑	–
PCS_6	LA	Min	$<10^{-4}$	$<10^{-4}$	Max	$<10^{-4}$	$<10^{-4}$	–	<0.001	–
	BA	←	Max	$<10^{-4}$	↑	Min	<0.05	↑	Min	<0.001
(γ *freq.*)	BUA	←	↑	–	–	←	↑	–	←	–
PCS_7	LA	–	–	<0.001	–	<0.001	–	–	<0.01	$<10^{-4}$
(α *LRTC*)	BA	–	–	$<10^{-4}$	↑	–	–	↑	Min	$<10^{-4}$
	BUA	←	←	Max	–	–	–	←	←	Max
PCS_8	LA	–	$<10^{-6}$	–	–	$<10^{-4}$	–	–	$<10^{-4}$	–
	BA	←	Max	$<10^{-6}$	←	Max	$<10^{-4}$	←	Max	$<10^{-4}$
(β *freq.*)	BUA	–	↑	–	–	↑	–	–	↑	–
PCS_9	LA	–	<0.01	–	Min	$<10^{-4}$	$<10^{-4}$	–	<0.001	–
	BA	←	Max	$<10^{-4}$	←	–	–	←	Max	$<10^{-4}$
(β, γ *freq.*)	BUA	–	↑	–	←	–	–	–	↑	–

Tukey variable is the task type (LA, BA, and BUA) and all possible pairs of means in each subtable are compared. Significant p-values are reported in upper triangular. Main diagonal denotes which task is the maximum (Max) and the minimum (Min) compared to other tasks in terms of a given PC score. Lower triangular arrows are directed toward the tasks which have higher PC scores, when comparing two tasks. The non-significant (p >0.05) differences are shown by dash signs. The type(s) and frequency band(s) associated with each component (using the features have the strongest impacts; refer to ***Figure 3C****) are reported in the first column.*

Each element of the sub-matrices is corresponding to the listening tasks labeled above and to the left side: LA, lecture attended, BA, background attended, BUA, background unattended; Each sub-matrix is corresponding to the background noise type: MT, multi-talker; HW, highway; FT, fluctuating traffic and EEG principal component scores (PCS_i) labeled above and to the left side, respectively.

of retained keywords in silence for different topics are different, and hence, the difficulty of retaining information in each topic is different.

In order to assess the effect of background noise type on predicting the exam z-scores, the exam z-scores were modeled using formula (10) and then, Tukey *post-hoc* multiple comparison testing was used to compare the background noise types. The statistical results are reported in **Table 3**. As can be seen, there are significant differences between pink and multi-talker, between pink and highway, and between fluctuating traffic and multi-talker noises. This means that the exam z-scores are higher in the pink noise (LA in silence) than those of in the multi-talker and highway background noise (as we expected). In addition, the fluctuating traffic background noise leads to the higher exam z-scores compared to those of the multi-talker background noise. Therefore, compared to the fluctuating traffic noise, the multi-talker noise leads to more difficult condition for information retention. Note that pink noise refers to a very low-level pink noise (see section 2.2) and means that subjects have listened to the lectures in silence.

TABLE 2 | Tukey *post-hoc* multiple comparison testing for within-task model.

		LA				BA			BUA		
		MT	**HW**	**FT**	**PK**	**MT**	**HW**	**FT**	**MT**	**HW**	**FT**
PCS_1	MT	Max	< 0.05	< 0.001	< 0.001	–	–	< 0.001	Min	< 0.001	< 0.01
	HW	↑	–	–	< 0.001	–	–	< 0.05	–	–	–
(α, $\frac{\theta}{\alpha}$ *pow.*)	FT	↑	–	–	< 0.05	↑	↑	Min	←	↑	–
	PK	↑	↑	↑	Min	–	–	–	–	–	–
PCS_2	MT	–	< 0.001	–	–	–	–	< 0.01	–	–	$< 10^{-4}$
	HW	↑	Min	< 0.001	< 0.001	–	–	–	–	–	$< 10^{-4}$
(δ, θ, WB pow.)	FT	–	←	–	–	←	–	–	←	←	Max
	PK	–	←	–	–	–	–	–	–	–	–
PCS_3	MT	–	–	–	–	Min	$< 10^{-4}$	$< 10^{-4}$	–	–	–
	HW	–	–	–	–	←	–	–	–	–	< 0.01
(α *freq.*)	FT	–	–	–	–	←	–	–	–	←	–
	PK	–	–	–	–	–	–	–	–	–	–
PCS_4	MT	–	–	< 0.05	–	–	< 0.001	–	–	< 0.05	< 0.001
	HW	–	–	< 0.001	< 0.001	←	–	–	←	Max	< 0.001
(δ *freq.*)	FT	↑	↑	–	–	–	–	–	↑	↑	Min
	PK	–	↑	–	–	–	–	–	–	–	–
PCS_5	MT	–	< 0.05	–	–	–	–	–	–	–	< 0.01
	HW	←	Max	< 0.05	< 0.001	–	–	–	–	–	< 0.001
(β *freq.*, γ *pow.*)	FT	↑	–	–	–	–	–	–	↑	↑	Min
	PK	–	↑	–	–	–	–	–	–	–	–
PCS_6	MT	Min	$< 10^{-5}$	$< 10^{-4}$	$< 10^{-5}$	Max	$< 10^{-4}$	$< 10^{-4}$	–	–	< 0.01
	HW	←	Max	$< 10^{-5}$	$< 10^{-5}$	↑	Min	$< 10^{-4}$	–	–	$< 10^{-4}$
(γ *freq.*)	FT	←	↑	–	–	↑	←	–	←	←	Max
	PK	←	↑	–	–	–	–	–	–	–	–
PCS_7	MT	Max	< 0.05	< 0.001	< 0.001	Max	< 0.05	< 0.001	Max	$< 10^{-5}$	$< 10^{-5}$
	HW	↑	–	< 0.001	< 0.001	↑	–	< 0.001	↑	Min	< 0.001
(α *LRTC*)	FT	↑	↑	–	< 0.05	↑	↑	Min	↑	←	–
	PK	↑	↑	↑	Min	–	–	–	–	–	–
PCS_8	MT	–	–	$< 10^{-4}$	$< 10^{-4}$	–	–	–	Max	< 0.05	< 0.001
	HW	–	–	$< 10^{-4}$	$< 10^{-4}$	–	–	< 0.01	↑	–	< 0.001
(β *freq.*)	FT	↑	↑	–	–	–	↑	–	↑	↑	Min
	PK	↑	↑	–	–	–	–	–	–	–	–
PCS_9	MT	–	–	–	–	–	–	–	–	–	–
	HW	–	–	–	< 0.01	–	–	–	–	–	–
(β, γ *freq.*)	FT	–	–	–	< 0.01	–	–	–	–	–	–
	PK	–	←	←	–	–	–	–	–	–	–

*Tukey variable is the background noise type and all possible pairs of means in each subtable are compared. Significant p-values are reported in upper triangular. Data can be decoded like **Table 1**. The type(s) and frequency band(s) associated with each component (using the features have the strongest impacts; refer to **Figure 3C**) are reported in the first column. Each element of the sub-matrices is corresponding to the background noise type labeled above and to the left side: MT, multi-talker; HW, highway; FT, fluctuating traffic; Each sub-matrix is corresponding to the listening tasks: LA, lecture attended, BA, background attended, BUA, background unattended and EEG principal component scores (PCS_i) labeled above and to the left side, respectively.*

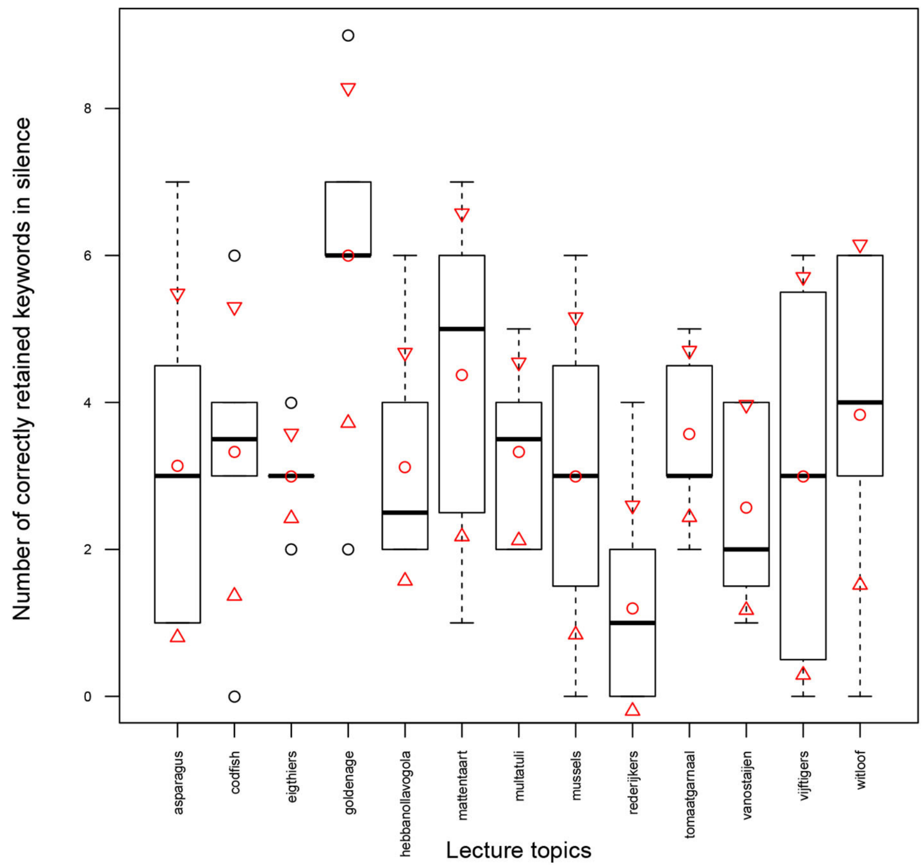


FIGURE 5 | Number of correctly retained keywords in silence (LA-PK) across 13 topics; mean and standard deviation values are shown by circles and triangles, respectively. The box-plots display the median marked as a bold line. The lower and upper whiskers represent another 50% data distributed outside the interquartile box.

TABLE 3 | Effect of background noise on exam z-score; Tukey *post-hoc* multiple comparisons between different types of background noise for modeling exam z-score in lecture attended task (using mixed-effect modeling).

Background type	Pink	Highway	Multi-talker	Fluctuating traffic
Pink	–	$p < 0.05$	$p < 0.01$	–
Highway	↑	–	–	–
Multi-talker	↑	–	–	$p < 0.05$
Fluctuating traffic	–	–	←	–

Tukey variable is the background noise type and all possible pairs of means are compared. Upper triangular elements indicate corresponding p-values (p) between two background noises (if p <0.05). Lower triangular arrows are directed toward the background noises which have higher (better) exam z-scores.

To identify the link between EEG-PC scores and the exam z-scores, both fixed and mixed-effect models were employed as presented in section 2.9. Note that the EEG-PC scores used in this section were obtained by averaging across the channels corresponding to the given subregions. The models were compared using two criteria. First, χ^2 test was used to compare between the two models using "anova" function in STATS v3.6.2 package of the statistical software R (R Core Team, 2019). A good model not only needs to fit data well—it also needs to be parsimonious. This criterion takes the model objects as arguments and returns an ANOVA testing whether or not the more complex model is significantly better at capturing the data than the simpler mode. If the resulting p-value is <0.05, we conclude that the more complex model is significantly better than the simpler model. If the p-value is >0.05, we should favor the simpler model.

The second criterion used to compare the fitted models was the Akaike information criterion (AIC) (Akaike, 1974). When comparing models fitted by maximum likelihood to the same data, a lower AIC value indicates a better fit. We have used "extractAIC" function in STATS v3.6.2 package of the statistical software R (R Core Team, 2019). The following equation is used to estimate AIC: $-2\log(L)+(k\times edf)$, where $k=2$, L refers to the likelihood, and edf stands for the equivalent degrees of freedom (i.e., the number of free parameters for the models) of fit.

Table 4A reports the predictability of exam z-scores based on linear fixed-effect modeling (without considering participant as a random factor). The following predictors (fixed factors) were used: (1) no fixed factor (constant), (2) background type, and (3) 54 EEG-PC scores as defined by formulas (6), (7), and (8), respectively. Furthermore, a stepwise fixed-model regression was performed to regress exam z-score using the most significant EEG-PC scores (refer to section 2.9). P-values shown on the

TABLE 4 | Predictability of exam z-scores using **(A)** fixed-effect and **(B)** mixed-effect models.

(A) Fixed-effect model: [Exam z-score ∼ 1 + Fixed Factor].

Fixed factor	**Constant**	**Background type**	**54 EEG-PC scores**	**Stepwise EEG-PC scores**[a]
Constant	$AIC = 134.26$	$p < 0.01$	$p < 10^{-12}$	$p < 10^{-15}$
Background type	←	$AIC = 128.67$	$p < 10^{-10}$	$p < 10^{-15}$
54 EEG-PC scores	←	←	$AIC = 89.73$	–
Stepwise EEG-PC scores	←	←	←	$AIC =$ **38.72**

[a]Contributing PC scores (PCSs):

$p < 10^{-4}$ → Parietal PCS 7 (−0.86);

$p < 10^{-3}$ → Central PCS1 (0.63);

$p < 0.01$ → Occipital PCS 1 (−0.51), Occipital PCS 2 (−0.33), Occipital PCS 7 (0.50), Occipital PCS 9 (−0.32), Frontal PCS 4 (−0.33), Central PCS 5 (0.57), Central PCS 8 (−0.41), Left Temporal PCS 4 (0.52);

$p < 0.05$ → Occipital PCS 4 (−0.25), Frontal PCS 3 (−0.16), Parietal PCS 1 (0.42), Parietal PCS 5 (−0.40), Parietal PCS 8 (0.25), Left Temporal PCS 2 (0.65), Left Temporal PCS 1 (−0.53), Left Temporal PCS 6 (0.26);

•$p < 0.2$ → Frontal PCS 6 (−0.14), Frontal PCS 7 (0.14), Left Temporal PCS 3 (0.12), Left Temporal PCS 5 (−0.18), Right Temporal PCS 2 (−0.42).

(B) Mixed-effect model: [Exam z-score ∼ (1|Participant) + 1 + Fixed Factor].

Fixed effects	**Constant**	**Background type**	**54 EEG-PC scores**	**Stepwise EEG-PC scores**[b]
Constant	$AIC = 839.75$	$p < 0.01$	$p < 10^{-3}$	$p < 10^{-7}$
Background type	←	$AIC = 830.94$	$p < 0.01$	$p < 10^{-5}$
54 EEG-PC scores	←	←	$AIC = 851.45$	–
Stepwise EEG-PC scores	←	←	←	$AIC =$ **806.76**

[b]Contributing PCSs:

$p < 10^{-5}$ → Parietal PCS 7 (−0.37)

$p < 10^{-4}$ → Occipital PCS 2 (−0.47);

$p < 0.001$ → Central PCS 1 (0.68), Left Temporal PCS 2 (0.35);

$p < 0.01$ → Central PCS 8 (−0.45);

$p < 0.05$ → Occipital PCS 1 (−0.17), Central (0.39) and Parietal (−0.38) PCS 5, Parietal PCS 4 (−0.12), Parietal PCS 8 (0.37), Left Temporal PCS 1 (−0.41).

Upper triangular elements are pairwise p-values (p) when two models are compared using χ^2 test (if p <0.05). Lower triangular arrows are directed toward the better models when comparing two models. If the resulting p-value is <0.05, the more complex model is significantly better at capturing the data than the simpler model. If the p-value is > 0.05, we favor the simpler model. Main diagonal elements indicate AIC values for given models. The lowest AIC value (corresponding to the best model) is shown in bold. The PC scores obtained by the stepwise method are reported below the tables and (•) denotes the regression coefficient (slope) of each factor.

upper diagonal of **Table 4A**, suggest that there are pairwise significant differences between all models except between two models which use 54 EEG-PC scores and stepwise EEG-PC scores as the fixed factors. This means that the stepwise model (simpler model) is better than the full model (more complex model) in terms of χ^2 test criterion.

AIC values shown on the main diagonal of **Table 4A**, suggest that stepwise EEG-PC scores can predict the exam z-scores better than other models (the lowest AIC). We found the 23 predictors that play more significant roles to predict the exam z-scores. The names of these predictors, their *p*-values (to predict the exam z-scores), and their coefficient (slope in regression) are reported below **Table 4A**. They were ordered according to their statistical significance. As can be seen from **Table 4**. the parietal PC score 7 (related to alpha LRTC), which is negatively correlated with exam z-scores, is the most important predictor to model the exam z-scores using the linear fixed-effect modeling.

The results of the mixed-effect models (formulas 9–11) to model exam z-scores are presented in **Table 4B**. By including the participant as a random factor, the models are less likely to be affected by individual differences. Therefore, those EEG features that contribute to differentiate between participants are expected to be less relevant in this modeling. In contrast to the fixed-effect model, in the mixed-effect model, background noise type can better predict the exam z-scores compared to 54 EEG-PC scores (in terms of AIC and not χ^2 test). However, the stepwise EEG-PC scores results in a significantly better model than knowing background noise type to predict exam z-scores (the lowest AIC).

According to the tables, in the both fixed and mixed-effect models, the modes which use stepwise EEG-PC scores predict the exam z-scores better than all other models. It is worth noting that unlike the fixed-effect model which all the components in the certain subregions play the significant roles in predicting, in the mixed-effect model, the most contributing predictors are limited to the PCS 1, 2, 4, 5, 7, and 8 in the particular subregions (as can be seen from below **Table 4B**). These results are consistent with the results of section 3.3, where the importance of these components (especially PCS 7) to distinguish between the background noises in the lecture attended task was shown (refer to **Table 2**). The relationship between these components and hypotheses presented in the introduction section and their underlying mechanisms will be discussed in the next section.

4. DISCUSSION

The present study used a single-trial 64-channel EEG measurement and ecologically valid stimuli to investigate the neural correlates of acquiring and retaining vocally presented information. To identify significant EEG components, a broad set of three listening tasks were performed: (1) attentive listening to 5-min lectures in the environmental sound (LA), (2) attentive listening to environmental sounds (BA), and (3) inattentive listening to environmental sounds (BUA). The environmental sounds included multi-talker, highway, and fluctuating traffic sounds. During this unsupervised learning step, a wide range of features of sensor-space EEG signals were collected and their principal component scores (PCSs) were calculated. Unlike the attention decoding studies that aim to explicitly decode an attended from unattended speech stream based on the supervised approach (Horton et al., 2014; O'Sullivan et al., 2014), we aimed to distinguish between attentive and inattentive listening conditions. To this end, we used an unsupervised learning method that, as such, did not require knowledge of the attended sound signal.

During the LA task, the mixture of verbal lectures and different types of background noise were presented. The lectures were related to topics for which prior knowledge is expected to be minimal. A written exam was taken after the experiment to quantify the amount of information that participants have acquired and retained from the lectures. Since the exam included the questions related to fact and insight, memory is expected to be more specifically involved. It is worth noting the following: (1) although the background sounds could distract the participants while listening to the speech, they did not mask the speech energetically, and (2) no visual distractor was presented during the experiment.

4.1. Essential EEG-PC Scores to Predict the Exam Results

The predictability of exam results of the LA task by the EEG-PC scores (EEG-PCSs) has been assessed by linear fixed and mixed-effect modeling of the exam z-scores. It is expected that differences in the exam z-scores can arise from the instantaneous listening state but also from the overall state, personal traits, physiology, and prior knowledge, hence both fixed and mixed-effect models were used to regress the exam z-scores. The fixed-effect model, not considering participant as a random factor, assumes that all relevant differences for predicting exam z-scores are visible in EEG-PCS, whereas the mixed-effect model, considering participant as a random factor, assumes some personal differences are not visible in the EEG-PCS. We first consider the latter approach.

Firstly, it could be confirmed that knowing the type of background sound improves the predictability of exam z-scores (refer to **Table 4**). Exams on information presented in background noise always gave significantly lower scores, except for fluctuating traffic noise that did not seem to significantly affect exam z-scores (refer to **Table 3**). Note that in our experiment, noise may affect speech perception, listening comprehension, distraction, and memory encoding. Speech perception in noise was found to be consistently worse in babble than in traffic noise in previous research (Shukla et al., 2018). For episodic memory tasks, it was found that encoding under traffic noise and meaningful irrelevant speech were worse than under silent conditions, but scores were lower for traffic noise than for competing meaningful speech (Hygge et al., 2003). Thus, our results seem to confirm previous works. We can now turn to the question of whether EEG allows us to disentangle the multitude of interacting effects that play a role.

A stepwise mixed-effect model identified that a few specific EEG-PCSs play a more significant role in modeling the exam z-scores (refer to **Table 4B**). These EEG-PCSs are the central, occipital, and left temporal PCS 1, the occipital and left temporal PCS 2, the parietal PCS 4, the central and parietal PCS 5, the parietal PCS 7, the central, and parietal PCS 8. The underlying mechanisms of these components and their links with our hypotheses are discussed based on the unsupervised learning phase and the previous studies as follows.

- The first component: overall attentive state

In general, the alpha-band activity has been assumed as an idling rhythm (Pfurtscheller et al., 1996) meaning the power of alpha activity increases during resting state and conditions of mental inactivity. During the cognitive effort, alpha activity usually diminishes, which is referred to as alpha desynchronization (Pfurtscheller and Da Silva, 1999; Sauseng et al., 2005). In addition, previous studies have argued increased occipital (task-irrelevant) and decreased frontal (task-relevant) alpha activity can reflect the distracted auditory attention (Pfurtscheller and Da Silva, 1999; Sauseng et al., 2005; Clayton et al., 2015). Our results showed the occipital and PCS 1 is negatively correlated with the exam z-scores ($p = 0.02$, $s = -0.17$, where s is the slope of corresponding factors in the linear regression). Based on the results yielded by PCA (refer to **Figure 3**), the alpha peak power and alpha bandwidth are the most positively and negatively contributing feature to this component, respectively. Therefore, an increase in the exam z-scores can be associated with a decrease in this component score due to overall mind wandering and distracted attention. This statement is in accordance with the unsupervised analysis results where the multi-talker and pink (lecture in silence) PCS 1 is the maximal (the least attention) and minimal (the highest attention) compared to other background sounds during the lecture attended task (see **Table 2**). In addition, the ratio of theta to alpha power (RPTA in **Figure 3**) also positively contributes to this component which also confirms that an increase in PCS 1 indicates the deterioration in attention (in agreement with Holm et al., 2009; Borghini et al., 2014).

- The fourth component: low-frequency speech envelope following

The parietal PCS 4 is negatively correlated with the exam z-scores ($p = 0.027$, $s = -0.12$). The fourth PC is strongly determined by various characteristics of the delta frequency band, such as bandwidth, central frequency, and spectral edge frequency (refer to **Figure 3**). This frequency band is observed during speech envelope following (Kerlin et al., 2010; Ding and Simon, 2014; Vanthornhout et al., 2019). In addition, the gamma central frequency and the alpha-band LRTC negatively contribute to the fourth PCS and are visible in the occipital, temporal, and parietal regions (see **Figures 3**, **4**). The unsupervised analysis revealed that the fourth PCS exhibits the highest and lowest values in background attended and unattended tasks, respectively (see **Table 1**). Therefore, the fourth PCS may reflect speech envelope following and listening attentively without necessarily linguistic processing or gating out (our third hypothesis). This interpretation could be consistent with the lower values (more negative values) of the parietal and occipital fourth PCS in fluctuating traffic noise compared to other background noises in the lecture attended and background unattended tasks (refer to **Figure 4**).

- The fifth component: decreased focusing during listening

The parietal fifth PCS exhibits a reverse relationship with the exam z-scores ($p = 0.020$, $s = -0.38$). The positively contributing EEG features to the fifth PCS include the beta central frequency and the gamma absolute power. Based on the unsupervised analysis, the fifth PCS is the lowest in the lecture attended (LA) task compared to other tasks for all background noises (refer to **Table 1**). Therefore, decreased fifth PCS is likely associated with more focus during listening, where the exam scores are expected to improve as well.

- The sixth component: cognitive prediction error

Although the sixth PCS is not obtained from the mixed-effect stepwise regression as a contributing component, the left temporal PCS 6 is the most significant component obtained from the full mixed model ($p = 0.02$, $s = 0.55$). The sixth PCS positively loads on the gamma spectral edge frequency, bandwidth, and central frequency. Moreover, the frontal and central sixth PCS is negatively correlated with the exam z-scores ($s = -0.10$ and $s = -0.20$). Based on the unsupervised analysis, the sixth PCS is more discriminating between the background noises. Its highest values are observed for attended speech in continuous highway sound (LA-HW) and for attended multi-talker sound (BA-MT) (refer to **Table 2**). These two conditions have in common that one may rely on linguistic processing and prediction to complete the information. This factor is therefore likely associated with predictive coding. Higher values of the sixth PCS result in lower exam z-scores which may be explained by the fact that a need for prediction to complete the information may result in poor encoding. This finding is in line with Bastos et al. (2012), Sedley et al. (2016), and Alexandrou et al. (2017) where has been shown the prediction violations or errors (our fifth hypothesis) are encoded by gamma-band activity (especially over higher brain areas). It was also found that this component over the left temporal region is positively correlated with the exam z-scores reflecting task-relevant gamma-band activity role on speech processing in alignment with Giraud et al. (2007), Morillon et al. (2012), and Alexandrou et al. (2017).

- The seventh component: alpha-as-inhibition and inhibition-excitation balance

The parietal seventh PCS, which positively loads on alpha-band LRTC, is negatively correlated with the exam z-scores ($p = 8 \times 10^{-6}$, $s = -0.37$). Interestingly, this PC score in multi-talker noise and independent of task type is significantly dominant compared to other background noises. Increased alpha-band LRTC reflects that the autocorrelations of alpha activity slower decay in power-law behavior and as a result, the self-similarity of alpha activity increases. In fact, high levels of alpha-band LRTC reflect the enduring alpha waves. In agreement with Poil et al. (2012), this increased self-similarity or long-lasting changes could reflect more balance between excitation and inhibition states of alpha-band activity during the auditory stimulus (our second hypothesis). Both excitation and inhibition sates are therefore involved during attentive listening to the lecture in multi-talker sound, which is required for more listening effort due to multi-talker distraction. In contrast to multi-talker, attentive listening to lectures in pink noise (lecture in silence), the alpha-band LRTC is the lowest compared to other noises due to less need for inhibition during listening. In fact, during this listening condition, the excitation state is more dominant than the inhibition state. Increased PCS 7 could thus be associated with a higher inhibition-excitation balance. This component can be linked to the alpha-as-inhibition (Clark, 1996; Uusberg et al., 2013) hypothesis (our first hypothesis) where alpha synchronization reflects suppression of irrelevant information (inhibition).

For the fixed-effect model, where all differences between people are assumed to be explainable through EEG, also adding second (over non-occipital regions), third, sixth, and ninth PCSs improves the predictability of exam results (refer to **Table 4A**). The second PCS loads strongly on the wide-band absolute power and absolute powers in the low-frequency bands (delta and theta). It is probably related to the observability of EEG for each specific person and may not indicate specific brain-related functions. The third PCS mainly loads on alpha peak frequency, alpha central frequency, and related factors. As PCS 1, the third PCS is significantly higher in the BUA task. Literature is not univocal on the expected trends in relation to tasks (Angelakis et al., 2004; Mierau et al., 2017) but points at a significant difference between persons (Klimesch et al., 1993; Haegens et al., 2014). The latter may explain why PCS 3 only occurs as a significant predictor in the fixed-effect model where it helps to differentiate between persons.

4.2. EEG-PC Scores Related to Task Difficulty-Based Cognitive Load

In this experiment, adding background sound to the lectures increases the effort needed to process the sound, but it may also affect cognitive load and task difficulty. The cognitive load of subjects has been assessed from different perspectives using EEG depending on the type of task. For instance, the task difficulty during the intelligence test (Friedman et al., 2019) and learning task (Mills et al., 2017) as the cognitive load has been linked to EEG features. Moreover, the cognitive load during a visual task has been associated with the attentional demand using an ERP analysis (Grassini et al., 2019). There is no unique EEG feature that is directly related to cognitive load. Theta power has been suggested as an indicator for the average cognitive load of subjects and the linguistic complexity of educational videos (Castro-Meneses et al., 2019). Mu rhythm oscillations (8 − 13 Hz over the sensorimotor cortex) could be affected by the cognitive load during speech perception due to attention and working memory processes (Jenson et al., 2019). In addition to the task difficulty, the listener's skill also may affect the cognitive load.

In this paper, although the cognitive load of listeners has not been explicitly investigated, some PCSs may reflect the task difficulty-based cognitive load, such as the sixth and seventh PCSs (reflecting the prediction error and the inhibition during listening, respectively). However, caution is needed to link neural results to these behavioral outcomes as this study is based on a sample of young adults only. Aging populations might react differently.

Since there are more noiseless gaps during fluctuating traffic sound compared to the highway sound (refer to **Figure 2**), it is expected that less mental resources are needed to predict the missing part (less PCS 6) during LA in fluctuating traffic sound. Therefore, LA in the highway sound (LA-HW) is likely more difficult task compared to LA in the fluctuating traffic sound (LA-FT). However, the task difficulty can be reflected either in the continuous inhibition by increased PCS7 (highway sound) or in the fluctuating inhibition by decreased PCS7 (fluctuating traffic sound). Moreover, in the BUA task, the fluctuating traffic sound is the most difficult sound to predict (the highest PCS6) compared to the other sounds. Although the BUA in the multi-talker sound exhibits more inhibition compared to the fluctuating traffic (higher PCS7), the multi-talker sound in the BUA can be easier predicted (lower PCS 6) compared to the fluctuating traffic sound. These findings may explain the impacts of different types of environmental sound during daily activities.

5. CONCLUSION

The current study showed that it is possible to predict beyond the chance level the amount of vocal information that participants acquire and retain from the lectures presented in different environmental sounds using 64-channel EEG. Five principal component scores of the EEG features obtained under different listening conditions and for different persons were essential for this prediction. Based on their loading on the spectral range and their ability to distinguish between listening tasks, we associate them with overall attentive state, speech envelope following (listening attentively without necessarily linguistic processing), focusing during listening, cognitive prediction error, and specific inhibition. Part of the variance between persons could further be explained by principal component scores that tend to relate to overall signal strength, an indication of observability of EEG signals, and person identification through inter-individual differences between typical alpha peak frequencies.

Inhibition-excitation balance (reflected by alpha-band representation) and predictive mechanisms (reflected by gamma-band representation) play a more important role than might have been expected and could be observed via EEG. Furthermore, the results of comparing the principal components scores of three different auditory tasks (attentive listening to the lecture in environmental noise, attentive listening to the environmental sound, and inattentive listening to the environmental sound) showed the extracted principal components scores are able to discriminate the different listening tasks and background noises. Specifically, (i) the sixth and seventh principal component scores, which reflect prediction error and inhibition-excitation balance, respectively, allow us to distinguish different types of background sound. Moreover, (ii) the type of listening tasks could be completely distinguished by the first and fifth principal component scores, which reflect the overall attentive state and decreased focusing, respectively.

In terms of methodology, by combining different listening conditions to train in an unsupervised way the definition of orthogonal features based on EEG, a more efficient supervised model for the prediction of the memorization of information could be obtained. This methodology could be relevant for assessing the impact of environmental sounds on daily activities, such as communicating, learning, and relaxing as some of the principal components identified could be related to increased cognitive load. They could also be relevant for future artificial intelligence communicating optimally with humans based on observed brain activity. The methodology also allows us to assess individual differences in the ability to process speech in noise.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because further analysis is ongoing. Requests to access the datasets should be directed to the first author. The Matlab® and R® codes implementing the algorithms and statistical analyses are publicly accessible on GitHub (https://github.com/EhsanEqlimi/EEG-Correlates-of-Learning-From-Speech-Presented-in-Environmental-Noise).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by International Laboratory for Brain, Music and Sound Research (BRAMS), Montreal, Canada. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

EE carried out the data analysis and interpretation, signal processing, statistical analysis, and writing of the manuscript. AB carried out the data acquisition, the experiment design, study idea, statistical analysis, data interpretation, and editing of the manuscript. BD carried out the data interpretation and the editing the manuscript. MS carried out the experiment design, data acquisition and interpretation, and editing of the manuscript. DT carried out the data interpretation and editing of the manuscript. DB carried out the original idea for study, data interpretation, experiment design, and editing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was part of DUCK project (Distraction from learning by Unrelated auditory events assessed by Computational modeling and Knowledge extraction from single-trial electroencephalography). This research received funding from the Flemish Government under the Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen programme and the Belgium Special Research Fund (Bijzonder Onderzoeksfonds, BOF).

ACKNOWLEDGMENTS

We gratefully acknowledge the assistance and support of International Laboratory for Brain, Music and Sound Research (BRAMS), Montreal, Canada for collecting the EEG data.

REFERENCES

Ahveninen, J., Huang, S., Belliveau, J. W., Chang, W.-T., and Hämäläinen, M. (2013). Dynamic oscillatory processes governing cued orienting and allocation of auditory attention. *J. Cogn. Neurosci.* 25, 1926–1943. doi: 10.1162/jocn_a_00452

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19, 716–723. doi: 10.1109/TAC.1974.1100705

Alain, C. (2007). Breaking the wave: effects of attention and learning on concurrent sound perception. *Hear. Res.* 229, 225–236. doi: 10.1016/j.heares.2007.01.011

Alexandrou, A. M., Saarinen, T., Mäkelä, S., Kujala, J., and Salmelin, R. (2017). The right hemisphere is highlighted in connected natural speech production and perception. *Neuroimage* 152, 628–638. doi: 10.1016/j.neuroimage.2017.03.006

Angelakis, E., Lubar, J. F., Stathopoulou, S., and Kounios, J. (2004). Peak alpha frequency: an electroencephalographic measure of cognitive preparedness. *Clin. Neurophysiol.* 115, 887–897. doi: 10.1016/j.clinph.2003.11.034

Barabell, A. (1983). "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *ICASSP'83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 8 (Boston, MA: Citeseer), 336–339. doi: 10.1109/ICASSP.1983.1172124

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003

Braboszcz, C., and Delorme, A. (2011). Lost in thoughts: neural markers of low alertness during mind wandering. *Neuroimage* 54, 3040–3047. doi: 10.1016/j.neuroimage.2010.10.008

Braeutigam, S., Bailey, A. J., and Swithenby, S. J. (2001). Phase-locked gamma band responses to semantic violation stimuli. *Cogn. Brain Res.* 10, 365–377. doi: 10.1016/S0926-6410(00)00055-0

Buzsaki, G. (2006). *Rhythms of the Brain*. Oxford: Oxford University Press.

Buzsáki, G., Logothetis, N., and Singer, W. (2013). Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron* 80, 751–764. doi: 10.1016/j.neuron.2013.10.002

Castro-Meneses, L. J., Kruger, J.-L., and Doherty, S. (2019). Validating theta power as an objective measure of cognitive load in educational video. *Educ. Technol. Res. Dev.* 68, 181–202. doi: 10.1007/s11423-019-09681-4

Cervenka, M. C., Nagle, S., and Boatman-Reich, D. (2011). Cortical high-gamma responses in auditory processing. *Am. J. Audiol.* 20, 171–180. doi: 10.1044/1059-0889(2011/10-0036)

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* 13:1428. doi: 10.1038/nn.2641

Chaytor, N., and Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. *Neuropsychol. Rev.* 13, 181–197. doi: 10.1023/B:NERV.0000009483.91468.fb

Cherry, C. (1978). *On Human Communication: A Review, A Survey, and A Criticism*. Cambridge, MA: MIT press.

Clark, C., and Sörqvist, P. (2012). A 3 year update on the influence of noise on performance and behavior. *Noise Health* 14:292. doi: 10.4103/1463-1741.104896

Clark, C. R., Veltmeyer, M. D., Hamilton, R. J., Simms, E., Paul, R., Hermens, D., et al. (2004). Spontaneous alpha peak frequency predicts working memory performance across the age span. *Int. J. Psychophysiol.* 53, 1–9. doi: 10.1016/j.ijpsycho.2003.12.011

Clark, J. M. (1996). Contributions of inhibitory mechanisms to unified theory in neuroscience and psychology. *Brain Cogn.* 30, 127–152. doi: 10.1006/brcg.1996.0008

Clayton, M. S., Yeung, N., and Kadosh, R. C. (2015). The roles of cortical oscillations in sustained attention. *Trends Cogn. Sci.* 19, 188–195. doi: 10.1016/j.tics.2015.02.004

Colosio, M., Shestakova, A., Nikulin, V. V., Blagovechtchenski, E., and Klucharev, V. (2017). Neural mechanisms of cognitive dissonance (revised): an EEG study. *J. Neurosci.* 37, 5074–5083. doi: 10.1523/JNEUROSCI.3209-16.2017

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Ding, N., Chatterjee, M., and Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage* 88, 41–46. doi: 10.1016/j.neuroimage.2013.10.054

Ding, N., and Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* 33, 5728–5735. doi: 10.1523/JNEUROSCI.5297-12.2013

Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8:311. doi: 10.3389/fnhum.2014.00311

Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85, 761–768. doi: 10.1016/j.neuroimage.2013.06.035

Doppelmayr, M., Klimesch, W., Pachinger, T., and Ripper, B. (1998). Individual differences in brain dynamics: important implications for the calculation of event-related band power. *Biol. Cybernet.* 79, 49–57. doi: 10.1007/s004220050457

Drummond, J., Brann, C., Perkins, D., and Wolfe, D. (1991). A comparison of median frequency, spectral edge frequency, a frequency band power ratio, total power, and dominance shift in the determination of depth of anesthesia. *Acta Anaesthesiol. Scand.* 35, 693–699. doi: 10.1111/j.1399-6576.1991.tb03374.x

Engel, A. K., and Fries, P. (2010). Beta-band oscillations-signalling the status quo? *Curr. Opin. Neurobiol.* 20, 156–165. doi: 10.1016/j.conb.2010.02.015

Eqlimi, E., Bockstael, A., De Coensel, B., and Botteldooren, D. (2019). "Evaluating potential EEG-indicators for auditory attention to speech in realistic environmental noise," in *23rd International Congress on Acoustics (ICA 2019)* (Aachen: Deutsche Gesellschaft für Akustik), 7631–7638.

Estrada, E., Nazeran, H., Nava, P., Behbehani, K., Burk, J., and Lucas, E. (2004). "EEG feature extraction for classification of sleep stages," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 1 (San Francisco, CA: Springer; IEEE), 196–199. doi: 10.1109/IEMBS.2004.1403125

Foxe, J. J., Simpson, G. V., and Ahlfors, S. P. (1998). Parieto-occipital 10 Hz activity reflects anticipatory state of visual attention mechanisms. *Neuroreport* 9, 3929–3933. doi: 10.1097/00001756-199812010-00030

Freeman, W., and Quiroga, R. Q. (2012). *Imaging Brain Function With EEG: Advanced Temporal and Spatial Analysis of Electroencephalographic Signals*. New York, NY; Heidelberg; Dordrecht; London: Springer Science & Business Media.

Friedman, N., Fekete, T., Gal, Y. K., and Shriki, O. (2019). EEG-based prediction of cognitive load in intelligence tests. *Front. Hum. Neurosci*. 13:191. doi: 10.3389/fnhum.2019.00191

Gazzaley, A., and Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends Cogn. Sci*. 16, 129–135. doi: 10.1016/j.tics.2011.11.014

Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., and Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56, 1127–1134. doi: 10.1016/j.neuron.2007.09.038

Goswami, U., and Leong, V. (2013). Speech rhythm and temporal structure: converging perspectives. *Lab. Phonol*. 4, 67–92. doi: 10.1515/lp-2013-0004

Graimann, B., Allison, B. Z., and Pfurtscheller, G. (2010). *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*. Heidelberg; Dordrecht; London; New York, NY: Springer Science & Business Media.

Grassini, S., Revonsuo, A., Castellotti, S., Petrizzo, I., Benedetti, V., and Koivisto, M. (2019). Processing of natural scenery is associated with lower attentional and cognitive load compared with urban ones. *J. Environ. Psychol*. 62, 1–11. doi: 10.1016/j.jenvp.2019.01.007

Gulbinaite, R., van Viegen, T., Wieling, M., Cohen, M. X., and VanRullen, R. (2017). Individual alpha peak frequency predicts 10 Hz flicker effects on selective attention. *J. Neurosci*. 37, 10173–10184. doi: 10.1523/JNEUROSCI.1163-17.2017

Haegens, S., Cousijn, H., Wallis, G., Harrison, P. J., and Nobre, A. C. (2014). Inter- and intra-individual variability in alpha peak frequency. *Neuroimage* 92, 46–55. doi: 10.1016/j.neuroimage.2014.01.049

Hald, L. A., Bastiaansen, M. C., and Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain Lang*. 96, 90–105. doi: 10.1016/j.bandl.2005.06.007

Hardstone, R., Poil, S.-S., Schiavone, G., Jansen, R., Nikulin, V. V., Mansvelder, H. D., et al. (2012). Detrended fluctuation analysis: a scale-free view on neuronal oscillations. *Front. Physiol*. 3:450. doi: 10.3389/fphys.2012.00450

He, B. J. (2014). Scale-free brain activity: past, present, and future. *Trends Cogn. Sci*. 18, 480–487. doi: 10.1016/j.tics.2014.04.003

He, B. J., Zempel, J. M., Snyder, A. Z., and Raichle, M. E. (2010). The temporal structures and functional significance of scale-free brain activity. *Neuron* 66, 353–369. doi: 10.1016/j.neuron.2010.04.020

Herrmann, B., Henry, M. J., Haegens, S., and Obleser, J. (2016). Temporal expectations and neural amplitude fluctuations in auditory cortex interactively influence perception. *Neuroimage* 124, 487–497. doi: 10.1016/j.neuroimage.2015.09.019

Holm, A., Lukander, K., Korpela, J., Sallinen, M., and Müller, K. M. (2009). Estimating brain load from the EEG. *Sci. World J*. 9, 639–651. doi: 10.1100/tsw.2009.83

Horton, C., Srinivasan, R., and D'Zmura, M. (2014). Envelope responses in single-trial EEG indicate attended speaker in a 'cocktail party'. *J. Neural Eng*. 11:046015. doi: 10.1088/1741-2560/11/4/046015

Houtgast, T., and Steeneken, H. J. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am*. 77, 1069–1077. doi: 10.1121/1.392224

Hutt, A., Mierau, A., and Lefebvre, J. (2016). Dynamic control of synchronous activity in networks of spiking neurons. *PLoS ONE* 11:e0161488. doi: 10.1371/journal.pone.0161488

Hygge, S., Boman, E., and Enmarker, I. (2003). The effects of road traffic noise and meaningful irrelevant speech on different memory systems. *Scand. J. Psychol*. 44, 13–21. doi: 10.1111/1467-9450.00316

Irrmischer, M., Poil, S.-S., Mansvelder, H. D., Intra, F. S., and Linkenkaer-Hansen, K. (2018). Strong long-range temporal correlations of beta/gamma oscillations are associated with poor sustained visual attention performance. *Eur. J. Neurosci*. 48, 2674–2683. doi: 10.1111/ejn.13672

Jenson, D., Thornton, D., Harkrider, A. W., and Saltuklaroglu, T. (2019). Influences of cognitive load on sensorimotor contributions to working memory: an EEG investigation of mu rhythm activity during speech discrimination. *Neurobiol. Learn. Mem*. 166:107098. doi: 10.1016/j.nlm.2019.107098

Jones, S. R., Pritchett, D. L., Sikora, M. A., Stufflebeam, S. M., Hämäläinen, M., and Moore, C. I. (2009). Quantitative analysis and biophysically realistic neural modeling of the MEG mu rhythm: rhythmogenesis and modulation of sensory-evoked responses. *J. Neurophysiol*. 102, 3554–3572. doi: 10.1152/jn.00535.2009

Kahneman, D. (1973). *Attention and Effort*, Vol. 1063. Englewood Cliffs, NJ: Prentice-Hall.

Kaiser, J., and Lutzenberger, W. (2005). Human gamma-band activity: a window to cognitive processing. *Neuroreport* 16, 207–211. doi: 10.1097/00001756-200502280-00001

Katsuki, F., and Constantinidis, C. (2014). Bottom-up and top-down attention: different processes and overlapping neural systems. *Neuroscientist* 20, 509–521. doi: 10.1177/1073858413514136

Kawabata, N. (1972). Nonstationary power spectrum analysis of the photic alpha blocking. *Kybernetik* 12, 40–44. doi: 10.1007/BF00289235

Kaya, E. M., and Elhilali, M. (2017). Modelling auditory attention. *Philos. Trans. R. Soc. B Biol. Sci*. 372:20160101. doi: 10.1098/rstb.2016.0101

Kayser, C., Wilson, C., Safaai, H., Sakata, S., and Panzeri, S. (2015). Rhythmic auditory cortex activity at multiple timescales shapes stimulus-response gain and background firing. *J. Neurosci*. 35, 7750–7762. doi: 10.1523/JNEUROSCI.0268-15.2015

Kelley, P., and Whatson, T. (2013). Making long-term memories in minutes: a spaced learning pattern from memory research in education. *Front. Hum. Neurosci*. 7:589. doi: 10.3389/fnhum.2013.00589

Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *J. Neurosci*. 30, 620–628. doi: 10.1523/JNEUROSCI.3631-09.2010

Klatte, M., Bergström, K., and Lachmann, T. (2013). Does noise affect learning? A short review on noise effects on cognitive performance in children. *Front. Psychol*. 4:578. doi: 10.3389/fpsyg.2013.00578

Klimesch, W. (1997). EEG-alpha rhythms and memory processes. *Int. J. Psychophysiol*. 26, 319–340. doi: 10.1016/S0167-8760(97)00773-3

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev*. 29, 169–195. doi: 10.1016/S0165-0173(98)00056-3

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci*. 16, 606–617. doi: 10.1016/j.tics.2012.10.007

Klimesch, W., Schimke, H., and Pfurtscheller, G. (1993). Alpha frequency, cognitive load and memory performance. *Brain Topogr*. 5, 241–251. doi: 10.1007/BF01128991

Lehmann, A., and Schönwiesner, M. (2014). Selective attention modulates human auditory brainstem responses: relative contributions of frequency and spatial cues. *PLoS ONE* 9:e85442. doi: 10.1371/journal.pone.0085442

Linkenkaer-Hansen, K., Nikouline, V. V., Palva, J. M., and Ilmoniemi, R. J. (2001). Long-range temporal correlations and scaling behavior in human brain oscillations. *J. Neurosci*. 21, 1370–1377. doi: 10.1523/JNEUROSCI.21-04-01370.2001

Linkenkaer-Hansen, K., Smit, D. J., Barkil, A., van Beijsterveldt, T. E., Brussaard, A. B., Boomsma, D. I., et al. (2007). Genetic contributions to long-range temporal correlations in ongoing oscillations. *J. Neurosci*. 27, 13882–13889. doi: 10.1523/JNEUROSCI.3083-07.2007

Ljung, R., Sörqvist, P., Kjellberg, A., and Green, A.-M. (2009). Poor listening conditions impair memory for intelligible lectures: implications for acoustic classroom standards. *Build. Acoust*. 16, 257–265. doi: 10.1260/135101009789877031

Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004

MacDonald, K. D., and Barth, D. S. (1995). High frequency (gamma-band) oscillating potentials in rat somatosensory and auditory cortex. *Brain Res.* 694, 1–12. doi: 10.1016/0006-8993(95)00662-A

Mahjoory, K., Cesnaite, E., Hohlefeld, F. U., Villringer, A., and Nikulin, V. V. (2019). Power and temporal dynamics of alpha oscillations at rest differentiate cognitive performance involving sustained and phasic cognitive control. *Neuroimage* 188, 135–144. doi: 10.1016/j.neuroimage.2018.12.001

Menzel, R., Manz, G., Menzel, R., and Greggers, U. (2001). Massed and spaced learning in honeybees: the role of cs, us, the intertrial interval, and the test interval. *Learn. Mem.* 8, 198–208. doi: 10.1101/lm.40001

Mierau, A., Klimesch, W., and Lefebvre, J. (2017). State-dependent alpha peak frequency shifts: experimental evidence, potential mechanisms and functional implications. *Neuroscience* 360, 146–154. doi: 10.1016/j.neuroscience.2017.07.037

Mills, C., Fridman, I., Soussou, W., Waghray, D., Olney, A. M., and D'Mello, S. K. (2017). "Put your thinking cap on: detecting cognitive load using EEG during learning," in *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (Vancouver, BC: ACM), 80–89. doi: 10.1145/3027385.3027431

Morillon, B., Liégeois-Chauvel, C., Arnal, L. H., Bénar, C. G., and Giraud, A.-L. (2012). Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study. *Front. Psychol.* 3:248. doi: 10.3389/fpsyg.2012.00248

Newman, M. E. (2005). Power laws, pareto distributions and zipf's law. *Contemp. Phys.* 46, 323–351. doi: 10.1080/00107510500052444

Nikulin, V. V., and Brismar, T. (2005). Long-range temporal correlations in electroencephalographic oscillations: relation to topography, frequency band, age and gender. *Neuroscience* 130, 549–558. doi: 10.1016/j.neuroscience.2004.10.007

Nikulin, V. V., and Brismar, T. (2006). Phase synchronization between alpha and beta oscillations in the human electroencephalogram. *Neuroscience* 137, 647–657. doi: 10.1016/j.neuroscience.2005.10.031

Obleser, J., Herrmann, B., and Henry, M. J. (2012). Neural oscillations in speech: don't be enslaved by the envelope. *Front. Hum. Neurosci.* 6:250. doi: 10.3389/fnhum.2012.00250

Obleser, J., and Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends Cogn. Sci.* 23, 913–926. doi: 10.1016/j.tics.2019.08.004

Oostenveld, R., and Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* 112, 713–719. doi: 10.1016/S1388-2457(00)00527-7

O'Sullivan, J., Herrero, J., Smith, E., Schevon, C., McKhann, G. M., Sheth, S. A., et al. (2019). Hierarchical encoding of attended auditory objects in multi-talker speech perception. *Neuron* 104, 1195–1209. doi: 10.1016/j.neuron.2019.09.007

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2014). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355

Palva, J. M., Zhigalov, A., Hirvonen, J., Korhonen, O., Linkenkaer-Hansen, K., and Palva, S. (2013). Neuronal long-range temporal correlations and avalanche dynamics are correlated with behavioral scaling laws. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3585–3590. doi: 10.1073/pnas.1216855110

Pashler, H. (1984). Processing stages in overlapping tasks: evidence for a central bottleneck. *J. Exp. Psychol. Hum. Percept. Perform.* 10:358. doi: 10.1037/0096-1523.10.3.358

Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., and Goldberger, A. L. (1994). Mosaic organization of dna nucleotides. *Phys. Rev. E* 49:1685. doi: 10.1103/PhysRevE.49.1685

Penolazzi, B., Angrilli, A., and Job, R. (2009). Gamma eeg activity induced by semantic violation during sentence reading. *Neurosci. Lett.* 465, 74–78. doi: 10.1016/j.neulet.2009.08.065

Pfurtscheller, G., and Da Silva, F. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/S1388-2457(99)00141-8

Pfurtscheller, G., Stancak, A. Jr., and Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band-an electrophysiological correlate of cortical idling: a review. *Int. J. Psychophysiol.* 24, 39–46. doi: 10.1016/S0167-8760(96)00066-9

Poil, S.-S., Hardstone, R., Mansvelder, H. D., and Linkenkaer-Hansen, K. (2012). Critical-state dynamics of avalanches and oscillations jointly emerge from balanced excitation/inhibition in neuronal networks. *J. Neurosci.* 32, 9817–9823. doi: 10.1523/JNEUROSCI.5990-11.2012

Pulvermüller, F., Birbaumer, N., Lutzenberger, W., and Mohr, B. (1997). High-frequency brain activity: its possible role in attention, perception and language processing. *Prog. Neurobiol.* 52, 427–445. doi: 10.1016/S.0301-0082(97)00023-3

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: RFoundation for Statistical Computing.

Riecke, L., Formisano, E., Sorger, B., Başkent, D., and Gaudrain, E. (2018). Neural entrainment to speech modulates speech intelligibility. *Curr. Biol.* 28, 161–169. doi: 10.1016/j.cub.2017.11.033

Ríos-López, P., Molnar, M. T., Lizarazu, M., and Lallier, M. (2017). The role of slow speech amplitude envelope for speech processing and reading development. *Front. Psychol.* 8:1497. doi: 10.3389/fpsyg.2017.01497

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 336, 367–373. doi: 10.1098/rstb.1992.0070

Rudner, M. (2016). Cognitive spare capacity as an index of listening effort. *Ear Hear.* 37, 69S–76S. doi: 10.1097/AUD.0000000000000302

Sauseng, P., Klimesch, W., Doppelmayr, M., Pecherstorfer, T., Freunberger, R., and Hanslmayr, S. (2005). Eeg alpha synchronization and functional coupling during top-down processing in a working memory task. *Hum. Brain Mapp.* 26, 148–155. doi: 10.1002/hbm.20150

Schack, B., Vath, N., Petsche, H., Geissler, H.-G., and Möller, E. (2002). Phase-coupling of theta-gamma EEG rhythms during short-term memory processing. *Int. J. Psychophysiol.* 44, 143–163. doi: 10.1016/S0167-8760(01)00199-4

Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychol. Rev.* 84:1. doi: 10.1037/0033-295X.84.1.1

Schroeder, C. E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18. doi: 10.1016/j.tins.2008.09.012

Sedley, W., Gander, P. E., Kumar, S., Kovach, C. K., Oya, H., Kawasaki, H., et al. (2016). Neural signatures of perceptual inference. *Elife* 5:e11476. doi: 10.7554/eLife.11476.017

Shukla, B., Rao, B. S., Saxena, U., and Verma, H. (2018). Measurement of speech in noise abilities in laboratory and real-world noise. *Indian J. Otol.* 24:109. doi: 10.4103/indianjotol.INDIANJOTOL_134_17

Siegel, M., Engel, A. K., and Donner, T. H. (2011). Cortical network dynamics of perceptual decision-making in the human brain. *Front. Hum. Neurosci.* 5:21. doi: 10.3389/fnhum.2011.00021

Snyder, A. C., and Foxe, J. J. (2010). Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: a high-density electrical mapping study. *J. Neurosci.* 30, 4024–4032. doi: 10.1523/JNEUROSCI.5684-09.2010

Spitzer, B., and Haegens, S. (2017). Beyond the status quo: a role for beta oscillations in endogenous content (RE) activation. *Eneuro* 4:ENEURO.0170-17.2017. doi: 10.1523/ENEURO.0170-17.2017

Strauß, A., Wöstmann, M., and Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Front. Hum. Neurosci.* 8:350. doi: 10.3389/fnhum.2014.00350

Szalma, J. L., and Hancock, P. A. (2011). Noise effects on human performance: a meta-analytic synthesis. *Psychol. Bull.* 137:682. doi: 10.1037/a0023987

Szeto, H. H. (1990). Spectral edge frequency as a simple quantitative measure of the maturation of electrocortical activity. *Pediatr. Res.* 27:289. doi: 10.1203/00006450-199003000-00018

Tetzlaff, C., Kolodziejski, C., Markelic, I., and Wörgötter, F. (2012). Time scales of memory, learning, and plasticity. *Biol. Cybernet.* 106, 715–726. doi: 10.1007/s00422-012-0529-z

Uusberg, A., Uibo, H., Kreegipuu, K., and Allik, J. (2013). EEG alpha and cortical inhibition in affective attention. *Int. J. Psychophysiol.* 89, 26–36. doi: 10.1016/j.ijpsycho.2013.04.020

Van Albada, S. J., and Robinson, P. A. (2013). Relationships between electroencephalographic spectral peaks across frequency bands. *Front. Hum. Neurosci*. 7:56. doi: 10.3389/fnhum.2013.00056

Vanthornhout, J., Decruy, L., and Francart, T. (2019). Effect of task and attention on neural tracking of speech. *Front. Neurosci*. 13:977. doi: 10.3389/fnins.2019.00977

Vural, C., and Yildiz, M. (2010). Determination of sleep stage separation ability of features extracted from EEG signals using principle component analysis. *J. Med. Syst*. 34, 83–89. doi: 10.1007/s10916-008-9218-9

Wascher, E., Rasch, B., Sänger, J., Hoffmann, S., Schneider, D., Rinkenauer, G., et al. (2014). Frontal theta activity reflects distinct aspects of mental fatigue. *Biol. Psychol*. 96, 57–65. doi: 10.1016/j.biopsycho.2013.11.010

Weiss, S., and Mueller, H. M. (2012). "Too many betas do not spoil the broth": the role of beta brain oscillations in language processing. *Front. Psychol*. 3:201. doi: 10.3389/fpsyg.2012.00201

Zhao, C., Zhao, M., Liu, J., and Zheng, C. (2012). Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Acc. Anal. Prev*. 45, 83–90. doi: 10.1016/j.aap.2011.11.019

Conflict of Interest: BD was employed by company ASAsense.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

MINI REVIEW
published: 11 November 2020
doi: 10.3389/fpsyg.2020.569056

Check for updates

Spatial Soundscapes and Virtual Worlds: Challenges and Opportunities

*Chinmay Rajguru, Marianna Obrist and Gianluca Memoli**

School of Engineering and Informatics, University of Sussex, Falmer, United Kingdom

There is increasing effort to characterize the soundscapes around us so that we can design more compelling and immersive experiences. This review paper focuses on the challenges and opportunities around sound perception, with a particular focus on spatial sound perception in a virtual reality (VR) cityscape. We review how research on temporal aspects has recently been extended to evaluating spatial factors when designing soundscapes. In particular, we discuss key findings on the human capability of localizing and distinguishing spatial sound cues for different technical setups. We highlight studies carried out in both real-world and virtual reality settings to evaluate spatial sound perception. We conclude this review by highlighting the opportunities offered by VR technology and the remaining open questions for virtual soundscape designers, especially with the advances in spatial sound stimulation.

Keywords: soundscape, virtual reality, sound perception, spatial audio, localization

OPEN ACCESS

Edited by:
Francesco Aletta, University College London, United Kingdom

Reviewed by:
Jooyoung Hong, Singapore University of Technology and Design, Singapore
Yue Wu, Harbin Institute of Technology, China

****Correspondence:***
Gianluca Memoli
g.memoli@sussex.ac.uk

Specialty section:
This article was submitted to Environmental Psychology, a section of the journal Frontiers in Psychology

Received: *02 June 2020*
Accepted: *16 April 2020*
Published: *11 November 2020*

Citation:
Rajguru C, Obrist M and Memoli G (2020) Spatial Soundscapes and Virtual Worlds: Challenges and Opportunities. Front. Psychol. 11:569056. doi: 10.3389/fpsyg.2020.569056

1. INTRODUCTION

The term "soundscape" was introduced in the 1970s, when Schafer (1976) considered the concept of "positive soundmarks." For many years, however, the word "soundscape" has been used to describe the recording and preservation of natural sounds. In projects like the UK Sound Map (The British Library, 2011), however, the concept has more recently evolved to include also the increasingly stimuli-rich acoustics of modern cities. According to this approach, our cities are not just reservoirs of unwanted sound (aka noise), but unexpectedly full of sounds that add to the immaterial heritage (Flesch et al., 2017).

In the soundscape approach, attention is shifted to the end users and communities within a modern city. The key difference from energy-focused descriptions (e.g., decibels) is that the different sounds present in a space are weighted by the listeners' perception.

The idea of managing soundscapes was introduced by the European Noise Directive (END) 2002/49/EC (European Commission, 2002). According to the END, unwanted sound, which has been a passively accepted aspect of Western societies since the Industrial Revolution, had now to be actively managed, even outside workplaces, to enhance citizens' well-being. The END also introduced a requirement to preserve quietness, a concept intended to have the widest possible meaning, and—at the time—left the definition to member states. After the END, planning the soundscape of future cities (i.e., the sources of sound present in a city) means not only reducing the intensity of sources labeled as noisy, but also considering positive sounds (Payne et al., 2009). While reducing the intensity of unwanted sounds is crucial near transport infrastructure (e.g., for houses facing a busy road where the impact on health may be severe), positive sounds dominate and may make a difference further from the road, where the focus is on self-reported well-being (Memoli and Licitra, 2012; Andringa et al., 2013; Aletta et al., 2019).

How to plan these soundscape changes has been addressed by the scientific community in two ways (Kang and Schulte-Fortkamp, 2016): (1) by evaluating in local communities physical indicators closely related to perception (Licitra et al., 2005; Memoli et al., 2008a; Kang et al., 2019) and (2) by standardizing acoustic surveys such that local residents are directly questioned to assess their perceptions of and expectations for the local acoustic climate; see Fields et al. (2001) and the ISO 12913 series (ISO/TC 43/SC 1 Noise, 2018). Armed with novel indicators, in the years following the END, different researchers—from the early days (Memoli et al., 2008b; Payne et al., 2009) to the most recent (Hong et al., 2020; Oberman et al., 2020) —have introduced a future where sounds can be added to an existing urban acoustic environment to change the perception of listeners.

This led to a number of studies where the soundscape of a place (and changes to it) is evaluated remotely in a laboratory through an immersive experience designed to recreate the acoustical feeling of "being there." Brambilla and Maffei (2010) pioneered this type of study, comparing their findings for two Italian cities (Rome and Naples) with laboratory experiences using 2D pictures and audio recordings. Similarly, Oberman et al. (2020) designed an auralization room for "virtual soundwalks" in three typologically unique cities (Graz, Zagreb, and Zadar), each with perceptible soundmarks (e.g., the Sea Organ in Zadar), using ambisonic audio (through loudspeakers) and 360° pictures on a screen.

This process, which in this work we will call "remote soundscape assessment," is also commonly used by global architecture firms. Arup's SoundLab (Forsyth, 2018), for instance, is an anechoic room where ambisonic audio is delivered through 12 speakers surrounding the listener. This setup allows Arup to evaluate the effects of noise action plans and has recently been used to evaluate Heathrow's updated respite procedure.

However, all these studies, which are often at the frontier between sonic art and immersive experiences, describe interventions on either the temporal aspect of an existing soundscape (e.g., whether we can add to a place a sequence of sounds that will be perceived as pleasant) or its frequency content (e.g., whether we can add sounds in a specific frequency range to alter perception). Moreover, they highlight the limits of just considering the temporal aspect, since multiple "non-acoustical" parameters can affect the judgments of a soundscape. These include the expectations of the listener (Miller, 2013; Sung et al., 2016; Aletta et al., 2018), and also the specific location, the local urban design and its visual appearance, the type of activities that happen there, and the listener's age, culture, and personal history (Kang and Schulte-Fortkamp, 2016).

In this mini-review, we address the often neglected impact on perception of where the sounds (appear to) come from in a remote assessment of soundscapes. Very little is, in fact, known about building and characterizing soundscapes from a spatial point of view. Even when spatialization is part of the soundscape design process, as described in the review by Hong et al. (2017), the typical conclusion is: take ambisonic recordings and deliver them through headphones.

Here, we explore alternative delivery methods. In section 2, we review works where sounds with a spatial connotation have been added to a visual stimulus to increase immersivity [1]. We mention cases using virtual reality (VR), augmented reality (AR), or mixed reality. These are different stages of a reality–virtuality continuum (Milgram and Takemura, 1994) of visual-based experiences produced by interactive displays, typically head-mounted. Experiences where sound is delivered either by loudspeakers or by headphones.

In section 3, we describe a selection of the increasing number of studies that use VR to evaluate potential changes to existing soundscapes. section 5 summarizes research on one critical aspect of soundscape design for VR: finding the optimal number of sources needed to maximize immersivity. In particular, we discuss studies that identify: (1) how close two sources can be for the listener to distinguish them, (2) the minimal number of sources to achieve a desired localization accuracy, and (3) the relation between localization accuracy and perceived immersivity. Section 6 summarizes our findings and highlights the unanswered questions.

Note that our analysis is limited, as it neglects the theory behind sound perception. The reader can find more about this subject and the success of 3D audio elsewhere (e.g., Begault, 2000; Hong et al., 2017; Roginska and Geluso, 2018).

2. VIRTUAL REALITY: A TOOL FOR EVALUATING SOUND PERCEPTION

Sound within human interfaces has not been developed anywhere near the level that visual interfaces have. Light-based special effects are common and light-based holograms are so well-known that they can be used as cheap souvenirs. However, as any theater or cinema director knows, achieving this level of control with sound is expensive. It requires either a large number of speakers or for everyone in the audience to wear headphones.

In VR, however, this level of control is more readily available. Moreover, audio is crucial for situational awareness. As summarized by Yung and Khoo-Lattimore (2019), the success of an immersive virtual environment (IVE) is based on three key elements: (1) the ability to look around (visualization), (2) a suspension of belief and physical representation of objects (immersion), and (3) a degree of control over the experience (interactivity). As reported by Hruby (2019) in a review on virtual cartography, different experiments support the positive impact of sound on spatial presence in IVEs. That is, sound is crucial for a user's feeling of being there. As an example, Kern and Ellermeier (2020) conducted a study where participants ($N = 36$) were asked to wear noise-canceling headphones and tasked to take a stroll in a VR park while walking on a treadmill in the real world. Different sounds were fed through the headphones. The conditions *steps with background* and *background sounds* scored more than 60% on the scales *presence* and *realism*, whereas *no-headphones*, *noise-canceling*, and *steps only* were around 40%.

Another example is virtual tourism. Since travelers are already happy to escape into alternate realities (e.g., theme parks), it is not surprising that a multitude of tourism-focused VR

[1] Often defined as the feeling of being there.

utilities are emerging. Games, educational tools, destination marketing, and virtual visits to cultural heritage sites aim to deliver selected visual, audio, and most importantly, spatial aspects of the destination without actually being there (Yung and Khoo-Lattimore, 2019). Virtual environments like *Second Life* (Linden Lab, 2020) have also gained momentum, especially when real travel is not possible (e.g., for someone confined at home to prevent the spread of disease). In these cases, adding audio to a 360° visualization may make the experience of a virtual visit almost indistinguishable from the real one (Wagler and Hanus, 2018).

The above considerations show that VR is the perfect tool for perception-focused acoustic experiments (**Figure 1A**). Even when cross-modal interactions have been detected (Malpica et al., 2020), there seem to be a prevalence of audio over visual cues in VR. This comes from one key advantage: not only can listeners experience virtual sound sources from anywhere in a 360° space, but these can be changed as required, whereas vision requires eye or hand movements (Madole and Begault, 1995). Like what happens to blindfolded people subject to acoustic cues (Tabry et al., 2013), the success of audio in VR is, however, underpinned by the ability to localize and position where the sounds come from in the 3D (virtual) environment. It is, therefore, paramount to deliver acoustic cues so that their location is perceived accurately (section 5).

3. SOUNDSCAPE EVALUATION IN VR

The low cost of VR and AR headsets and even the possibility of transforming a mobile phone into a VR visor has allowed more ambitious projects, as anticipated by Miller (2013). Lugten et al. (2018), for instance, explored the effect of adding sounds from moving water (e.g., fountains or ponds) or sounds from vegetation (e.g., sounds made by birds or the wind) in areas exposed to aircraft sounds. In their study, participants from the local community ($N = 41$) were exposed to eight different VR scenes and answered a questionnaire after each test. Lugten et al. showed that there was a marginal effect from adding only sound from vegetation, but that there was a massive positive effect for both the pleasantness and eventfulness ratings of the soundscape due to adding sound from water features or adding a combination of visual and audio aspects of water features. In particular, the eventfulness of the soundscapes increased by 26% for a sound pressure level of 70 dB and by 20% for 60 dB. Arntzen et al. (2011) went even further, with a VR-based method to asses the impact of aircraft sounds on community well-being.

Stevens et al. (2018) used VR to show how visual stimuli can alter the perception and categorization of a soundscape. In their experiments, they first presented participants ($N = 31$) with recordings taken in the North of England (in rural, urban, and suburban environments), each comprising a mixture of natural, human, and mechanical sounds. Then they presented the sounds with a visual accompaniment to represent: (1) a forest (natural), (2) a rural or suburban setting, and (3) a city center (urban). These were delivered as 360° videos through a head-mounted display (HMD). On comparing the subjective evaluations (valence, arousal, and dominance), they found a significant difference in the emotion and category ratings with and without the visuals. Again, the registered impact of vegetation was very similar to what has been found with standard surveys (Watts et al., 2011). Conversely, no significant correlation between visual and sound cues was found by Echevarria Sanchez et al. (2017). These authors used binaural sound and 360° images, delivered through a VR headset, to evaluate the renovation designs for a bridge connecting the inner city of Ghent to a large park ($N = 75$).

4. SOUNDSCAPE IMMERSIVITY AND DELIVERY

As shown by Jiang et al. (2018), reproducing both the visual and the audio experience is not trivial. These authors created a virtual reproduction of a Neapolitan square (Piazza Vittoria) using captured 360° images (delivered through a VR headset) and sound recordings (e.g., human voices, bird sounds, fountain sounds, sea waves, and background urban sounds) attached to corresponding objects in the VR environment. They asked participants ($N = 100$) to evaluate the IVE on 7-point scales (ranging from *poor and unrealistic* to *good and realistic*) and to leave comments about it. In this study, 62% of participants voted *good and realistic* for the visuals, but only 51% did so for the audio.

Jiang et al. (2018) concluded that, to improve the sound realism, more sound sources are needed in a VR simulation. Having more sources, however, requires more rendering power: what is gained in immersivity through the audio may be lost in the visuals. Part of the delivery could, therefore, be delegated to external real sources located around the user (e.g., loudspeakers, as in Forsyth, 2018). According to Hong et al. (2019), this may be very effective. These authors asked participants ($N = 30$) to evaluate soundscape quality and perceived spatial qualities for three spatial sound reproduction methods (static binaural, tracked binaural, and 2D octagonal speaker) and found no perceived difference. Conversely, Hruby (2019) found that loudspeakers may reduce immersivity, because headphones allow the complete exclusion of background sounds.

Choosing how the soundscape is delivered is, therefore, difficult, as there is a gap between what can be programmed (e.g., with commercial packages like Unity or Unreal) or recorded (e.g., ambisonics) and what is actually perceived by the user through arrays of speakers or headphones.

5. SPATIAL SOUND: FROM DELIVERY TO PERCEPTION

We are familiar with surround sound, a technology where an array of loudspeakers around the listener is used to deliver sound from 360°. Most of these setups are basically an extension of the stereo concept, where different mono audio channels are sent to each speaker and integrated in the brain of the listener. Surround sound can be found in cinemas, home theaters, and many of the soundscape studies cited earlier. In this approach, which we call fixed-speaker sources in the following (section 5.1), more loudspeakers typically lead to a more precise delivery. We also consider that Ambisonic methods are in this group. In this

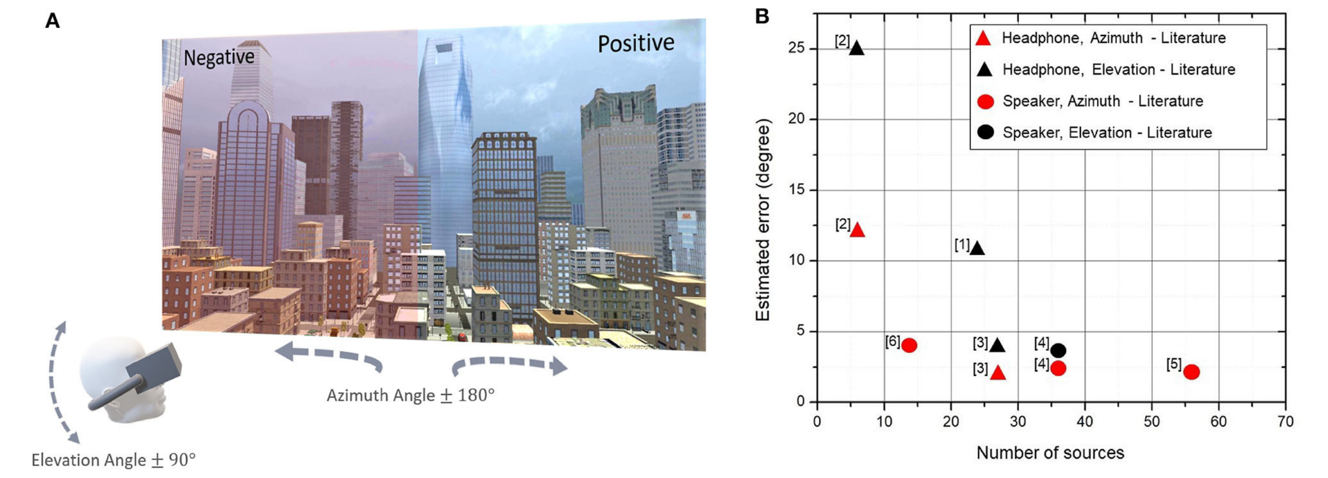


FIGURE 1 | (A) An example of how VR could be used to evaluate changes in a specific soundscape (e.g., due to works), highlighting the role of elevation and azimuth in source localization. The 3D cityscape in this figure was created by the authors for Oculus Quest using a free city asset for Unity (Dactilardesign, 2019). **(B)** Graphical representation of estimated error found for azimuth and elevation angles from a selection of the studies in **Table 1**, with numbers among parenthesis referring to entries in the Table.

recording technique, first-order spherical harmonics are used to interpret the sound reaching the recording microphone so that it can be delivered into four or eight channels.

More recent methods (e.g., wave-field synthesis) try, instead, to recreate a 3D sound field physically. With 3D sound systems, it is possible to create (using interference) the field corresponding to a source positioned between two physical speakers (i.e., virtual-speaker sources in section 5.2). In modern sound bars and linear arrays, the same signal is used to feed all the loudspeakers simultaneously, with differences in amplitude and phase. These emissions combine in real space and the wavefront that reaches the listener has the right 3D information (Begault, 2000; Roginska and Geluso, 2018). However, the relation between precise delivery and the number of speakers is not straightforward.

HMDs, however, come with headphones, and this is the preferred method for delivering audio in VR and AR (see also Hong et al., 2017). In this case, the sound sources are objects placed within the simulation by a programmer. Signal processing is used to calculate what needs to be delivered to each ear of the listener, after weighting for a standardized geometry of their head (see below for head-related transfer functions). This technology, like the beamforming used in radar and medical ultrasound scanners, is typically called spatial sound and gives the listener the impression that the sound comes from all around them (Begault, 2000). Although headphone sources are all virtual, in this mini-review we distinguish two sub-categories: fixed-headphone sources where the audio object stays in a fixed position relative to the visual environment and dynamic headphone sources where the audio is linked to a movable item.

5.1. Delivery Using Arrays (Fixed-Speaker Sources)

In a typical experiment with speaker arrays, the loudspeakers are positioned at a fixed distance D from the listener, either along the azimuth or the elevation direction, and the participants experience different acoustic stimuli, apparently coming from positions on a sphere of diameter D (**Table 1**). According to Guastavino et al. (2005), however, how the sounds are recorded is crucial, as this is the first step in a process later formalized by Hong et al. (2017). Guastavino et al. (2005) captured city sounds, using either stereo or ambisonic recordings, and delivered the acoustic experience using eight fixed-speaker sources. The participants ($N = 29$ for stereo and $N = 27$ for ambisonics) were asked which setup sounded more like an everyday experience. The results showed that 2D configurations of speakers are sufficient outdoors, whereas 3D configurations should be preferred indoors.

The first parameter for characterizing these systems is the minimum audible angle (MAA), which is the smallest difference in the azimuth direction of two equal sound sources that can be reliably separated (Mills, 1958). This is assumed to be about 2° in front of the listener, which was confirmed by the experiments of Kühnle et al. (2013), who found a median value of $2.5 \pm 1.1°$ for sources at $\varphi \approx 0°$ (14 speakers, $N = 136$ participants). The MAA, however, is thought to degrade as the angle increases, and in fact, a median of $5.3 \pm 2.5°$ was found for sources at $\varphi \approx 90°$ (Kühnle et al., 2013). The MAA is a psychoacoustic quantity and does not seem to depend on visual stimuli (Rummukainen et al., 2020).

The second parameter is the localization accuracy, which is the maximum difference between the programmed position of the sound source and its perceived position. Although this quantity cannot be larger than the MAA, accuracy depends on the number of speakers and their positions in space. Just like changes in the output of an optical display need to be delivered quicker than the eye can perceive to produce fluid images (i.e., 0.1 s), the MAA gives acoustic designers a target for spatial accuracy.

Makous and Middlebrooks (1990), for instance, used 36 real speakers spaced at 10° intervals around a circular hoop (1.2 m radius). The participants ($N = 6$) were asked to turn their head

TABLE 1 | A summary of the literature considered in this mini-review on the capability of locating sound sources.

Reference	Environment	Sound Stimuli	Sound Stimuli distance	Sound delivery method	No. and type of sources	No. of participants	Error found
[1] Sodnik et al. (2006)	AR	Engine sound	min. 15 cm max. 80 cm	Headphones	24 fixed sources	10	Distance between perceived and real source <15 cm (i.e., 10.8°)
Rungta et al. (2017)	VR	Recorded human clapping	1.7 m	Headphones	7 fixed sources	17	Users overestimated distances <1 m and underestimated distances >1 m
Kose et al. (2018)	VR	Audio clip from Modern music	1.5 m	Headphones	1 dynamic source	n/a	Elevation was misjudged
[2] Yang et al. (2019)	VR	Synthesized sounds	Arbitrary distribution	Headphones	6 fixed sources	21	Mean azimuth error: 12.07° Mean elevation error: 25.06°
[3] Ahrens et al. (2019)	VR and Real	Pink noise burst	2.4 m 15° separation	27 loudspeakers	27 fixed sources: 13 along azimuth 7 each for elevation ± 28°	10	Elevation error (max. 2°) was larger when using a head mounted display (HMD)
[4] Makous and Middlebrooks (1990)	Real	System generated signals	1.2 m	36 loudspeakers	36 fixed sources	6	Azimuth error 2° Elevation error 3.5°
[5] Müller et al. (2014)	Real	Pulsed noise, speech, guitar tones	3 m	56 real loudspeakers	Multiple virtual sources	17	Azimuth error <11.5 cm (i.e., 2.2°)
Sato et al. (2020)	Real	Low-frequency noise (100 Hz 500 Hz)	1.5 m	4 real loudspeakers	Multiple virtual sources	7	Performance of judging elevation reduced after 65°
[6] Kühnle et al. (2013)	Real	Gaussian noise bursts (250 ms)	2.35 m	14 loudspeakers	14 fixed sources	136	2 ± 1° near the front 4 ± 2° at 90° from front
Litovsky et al. (2004)	Real	Pink noise bursts (170 ms, 65 dB)	1.4 m	8 loudspeakers	8 fixed sources	17	Root mean square error for bilateral signals approx. 30°

In the typical experiment, the user is in a virtual or real environment and receives sound cue(s) from a specific distance. The stimuli are produced by a certain number of sources, located all around the user, both in the azimuth and the elevation directions. The study realizes a specific angular error in locating the sound source. References with number among parenthesis are represented graphically in the ***Figure 1B****.*

to look in the direction of the sound while being tracked using an electromagnetic device. It was found that the performance of listeners was better in front than any other direction, with an average sound localization error of 2° in azimuth and of 3.5° in elevation.

Ahrens et al. (2019) used more speakers. They built a full sphere with 64 loudspeakers (2.4 m from the listening position) and used it to highlight the challenge of aligning the real world (i.e., the fixed loudspeakers) and the virtual world (i.e., the VR simulation) to achieve accurate spatial auditory perception. These authors used only the frontal 27 speakers and conducted tests for different user conditions, such as with or without a blindfold as well as with a HMD. With visible sources, the accuracy was very close to the MAA. However, the azimuth and elevation localization errors increased by 3° and 1.5° when the subjects were blindfolded. When participants were wearing a HMD, the azimuth error was the same but the elevation error was 2° larger. Interestingly, users performed better on the right-hand side by ≈1°.

5.2. Delivery Through Virtual-Speaker Sources

Müller et al. (2014) used a "BoomRoom"—a room containing a ring of 56 real loudspeakers (diameter 3 m, positioned at the ear level of the user) and 16 suspended cameras—to track a user's position and deliver virtual-speaker sources. Müller et al. (2014) created an AR experience using wave-field synthesis to create virtual sound sources originating from real objects (bottles or bowls). The participants ($N = 17$) were asked to determine the sound source location, which was accurate within 2.2° (azimuth), very close to the MAA.

More recently, Sato et al. (2020) investigated the relation between the perception of azimuth and elevation angles using only four speakers. These authors tested 40 different configurations with two acoustic signals (wideband noise: 100 Hz–20 kHz or low-pass noise: 100–500 Hz), four elevation angles (55°, 65°, 75°, and 80°), and five initial azimuth angles (0°, 45°, 90°, 180°, and −135°). The participants ($N = 7$) were asked to find the direction of the sound source (sound pressure level of 65 dB and duration of 1,600 ms) by pressing a button. The procedure was repeated 160 times. Their algorithms worked effectively when the height of the sound source was lower than 3 m and horizontally farther than 1 m. Also, 65° was the upper limit of the elevation angle.

5.3. Delivery Through Headphone Sources

Spatial audio uses the time and intensity differences between the signals to each ear, which underpin our ability to position a source in the horizontal plane (azimuthal angle $\varphi = 0°$ to

180°, where 0° corresponds to the front of the listener) and at a certain distance (Rayleigh, 1907; Bronkhorst and Houtgast, 1999). Once the position of listener relative to the sources is known, spatial audio is relatively easy to implement. In contrast, the ability to locate a source in the vertical plane (elevation angle, $\theta = 0°$ to $90°$) depends on the direction-dependent filtering of the outer ear (Roffler and Butler, 1968). There may be larger localization uncertainties for elevation, as perception depends on the individual.

In this approach, there are a few positions where the sound localization is difficult. Two obvious positions are the points just in front or just behind a user, since the signals reach each ear at the same time. To overcome this "cone of confusion" (Aggius-Vella et al., 2017), listeners simply need to move their heads, so that the angle changes until a perceivable difference is created. When the head is fixed, however, listeners hearing sounds through headphones that are processed to appear as if they come from behind them can experience localization inaccuracies as large as 45° (Steadman et al., 2019). The temporal dynamic of the source, however, is only one aspect of perception. As shown in **Table 1**, assumptions about the listener, the number of virtual objects producing the sound, and the amount of training received are also crucial.

5.3.1. Assumptions About the User

The frequency filtering due to diffraction from the pinna, head, and torso is usually described by a head-related transfer function (HRTF) (Begault, 2000). Virtual audio systems are based on the assumption that, once the HRTF is known, any sound can be processed so that, when delivered through headphones or an array of speakers, it is perceived as coming from any desired position in 3D space (Wightman and Kistler, 1989). HRTFs are parameters of the individual, but because commercial solutions use standard functions measured with dummies (Wenzel et al., 1993), there is a potential reduction in localization accuracy (Ben-Hur et al., 2020). In addition, since HRTFs are not always recorded in non-anechoic environments, any differences between the space where the sound was recorded and the one where it is played may result in further inaccuracies, especially in terms of the perceived distance of the source, which may be as low as ≈18% of the correct distance, according to Gil-Carvajal et al. (2016).

5.3.2. Source Distance and Movement

It is very challenging to estimate the distance of a source, such as the altitude of an overflying plane, and this results in large errors in real life (Memoli et al., 2018). Localization improves when a user is allowed to move towards the source, which is crucial for audio-only AR environments (e.g., a voice describing places).

Rungta et al. (2017) compared experimentally the performance of analytical algorithms (based on parametric filters) with ray tracing in rendering the distance of an acoustic source. Participants ($N = 17$) were trained blindfolded and then tasked to judge the distance to different sources, while walking along a path in VR. This study found that, although the actual distance is directly proportional to the perceived one, subjects tended to overestimate distances <1 m and underestimate distances >1 m.

Yang et al. (2019) placed everyday objects around users as spatial audio (virtual) sources in AR. These authors describe an experiment ($N = 21$) consisting of three parts. First, they delivered 3D sound through headphones, while the user was standing stationary, facing the direction of the sound. Second, they delivered sound with a visual cue (real paper boxes). For these two parts, the user was asked to find the source location while remaining in the same spot. Third, the user was allowed to walk towards the virtual sound source to identify its location. For the first part, they found a maximum azimuth localization error of 30° and an average error of 12.07°. In the second case, only 4 tests were answered incorrectly out of 168. All the users were able to find the objects accurately in the third case. Interestingly, Yang et al. (2019) noted that participants first walked in the direction of the sound to reduce the angle or distance error.

5.3.3. Training

When required to localize an acoustic source, participants benefit from training. Sodnik et al. (2006), for instance, found a net improvement in performance after three attempts. Steadman et al. (2019) showed that using a game-like environment to train users can improve their ability to localize sounds. The participants ($N = 36$) listened to 19 acoustically complex stimuli positioned along a hemisphere centered on the listener. Stimuli were delivered using headphones. The tests were conducted over 3 days, during which each participant was randomly assigned to one of four groups. The control group ($N = 9$) did not receive any training while the other three groups were trained using increasing gamification elements. The perception errors for azimuth and elevation (i.e., the angular difference between where the sound was actually delivered and where it was perceived) were highly skewed, but the majority of errors for the control group were below <90°, with an average of ≈40°. All participants undergoing training had lower localization errors than the control group. The error decreased with the number of training sessions. When the participants were allowed to turn their heads, the average error was reduced to 20% after only six sessions.

5.3.4. Presence of Visual Stimuli

There is often a correlation between the acoustic and visual judgments of a soundscape (Watts et al., 2011; Memoli et al., 2018; Ahrens et al., 2019). Since in VR the balance may be altered, e.g., by presenting visual and acoustic stimuli at different times, it is, therefore, important to quantify their relative weights.

Sodnik et al. (2006) highlighted the role of cross-modal correspondences using AR by delivering 3D sound through headphones. Their virtual sources were coincident with 24 identical aircraft models, randomly placed in a tabletop-sized environment (100 × 60 × 60 cm), at distances between 15 cm and 80 cm from the listener. Participants ($N = 10$) were asked to indicate a noisy object by turning their head in its direction. These authors found that the minimum distance between the head and the target was 15 cm (corresponding to a 10.8° localization error) and observed that participants performed sound localization first in the horizontal plane and then in vertical plane. They also noted that localization can be improved if there is some regularity in the distribution of the sound sources.

Kose et al. (2018) ran a similar experiment, simultaneously presenting acoustic and visual stimuli in a VR environment. They used a single speaker moving along a rectangular path. In the first part of the experiment, they localized the source using microphones and a triangulation algorithm. In the second part, they used the captured sounds to deliver 3D sound to participants in a VR environment, where the source appeared like a group of spheres. They found that the microphones often misjudged the up–down direction. Also, the users often got distracted by the delay of 500 ms due to processing.

6. DISCUSSION AND CONCLUSIONS

In comparing judgments on soundscapes obtained in situ with those obtained remotely [i.e., in a dedicated room with sound and visual stimuli (Brambilla and Maffei, 2010; Oberman et al., 2020)], we highlighted that only the right number of sources can recreate the auditive feeling of being there (Jiang et al., 2018), which corresponds to hearing sound all around.

As reported by Guastavino and Katz (2004), however, there is not an optimal reproduction method that works for arbitrary audio material. The choice is left to the designer, who often has only qualitative information on the differences between the delivery methods (e.g., Hong et al., 2017). Indeed, even design indications for loudness are inadequate. The few researchers who have looked into this parameter (e.g., by measuring the sound level threshold) seem to agree that the threshold does not depend on loudness (Makous and Middlebrooks, 1990; Litovsky et al., 2004; Rungta et al., 2017).

In this mini-review, we focused on one key design parameter: identifying the minimum number of sources needed for a given localization accuracy. We analyzed the literature on the spatial perception of sound and in particular the studies that use either fixed-speaker sources or fixed-headphone sources (i.e., acoustic objects fixed in a virtual landscape and delivered through headphones). In these studies—a subset of those reported in **Table 1**—one single speaker was active at any moment in time but, as shown in **Figure 1B**, we found that the localization error decreased with the number of sources. An MAA of 2° was reached in studies with at least 27 sources [13 in azimuth and 7 in ±28° elevation angles (Ahrens et al., 2019)]. Note that **Figure 1B** plots the values for fixed-headphone sources and fixed-speaker sources with the same color, as they seem to follow the same trend. This suggests that to achieve a seamless spatial delivery, at least 27 sound objects are needed in a VR simulation, making the latter difficult for a portable headset. This finding needs further investigation.

Table 1 also highlights how the sound sources positioned along azimuth angles are easier to localize than sources at different elevations. This is crucial for 3D soundscape designers (Hruby, 2019) and for interpreting data in a cityscape (Memoli et al., 2018).

Once the right level of immersivity is reached, however, the soundscape evaluations obtained in VR-based experiences seem to be similar to those obtained—in the same location—from standard questionnaire surveys (Hong et al., 2017; Wagler and Hanus, 2018; Oberman et al., 2020). This observation, if confirmed by more studies, indicates that VR could be used to evaluate the impact of changes to a soundscape in urban areas, with highly reduced costs compared to field studies. Recent studies are already starting to transfer the instruments typically used for assessing real soundscapes into VR environments (Lam et al., 2020).

Before getting there, however, it is necessary to confirm whether audio in VR has the same impact as in the real world, as some of the studies we reviewed emphasize that the role of visual stimuli is at the same level observed in standard questionnaire surveys (Stevens et al., 2018; Wagler and Hanus, 2018). In contrast, other researchers report that audio is important in VR (Malpica et al., 2020) and others report no correlation between the two stimuli, at least in certain tasks (Echevarria Sanchez et al., 2017; Rummukainen et al., 2020). Equally important is to understand the positive role of training in VR, which has not been discussed in real soundscape evaluations, to date.

6.1. Recent Work and Future Opportunities

We are living a spring of creativity, as artists create acoustic live performances (Melody, 2020), shared spaces to meet, as well as live musical demos using combinations of virtual and real stimuli (Sra et al., 2018). One example of these mixed reality events is *Out There* (WilkinsAvenue, 2019), the first location-based immersive musical experience with a lyrical narrative. Another is the *Star Wars: Bose AR Experience* (Bose, 2019), in which users had to walk around in a physical space to explore the virtual sound and the events happening in the story. According to some authors, we will soon be able to purchase personalized soundscapes (Kiss et al., 2020). With 3D audio coming into the real world (Neoran et al., 2020), the boundaries between real and virtual are blurring.

As described in this mini-review, however, there are (potentially) large uncertainties associated with using headphones, underpinning the general consensus that external loudspeakers may create more immersive experiences, even in VR (Hruby, 2019). In addition, headphones do not allow users to comment on a shared experience in real space, reducing interactions. Further studies on alternative ways of delivering location-based sound cues—e.g., directional speakers (Yoneyama et al., 1983; Obrist et al., 2017; Ochiai et al., 2017), immersive audio domes (Ott et al., 2019), and, more recently, acoustic projectors (Rajguru et al., 2019)—would therefore be highly desirable. These methods create real sources around the listener and may, therefore, complement headphones in the virtual delivery of soundscape experiences.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors contributed equally to this manuscript. GM as an acoustician working on soundscapes, CR as an expert on virtual and augmented realities. MO as an expert in multi-sensory immersive experiences.

FUNDING

The authors acknowledge funding from UK Research and Innovation through Grant No. EP/S001832/1. MO contribution to this review is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement No. 638605.

REFERENCES

Aggius-Vella, E., Campus, C., Finocchietti, S., and Gori, M. (2017). Audio spatial representation around the body. *Front. Psychol*. 8:1932. doi: 10.3389/fpsyg.2017.01932

Ahrens, A., Lund, K. D., Marschall, M., and Dau, T. (2019). Sound source localization with varying amount of visual information in virtual reality. *PLoS ONE* 14:e0214603. doi: 10.1371/journal.pone.0214603

Aletta, F., Oberman, T., Mitchell, A., Erfanian, M., Lionello, M., Kachlicka, M., et al. (2019). Associations between soundscape experience and self-reported wellbeing in open public urban spaces: a field study. *Lancet* 394:S17. doi: 10.1016/S0140-6736(19)32814-4

Aletta, F., Van Renterghem, T., and Botteldooren, D. (2018). Influence of personal factors on sound perception and overall experience in urban green areas. a case study of a cycling path highly exposed to road traffic noise. *Int. J. Environ. Res. Public Health* 15:1118. doi: 10.3390/ijerph15061118

Andringa, T. C., Weber, M., Payne, S. R., Krijnders, J. D. D., Dixon, M. N., Linden, R., et al. (2013). Positioning soundscape research and management. *J. Acoust. Soc. Am*. 134, 2739–2747. doi: 10.1121/1.4819248

Arntzen, M., van Veen, T. A., Visser, H. G., and Simons, D. G. (2011). "Aircraft noise simulation for a virtual reality environment," in *17th AIAA/CEAS Aeroacoustics Conference 2011 (32nd AIAA Aeroacoustics Conference)*, Portland. doi: 10.2514/6.2011-2853

Begault, D. R. (2000). *3-D Sound for Virtual Reality and Multimedia*. Moffett Field, CA: National Aeronautics and Space Administration NASA.

Ben-Hur, Z., Alon, D., Robinson, P. W., and Mehra, R. (2020). "Localization of virtual sounds in dynamic listening using sparse HRTFs," in *Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality*.

Bose (2019). *Star Wars App: Enhanced With Bose AR*. Massachusetts. Available online at: https://www.bose.com

Brambilla, G., and Maffei, L. (2010). Perspective of the soundscape approach as a tool for urban space design. *Noise Control Eng. J*. 58, 532–539. doi: 10.3397/1.3484180

Bronkhorst, A. W., and Houtgast, T. (1999). Auditory distance perception in rooms. *Nature* 397, 517–520. doi: 10.1038/17374

Dactilardesign (2019). *Detailed City Pack*. Unity Asset Store. (accessed December 2019).

Echevarria Sanchez, G. M., Van Renterghem, T., Sun, K., De Coensel, B., and Botteldooren, D. (2017). Using virtual reality for assessing the role of noise in the audio-visual design of an urban public space. *Landsc. Urban Plan*. 167, 98–107. doi: 10.1016/j.landurbplan.2017.05.018

European Commission (2002). *Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 Relating to the Assessment and Management of Environmental Noise*. Technical report.

Fields, J., Jong, R. D., Gjestland, T., Flindell, I., Job, R., Kurra, S., et al. (2001). Standardized general-purpose noise reaction questions for community noise surveys: research and a recommendation. *J. Sound Vibr*. 242, 641–679. doi: 10.1006/jsvi.2000.3384

Flesch, S., Gutsche, A.-S., and Paschen, D. (2017). "Exploring the untapped potential of sound maps," in *Proceedings of Visualizing Cultural Collections 2017*, Potsdam.

Forsyth, C. (2018). "A method for virtual acoustic Auralisation in VR," in *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion, AM'18* (New York, NY: Association for Computing Machinery). doi: 10.1145/3243274.3243304

Gil-Carvajal, J. C., Cubick, J., Santurette, S., and Dau, T. (2016). Spatial hearing with incongruent visual or auditory room cues. *Sci. Rep*. 6:37342. doi: 10.1038/srep37342

Guastavino, C., Katz, B. F., Polack, J. D., Levitin, D. J., and Dubois, D. (2005). "Ecological validity of soundscape reproduction," in *Acta Acustica United With Acustica*, (Stuttgart: S. Hirzel Verlag).

Guastavino, C., and Katz, B. F. G. (2004). Perceptual evaluation of multi-dimensional spatial audio reproduction. *J. Acoust. Soc. Am*. 116, 1105–1115. doi: 10.1121/1.1763973

Hong, J. Y., He, J., Lam, B., Gupta, R., and Gan, W. S. (2017). Spatial audio for soundscape design: record. reproduction. *Appl. Sci*. 7:627. doi: 10.3390/app7060627

Hong, J. Y., Lam, B., Ong, Z. T., Ooi, K., Gan, W. S., Kang, J., et al. (2019). Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications. *Build. Environ*. 149, 1–14. doi: 10.1016/j.buildenv.2018.12.004

Hong, J. Y., Ong, Z.-T., Lam, B., Ooi, K., Gan, W.-S., Kang, J., et al. (2020). Effects of adding natural sounds to urban noises on the perceived loudness of noise and soundscape quality. *Sci. Tot. Environ*. 711:134571. doi: 10.1016/j.scitotenv.2019.134571

Hruby, F. (2019). The sound of being there: Audiovisual cartography with immersive virtual environments. *J. Cartogr. Geogr. Inform*. 69, 19–28. doi: 10.1007/s42489-019-00003-5

ISO/TC 43/SC 1 Noise (2018). *ISO/TS 12913-2:2018 Acoustics Soundscape Part 2: Data Collection and Reporting Requirements*. Technical report.

Jiang, L., Masullo, M., Maffei, L., Meng, F., and Vorländer, M. (2018). A demonstrator tool of web-based virtual reality for participatory evaluation of urban sound environment. *Landsc. Urban Plan*. 170, 276–282. doi: 10.1016/j.landurbplan.2017.09.007

Kang, J., Aletta, F., Oberman, T., Erfanian, M., Kachlicka, M., Lionello, M., et al. (2019). "Towards soundscape indices," in *Proceedings of the 23rd International Congress in Acoustics (ICA)*, AAchen.

Kang, J., and Schulte-Fortkamp, B. (eds.). (2016). *Soundscape and the Built Environment*. Boca Raton, FL: CRC Press.

Kern, A. C., and Ellermeier, W. (2020). Audio in VR: effects of a soundscape and movement-triggered step sounds on presence. *Front. Robot. AI*. 7:20. doi: 10.3389/frobt.2020.00020

Kiss, F., Mayer, S., and Schwind, V. (2020). Audio VR: did video kill the radio star? *Interactions* 27, 46–51. doi: 10.1145/3386385

Kose, A., Tepljakov, A., Astapov, S., Draheim, D., Petlenkov, E., and Vassiljeva, K. (2018). Towards a synesthesia laboratory: real-time localization and visualization of a sound source for virtual reality applications. *J. Commun. Softw. Syst*. 14, 112–120. doi: 10.24138/jcomss.v14i1.410

Kühnle, S., Ludwig, A., Meuret, S., Küttner, C., Witte, C., Scholbach, J., et al. (2013). Development of auditory localization accuracy and auditory spatial discrimination in children and adolescents. *Audiol. Neurotol*. 18, 48–62. doi: 10.1159/000342904

Lam, B., Tan, Y. X., Heng, J., Hong, J. Y., Ong, Z.-T., and Gan, W.-S. (2020). "Virtual, augmented and mixed reality tools in soundscape evaluation and design," in *Proceedings on Internoise 2020*, Seoul.

Licitra, G., Memoli, G., Botteldooren, D., and Coensel, B. D. (2005). "Traffic noise and perceived soundscapes: a case study," in *Proceedings of Forum Acousticum 2005*, Budapest.

Linden Lab (2020). Have an Adventure. San Francisco, CA. Available online at: https://www.lindenlab.com

Litovsky, R. Y., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., et al. (2004). Bilateral cochlear implants in adults and children. *Arch. Otolaryngol. Head Neck Surg*. 130, 648–655. doi: 10.1001/archotol.130.5.648

Lugten, M., Karacaoglu, M., White, K., Kang, J., and Steemers, K. (2018). Improving the soundscape quality of urban areas exposed to aircraft noise by adding moving water and vegetation. *J. Acoust. Soc. Am*. 144:2906. doi: 10.1121/1.5079310

Madole, D., and Begault, D. (1995). 3-D sound for virtual reality and multimedia. *Comput. Music J*. 19, 99–101. doi: 10.2307/3680997

Makous, J. C., and Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am*. 87, 2188–2200. doi: 10.1121/1.399186

Malpica, S., Serrano, A., Allue, M., Bedia, M. G., and Masia, B. (2020). Crossmodal perception in virtual reality. *Multim. Tools Appl*. 79, 3311–3331. doi: 10.1007/s11042-019-7331-z

Melody, V. R. (2020). *Live Music in Virtual Reality*. London. Available online at: https://www.melodyvr.com

Memoli, G., Bloomfield, A., and Dixon, M. (2008a). "Soundscape characterization in selected areas of Central London," in *Proceedings - European Conference on Noise Control*, Paris. doi: 10.1121/1.2935534

Memoli, G., Hamilton-Fletcher, G., and Mitchell, S. (2018). Soundscape assessment of aircraft height and size. *Front. Psychol*. 9:2492. doi: 10.3389/fpsyg.2018.02492

Memoli, G., and Licitra, G. (2012). *From Noise to Annoyance Mapping: A Soundscape Approach*. Reading: CRC Press, 371. doi: 10.1201/b12885-24

Memoli, G., Licitra, G., Cerchiai, M., Nolli, M., and Palazzuoli, D. (2008b). "Measuring soundscape improvement in urban areas," in *Proceedings of the Institute of Acoustics*.

Milgram, P., and Takemura, H. (1994). "Augmented reality: a class of displays on the reality-virtuality continuum," in *Proc. SPIE 2351, Telemanipulator and Telepresence Technologies*, Boston, MA. doi: 10.1117/12.197321

Miller, N. (2013). Understanding soundscapes. *Buildings* 3, 728–738. doi: 10.3390/buildings3040728

Mills, A. W. (1958). On the minimum audible angle. *J. Acoust. Soc. Am*. 30, 237–246. doi: 10.1121/1.1909553

Müller, J., Geier, M., Dicke, C., and Spors, S. (2014). "The BoomRoom: mid-air direct interaction with virtual sound sources," in *Conference on Human Factors in Computing Systems - Proceedings*, Toronto, ON. doi: 10.1145/2556288.2557000

Neoran, I., Ben-Asher, M., and Alchanati, G. (2020). "Virtual reality music in the real world," in *Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality*.

Oberman, T., Jambrošić, K., Horvat, M., and Šcitaroci, B. (2020). Using virtual soundwalk approach for assessing sound art soundscape interventions in public spaces. *Appl. Sci*. 10:2102. doi: 10.3390/app10062102

Obrist, M., Boyle, G., van Brakel, M., and Duerinck, F. (2017). "Multisensory experiences & spaces," in *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, ISS '17* (New York, NY: ACM), 469–472. doi: 10.1145/3132272.3135086

Ochiai, Y., Hoshi, T., and Suzuki, I. (2017). "Holographic whisper: rendering audible sound spots in three-dimensional space by focusing ultrasonic waves," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17* (New York, NY: ACM), 4314–4325. doi: 10.1145/3025453.3025989

Ott, J., Tutescu, A.-S., Wienboker, N., Rosenbauer, J., and Gorne, T. (2019). "Spatial audio production for immersive fulldome projections," in *5th International Conference on Spatial Audio*, Ilmenau.

Payne, S. R., Davies, W. J., and Adams, M. D. (2009). *Research Into the Practical and Policy Applications of Soundscape Concepts and Techniques in Urban Areas (NANR 200)*. Technical report, Salford University.

Rajguru, C., Blaszczak, D., PourYazdan, A., Graham, T. J., and Memoli, G. (2019). "Audiozoom: location based sound delivery system," in *SIGGRAPH Asia 2019 Posters, SA'19* (New York, NY: Association for Computing Machinery). doi: 10.1145/3355056.3364596

Rayleigh, L. (1907). XII. On our perception of sound direction. *Lond. Edinburgh Dublin Philos. Mag. J. Sci*. 13, 214–232. doi: 10.1080/14786440709463595

Roffler, S. K., and Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane. *J. Acoust. Soc. Am*. 43, 1255–1259. doi: 10.1121/1.1910976

Roginska, A., and Geluso, P., (eds.). (2018). *Immersive Sound*. New York, NY; London: Taylor and Francis. doi: 10.4324/9781315707525

Rummukainen, O. S., Schlecht, S. J., and Habets, E. A. P. (2020). No dynamic visual capture for self-translation minimum audible angle. *J. Acoust. Soc. Am*. 148, EL77–EL81. doi: 10.1121/10.0001588

Rungta, A., Rewkowski, N., Klatzky, R., Lin, M., and Manocha, D. (2017). Effects of virtual acoustics on dynamic auditory distance perception. *J. Acoust. Soc. Am*. 141:EL427. doi: 10.1121/1.4981234

Sato, H., Morimoto, M., and Sato, H. (2020). Perception of azimuth angle of sound source located at high elevation angle: effective distance of auditory guide signal. *Appl. Acoust*. 159:107084. doi: 10.1016/j.apacoust.2019.107084

Schafer, M. (1976). *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester: Destiny Books.

Sodnik, J., Tomazic, S., Grasset, R., Duenser, A., and Billinghurst, M. (2006). "Spatial sound localization in an augmented reality environment," in *ACM International Conference Proceeding Series*. (Sydney), 111–118. doi: 10.1145/1228175.1228197

Sra, M., Mottelson, A., and Maes, P. (2018). "Your place and mine: designing a shared VR experience for remotely located users," in *DIS 2018 - Proceedings of the 2018 Designing Interactive Systems Conference*. doi: 10.1145/3196709.3196788

Steadman, M. A., Kim, C., Lestang, J. H., Goodman, D. F., and Picinali, L. (2019). Short-term effects of sound localization training in virtual reality. *Sci. Rep*. 9:18284. doi: 10.1038/s41598-019-54811-w

Stevens, F., Murphy, D. T., and Smith, S. L. (2018). "Soundscape auralisation and visualisation: a cross-modal approach to soundscape evaluation," in *DAFx 2018 - Proceedings: 21st International Conference on Digital Audio Effects*. Aveiro.

Sung, W.-P., Zhou, J.-l., and Xie, L.-F. (2016). "Application of soundscape in city parks by case study," in *Proceedings of the 2016 International Conference on Green Building, Environment, Energy and Civil Engineering*. 113. doi: 10.1201/9781315375106-24

Tabry, V., Zatorre, R., and Voss, P. (2013). The influence of vision on sound localization abilities in both the horizontal and vertical planes. *Front. Psychol*. 4:932. doi: 10.3389/fpsyg.2013.00932

The British Library (2011). *UK Sound Map*. Available online at: https://sounds.bl.uk/Sound-Maps/UK-Soundmap.

Wagler, A., and Hanus, M. D. (2018). Comparing virtual reality tourism to real-life experience: effects of presence and engagement on attitude and enjoyment. *Commun. Res. Rep*. 35, 456–464. doi: 10.1080/08824096.2018.1525350

Watts, G. R., Pheasant, R. J., and Horoshenkov, K. V. (2011). Predicting perceived tranquillity in urban parks and open spaces. *Environ. Plann. B* 38, 585–594. doi: 10.1068/b36131

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). Localization using non-individualized head related transfer functions. *J. Acoust. Soc. Am*. 94, 111–123. doi: 10.1121/1.407089

Wightman, F. L., and Kistler, D. J. (1989). Headphone simulation of free field listening. i: Stimulus synthesis. *J. Acoust. Soc. Am*. 85, 858–867. doi: 10.1121/1.397557

WilkinsAvenue (2019). *Media Kit*.

Yang, J., Frank, Y., and Soros, G. (2019). "Hearing is believing: synthesizing spatial audio from everyday objects to users," in *ACM International Conference Proceeding Series*. (Reims) 1–9. doi: 10.1145/3311823.3311872

Yoneyama, M., Fujimoto, J., Kawamo, Y., and Sasabe, S. (1983). The audio spotlight: an application of nonlinear interaction of sound waves to a new type of loudspeaker design. *J. Acoust. Soc. Am*. 73, 1532–1536. doi: 10.1121/1.389414

Yung, R., and Khoo-Lattimore, C. (2019). New realities: a systematic literature review on virtual reality and augmented reality in tourism research. *Curr. Issues Tour*. 22, 2056–2081. doi: 10.1080/13683500.2017.1417359

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

CONCEPTUAL ANALYSIS
published: 20 November 2020
doi: 10.3389/fpsyg.2020.573041

Assessments of Acoustic Environments by Emotions – The Application of Emotion Theory in Soundscape

André Fiebig[1], Pamela Jordan[2] and Cleopatra Christina Moshona[1]*

[1] *Engineering Acoustics, Institute of Fluid Dynamics and Technical Acoustics, Technische Universität Berlin, Berlin, Germany,* [2] *Amsterdam Centre for Ancient Studies and Archaeology, University of Amsterdam, Amsterdam, Netherlands*

Human beings respond to their immediate environments in a variety of ways, with emotion playing a cardinal role. In evolutionary theories, emotions are thought to prepare an organism for action. The interplay of acoustic environments, emotions, and evolutionary needs are currently subject to discussion in soundscape research. Universal definitions of emotion and its nature are currently missing, but there seems to be a fundamental consensus that emotions are internal, evanescent, mostly conscious, relational, manifest in different forms, and serve a purpose. Research in this area is expanding, particularly in regards to the context-related, affective, and emotional processing of environmental stimuli. A number of studies present ways to determine the nature of emotions elicited by a soundscape and to measure these reliably. Yet the crucial question—which basic and complex emotions are triggered and how they relate to affective appraisal—has still not been conclusively answered. To help frame research on this topic, an overview of the theoretical background is presented that applies emotion theory to soundscape. Two latent fundamental dimensions are often found at the center of theoretical concepts of emotion: valence and arousal. These established universal dimensions can also be applied in the context of emotions that are elicited by soundscapes. Another, and perhaps more familiar, parallel is found between emotion and music. However, acoustic environments are more subtle than musical arrangements, rarely applying the compositional and artistic considerations frequently used in music. That said, the measurement of emotion in the context of soundscape studies is only of additional value if some fundamental inquiries are sufficiently answered: To what extent does the reporting act itself alter emotional responses? Are all important affective qualities consciously accessible and directly measurable by self-reports? How can emotion related to the environment be separated from affective predisposition? By means of a conceptual analysis of relevant soundscape publications, the consensus and conflicts on these fundamental questions in the light of soundscape theory are highlighted and needed research actions are framed. The overview closes with a proposed modification to an existing, standardized framework to include the meaning of emotion in the design of soundscapes.

Keywords: soundscape, emotion, mood, appraisal, soundscape descriptors, affect, affective quality

OPEN ACCESS

Edited by:
Bert De Coensel, Ghent University, Belgium

Reviewed by:
Rory Allen, Goldsmiths, University of London, United Kingdom
Jooyoung Hong, Singapore University of Technology and Design, Singapore

***Correspondence:**
André Fiebig
andre.fiebig@tu-berlin.de

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *16 June 2020*
Accepted: *29 October 2020*
Published: *20 November 2020*

Citation:
Fiebig A, Jordan P and Moshona CC (2020) Assessments of Acoustic Environments by Emotions – The Application of Emotion Theory in Soundscape. Front. Psychol. 11:573041. doi: 10.3389/fpsyg.2020.573041

INTRODUCTION

The field of soundscape focuses on how people experience their surrounding acoustic environments. This disciplinary position stands in contrast to the field of noise control, which focuses on human response to loudness and annoyance derived from environmental noise exposure. Soundscape's broader view of sonic experience naturally points to the potential of incorporating findings from affect, emotion and appraisal research, particularly as both noise and soundscape fields already borrow related language and concepts (e.g., annoyance as a metric). Human responses to the (acoustic) environment may even be a reflection of evolved motivational and affective systems, promoting survival through preferences for certain environments and avoidance of others (van den Bosch et al., 2018). In order to place potential benefits stemming from emotion theory within the context of soundscape research and assessment, a brief review of emotion theory is first necessary.

Emotion Theory and Research

Emotions are a nearly constant aspect of the human phenomenal experience (Nielsen and Kaszniak, 2007), with states such as fear, happiness, boredom or amusement arising without conscious effort. With a subject–and lived experience–so familiar to everyone, the scientific approach to the study of affective and emotional states[1] faces a challenge: any emotion theory must stand up to scientific rigor alongside any individual's common-sense examination. This dual standard for research on emotion is likely one reason why an established theory of emotion does not yet exist (Müller and Reisenzein, 2013).

Even so, research abounds. Rottenberg et al. (2007) have traced the explosive growth of research during the past few decades, leading to new theories, methods, and findings. Coan and Allen (2007) substantiate this, highlighting the great diversity of methodological approaches that are currently driving emotion science. Many researchers address the issue of separating emotion from cognition, the relation of cause and effect, the distinction between basic and complex emotions, conscious and unconscious aspects of emotions, the relation between rationality and emotion, and the true origin of emotion. Some key texts along these lines will be highlighted in the discussion that follows. Overall, emotions seem to be an integral concept that subsumes psychological stress and coping, uniting motivation, cognition and adaptation in a complex configuration (Lazarus, 1991).

As such, emotion is difficult to tackle by a single traditional psychological theory. Yet the study of the nature and structure of emotion has a long tradition that is still developing. It was recognized in the 19th century, the early days of psychophysics as a field, that body and mind are deeply intertwined. James had concluded that, if we consider a strong emotion and try to partition the feelings of its characteristic bodily symptoms from our consciousness of the emotion, there remains no "mind-stuff" from which the emotion can be constituted (James, 1884). Over 100 years later, Gross acknowledged that this tension remains rather unresolved: the "*definition of the construct emotion is* [still] *in a state of conceptual and definitional chaos and remains a heavily freighted term full of imprecision*" (Gross, 2010). Therefore, deriving a definition of emotion remains "[...] *a difficult matter [*and*] a definition of emotion can only be a product of theory*" (Frijda, 1986). Furthermore, the attempts by different disciplines to access emotion research via their own concepts and methods seems to impede the development of a universal view on emotion (Müller and Reisenzein, 2013). But Ekman has pointed out that what is really needed–rather than a comprehensive, universal theory of emotion–is to have a separate theory for each emotion in order to capture its unique aspects (Ekman, 1994).

There are a few aspects of emotion that appear to be recognized across disciplinary borders, which will be addressed in more detail:

- Emotions are internal, mostly conscious, and relational.
- Emotion can manifest in different forms. Frequently, the emotion phenomena are differentiated according to physiological responses, experiences, and behavior.
- Emotions are short-lived phenomena and must be distinguished from mood and attitude by means of duration. As emotions are short-lived processes affected by the moment, mood and attitude are more stable, less affected by the moment, and long-lasting.
- Emotions serve a purpose.

Relational Aspects

Emotions are internal psychological experiences, yet they are both relational and elicited by others or a specific encounter with an environment (Lazarus, 1991). The experience of an emotion can usually be linked to a specific, defining moment and triggered by a specific object, which makes emotions different from mood and attitude (Gray and Watson, 2007). The events that elicit emotions also appear to fulfill a special role – they are not simply stimuli. In fact, they appear to act through their significance, their meaning, their rewarding or aversive nature (Frijda, 1986).

As mentioned, the primary function of emotions is to provide feedback for reacting effectively to the environment (Clore, 1994). The processes of appraisal can be consciously controlled in part, but elements of the appraisal process, such as basic emotions (e.g., happiness, sadness, anger, and fear) and their functions, remain closed to comprehensive cognitive penetration (Frijda, 1986). The magnitude of emotional response an individual experiences is strongly related to the magnitude of emotional stimulus, and in this sense emotion is relational. But individuals experience emotions differently, attributable to a person's inherent emotional predispositions, what Larsen and Diener call personal emotionality (Larsen and Diener, 1987). Thus, the element of stimulus is always an intrinsic property that affects a human's emotions. At the

[1] *Affect* and *emotion* are often used interchangeably in literature, although these terms intend to denote different phenomena. The majority of theories considers emotion an integral part of the superordinate category of affect (Gross, 2010). Affective phenomena thus go beyond emotion and incorporate further aspects like personality traits or well-being. In the following text, there is no sharp distinction made between these terms – as no uniformly accepted distinction is available, both aspects are highly connected and of utmost importance to soundscape perception.

same time, knowledge about a stimulus' significance to well-being, inherent in the concept of appraisal, contributes to one's personal meaning that also drives emotional responses (Lazarus and Smith, 1988).

Although it is widely assumed that humans have access to their emotions and can report on them (cf. Müller and Reisenzein, 2013), it is likely that they have no direct access to the causal connections between external forces and internal responses. Individuals are simply limited in their ability to track the complex causal story of their emotions (Russell, 2003); even though emotions themselves are conscious, any appraisals leading to them are often unconscious (Clore, 1994). Sometimes the cause is obvious, but at other times individuals experience a change in affect without exactly knowing why (Russell, 2003). While understanding the stimulus and context for a response is an intrinsic feature of soundscape work, the difficulty in identifying the causes of emotions presents a challenge for emotion theory and soundscape research alike.

Forms

Emotions can be triggered by all human sensory systems, demonstrating an intrinsic link between emotional and physiological responses (Hume and Ahtamad, 2013). When looking at affective pictures, patterns of physiological change are found that vary with reports of affective valence and arousal (Bradley and Lang, 2000). Similar patterns of physiological reactions are elicited by affective pictures (Lang et al., 1993), affective sounds (Choi et al., 2015) and films (Fredrickson and Kahneman, 1993). In another illustrative study, when individuals viewed unpleasant pictures, a cardiac deceleration, a large skin conductance response, observable increases in corrugator (frown) electromyogram (EMG), a larger scalp-recorded positivity, and a potentiation of the startle reflex were observed (Gray and Watson, 2007). As emotions manifest in varying forms with many sub-components (Juslin, 2013a), the emotional response can be observed and measured in different ways, such as affective reports, physiological reactivity, and overt behavioral acts (Bradley and Lang, 1994). Distinguishing the various forms of emotional response and developing research methodologies to measure them is an important consideration for formulating appropriate soundscape studies.

Duration

In contrast to longer-lived moods, which can last hours or even days, emotions are intensive yet brief (Gray and Watson, 2007). Emotions and mood can be linked by duration in some circumstances, such as when a series of mild positive events together result in a positive mood over time (Davidson, 1994). So moods and emotions can be seen to interact dynamically (though the duration criterion only applies to a limited extent). As moods last longer, their causes are more remote in time and less salient compared to emotions, which are closer to the cause and thus seem to be more conscious (Clore, 1994). **Figure 1** illustrates the role of time in distinguishing emotion, mood, and attitude, distinctions that are especially relevant in soundscape studies. Moreover, the phenomenon of *duration neglect* is frequently discussed. This term refers to the insignificance of duration for reporting summarized affect of longer periods. The irrelevance of duration was observed in several empirical contexts, like pain or loudness perception or the displeasure of movie clips, and is important for reporting about emotions as well.

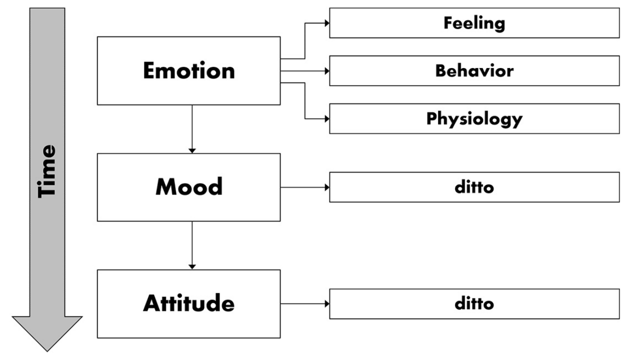


FIGURE 1 | Simplified model of emotion adapted from Gross (Gross, 2010).

Purpose

By means of elicited emotions, humans can rapidly recognize and quickly adapt necessary behavioral responses. Emotion thus can be understood as a driver of behavior. Emotions most often arise in situations where adaptive action is required (Davidson, 1994); they provide a means for dealing with fundamental situations quickly without much elaborate planning (Ekman, 1992). This could help explain the observation that emotional stimuli are prioritized in perception, are detected more rapidly, and gain access to conscious awareness more easily than non-emotional stimuli (Brosch et al., 2010). Such dynamics are reflected in the ability of humans to detect even subtle emotional nuances in speech and to adapt to them accordingly, for instance (Paeschke, 2003). Going further, Lang et al. (1993) state that *valence* and *arousal* represent primitive motivational parameters that define a general disposition to approach or avoid stimulation. Because judgments of pleasure and arousal reflect (in part) this motivational imperative, Lang et al. (1993) postulate a correlation between brain state and evaluation. Many researchers addressing emotion theories agree that emotional stimuli and emotional responses represent a special type of input – they both represent high relevance for survival and well-being by preparing the organism for action or decision (Gray and Watson, 2007; Brosch et al., 2010). Emotions inform the individual of the nature and importance of events, and the magnitude of feelings motivates an individual to focus quickly on relevant considerations (Clore, 1994). Lang et al. (1993) have proposed that the multi-dimensional emotional experience underlying affective judgments represents a bi-motivational structure involving two systems of appetitive and defensive motivation in life. However, Hall et al. (2013) believe that the emotional experience underlying real environments is perhaps too complex to be captured by only two motivational factors. This framing supports the idea that a number of physiological systems are primarily sensitive to emotional activation across sensory modalities rather than to a specific mode of presentation of stimuli, such as images or sounds (Bradley and Lang, 2000).

Though a consolidated theory of emotion is the subject of ongoing research, the specific role of emotion for managing inner and outer worlds, including characteristics and features, is well acknowledged and bears significance for soundscape research.

Introduction of Soundscape

The idea of soundscape was introduced in the late sixties as a contrast to the conventional perspectives of noise control and environmental policies at that time. According to Schafer (2012) (one of the founders of soundscape) all urban sounds should be the subject of study, "not merely those that were unpleasant or dangerous." This position significantly broadened the view on the distinct effects of sound on humans beyond the environmental noise abatement paradigm, which considered noise solely as a waste and the least annoying acoustic environment to be one free of any (unwanted) noise. Today it is known that the mere reduction of noise levels does not necessarily lead to more positive appraisals of an environment (van den Bosch et al., 2018).

The widening of scope to include both positive and negative sonic effects has led to a research shift from physical stimulus alone to human auditory sensation and its interpretation. The first concepts of soundscape emphasized that an acoustic environment is understood by those living within it and creating it (Truax, 1984). This early notion of soundscape was echoed in the recent international standard on soundscape, ISO 12913-1: "*Soundscape is an acoustic environment as perceived or experienced and/or understood by a person or people, in context*" (International Organization for Standardization, 2014). The recognized term soundscape thus refers to the perceived acoustic environment of a place, whose character is the result of the action and interaction of natural and/or human factors (Kang et al., 2016). Soundscape research focuses on perception under contextual conditions.

Research on soundscape has become more and more popular, and the field continues to explore new facets of how acoustic environments affect human perception. The overarching aim of soundscape research is to understand the relationship of people and their acoustic environment, examining the sounds that people value or oppose as well as the shifts in reaction due to changing location and activity (Kamp et al., 2016). For that purpose, various approaches have been proposed for studying the meaning of (environmental) sounds for humans and for determining the specific characteristics of perception; one of the most important and relevant for soundscape study is the verbal report.

VERBAL REPORTS TO STUDY HOW HUMANS EMOTIONALLY REACT TO ENVIRONMENTS

In the late nineteenth century, Wundt recognized that emotions are composed of three major dimensions — "*Lust*" and "*Unlust*" (pleasure and displeasure), "*Erregung*" and "*Beruhigung*" (excitement and tranquilization), and "*Spannung*" and "*Lösung*" (tension and relaxation) (Wundt, 1906), terms which still seem current. Many psychologists since Wundt have agreed that the dimensional concept of emotion is a useful approach to provide a taxonomy of emotions and have searched for broadly applicable generic labels (Gehm and Scherer, 1988). Dimensional verbal reports of this variety would be also familiar to recent soundscape researchers (Axelsson et al., 2010) and will be addressed later on.

However, there continues to be a lively debate about the fundamental dimensions that characterize the phenomenal space of emotion experience (Nielsen and Kaszniak, 2007). Many researchers have followed the dimensional theory approach in the belief that affect and emotion are composed of a small number of general dimensions that are usually thought to be independent of each other. Gray and Watson (2007) pointed out that "researchers began to adopt models that bypassed these discrete affects and posited few underlying dimensions." As discussed in the introduction, emotions present a complex mixture of consciously accessible and intuitive responses that are captured in dimensional models. Although emotions have both behavioral and physiological characteristics, Lazarus (1991) concluded that emotions are above all psychologica+l. Clore (1994) emphasizes that "[. . .] *one cannot have an unconscious emotion, because emotion involves an experience, and one cannot have an experience that is not experienced*." As psychological states that are consciously accessible by their receivers, emotions can thus be effectively studied using participatory self-report methods. Intriguingly, such assessments appear to be stable over time: considering retrospective reporting of emotions over specific time intervals, it seems that participants have little trouble giving relatively reliable and valid emotion ratings (Robinson and Clore, 2002).

Continuing the explorations of dimensional models, Osgood et al. (1975) observed fundamental semantic dimensions such as *evaluation*, *activity* and *potency* by investigating the nature of meaning of languages using the semantic differential method. The dimensions across later research using the semantic differential method frequently bear a striking resemblance to the dimensions observed by Osgood et al. (1975) — *hedonic valence*, *activity* and *potency* (Gehm and Scherer, 1988). An influential work in line with Osgood et al. (1975) was later published by Mehrabian and Russell (1974) using the multivariate research on affective language, finding that the principal variance in emotional meaning appears to be sufficiently explained by a limited set of basic emotional responses to all situations: the main independent factors *pleasure, arousal,* and *dominance*. *Pleasure* must be distinguished from preference or liking, while *arousal* describes a single dimension ranging from sleepy to excitement. However, less attention is paid to dominance in research and models are used with only two axes: the degree of pleasure oriented horizontally, and the degree of arousal oriented vertically (Bakker et al., 2014). These terms have recently been adopted within soundscape, so their application in emotion research bears a moment of further consideration.

The dimensions identified as *pleasure* and *arousal* are frequently obtained in factor analytic solutions based on a set of data consisting of a heterogeneous sample of adjective items and a set of rated stimuli. Factors that emerge are expected to denote fundamental affective or perceptual components. Russell et al. (1981), building on the work from Mehrabian and Russell, developed a circumplex model of affective states elicited by environments, a circle in a two-dimensional bipolar space based on the dimensions of *pleasure-displeasure* and *arousal-sleep*. In a circumplex model, descriptors are systematically arranged around the perimeter of a circle leading to bipolar dimensions, revealing the relationships between two separate dimensional scales. Bakker et al. (2014) refer to the underlying mechanism to explain *pleasure* and *arousal* as related to the *degree of order* and *variation*.

The two-dimensional model has received extensive empirical support as the same basic two-dimensional structure consistently emerged in self-report data (Gray and Watson, 2007). A similar, though not identical, model receiving attention was proposed by Watson and Tellegen, who emphasized the importance of *negative affect* and *positive affect* as independent dimensions. The *negative affect* reflects unpleasant affective states with low or high arousal states, whereas the *positive affect* dimension ranges from enthusiastic and excited to sleepy and drowsy (Watson and Tellegen, 1985). There are some debates surrounding the bipolarity and independence of dimensions implied in the different models [e.g., Is positive affect the bipolar opposite of, or is it independent of, negative affect? (Feldmann Barrett and Russell, 1998)]. Yet Russell and Carroll (1999) detected no substantive controversy and a consensus on a descriptive structure of current affect seems imminent (Feldmann Barrett and Russell, 1998). Although meaning attributed to environments contains both affective and perceptual-cognitive components with the two highly interrelated, the detected latent fundamental dimensions focus specifically on emotions (Russel and Pratt, 1980). The identified dimensions of affective qualities are currently applied by numerous researchers, though there is and will be a continuing debate about the interpretation of the dimensions with their underlying mechanisms (Bakker et al., 2014). Russell himself acknowledged that his own dimensional model of emotion fails to provide a sufficiently rich account of *prototypical emotional episodes* such as distinguishing between *fear* or *anger* (Russell, 2003), fuelling the debate about the dimensional or categorical nature of emotions. However, the extensive evidence from similarity judgments between emotion related adjectives, judgments of facially expressed emotions, self-reported mood, and psychophysiological measurements indicates that two dimensions are usually considered to be sufficient (Västfjäll et al., 2002). The same basic two-dimensional structure consistently emerges in self-report data, leading to the conclusion that this structure is considered fundamental or *basic* as described by Watson et al. (1999).

As verbal reports frequently refer to a certain period experienced in the past, aspects of duration that were discussed in the introduction might become relevant. Delayed judgments of a past episode reduce the relevance of the episode's total duration, salient single moments become even more important, and at the same time other distinct emotions are glossed over during the episode (Fredrickson and Kahneman, 1993). Emotion reporting often requires participants to remember and summarize their experiences when giving an account of past emotions. Retrospective biases, such as recollection and weighing of specific moments of an experience or belief-based reconstruction, must be considered in such reporting (Robinson and Clore, 2002). It is very likely that, if retrospective measures of emotion experiences are requested, respondents create emotion reports using different types of processing strategies – retrieval of prior experiences versus reconstruction of the past experiences, for instance (Feldmann Barrett, 1997). Altogether, it appears that two distinct emotional selves are available: one that lives in the moment and one that lives in the abstract, which means that distinct sources of self-knowledge are accessed under different reporting conditions when referring to ongoing or to retrospective emotions (Robinson and Clore, 2002). According to Gärling et al. (2020), in the context of emotional wellbeing, the most valid and reliable method is the self-report on momentary states (e.g., *How do you right feel now?*), because instantaneous self-report measures are barely influenced by memory distortions and subject of meta-analyses. There is still a significant lack of understanding in the role of duration, memory, and integration heuristics on environmental sound-induced emotion and its reporting; systematic investigations on these issues are rarely conducted. However, the general value of self-reports for emotion research cannot be questioned and prove essential for soundscape research as well.

EMOTIONS AND THEIR DIMENSIONS IN SOUNDSCAPE RESEARCH

Soundscape research generally acknowledges that the process of perceiving and assessing environmental sound is multi-dimensional and the simplifying concept of annoyance is insufficient for thorough analysis (Schulte-Fortkamp and Fiebig, 2016; Jordan, 2019). Therefore, the consideration of basic and complex emotions within soundscape work is logical, and emotion theory is increasingly gaining significance in applied soundscape research. Sounds have been demonstrated to elicit emotional processes in experimentally controlled laboratory contexts with standardized affective stimulus databases, i.e., the International Affective Digitized Sounds IADS-I (Bradley and Lang, 1999) and IADS-II (Bradley and Lang, 2007b). It seems that environmental sounds carry biologically significant information reflected in human emotional responses, and that emotions work to optimize adaptive responses to biologically meaningful events (Ma and Thompson, 2015). However, research on the auditory system has been less intensively performed in the past than research on the visual system (Yang et al., 2018).

Emotion in Music Versus Soundscape

One sound-related area that has not been neglected by emotion is music. It seems beyond question that music as an auditory event can provoke emotions. According to Juslin and Västfjäll (2008), emotions can be evoked in different ways and to different degrees by different stimuli, and music is no exception. The dominant approaches to conceptualize emotions are classified as *categorical* and *dimensional*. As such, Juslin (2013b) concluded based on empirical evidence that musical expression of emotion is likely to involve emotion categories rather than mere dimensions. The ability of music to affect human emotions is derived from its arrangement and not just from its sonic material. Music seems to channel the complexity of the acoustic world into an ordered form (Truax, 1984). This suggests that acoustic environments are capable of evoking a similar set of emotions. In one study, changes in acoustic attributes that evoke emotional responses in speech and music (e.g., frequency spectrum, intensity, and rate) were observed also to induce emotions when perceived in environmental sounds (Ma and Thompson, 2015). According to Ma and Thompson (2015), this observation aligns with the *musical protolanguage hypothesis* that speech and music originated from a common emotional signal system based on the imitation and modification of sounds in the environment. Truax (2016) has observed that, although intense affective responses as expressions of emotions through speech and music have been studied extensively, the equivalent role of environmental sounds has unfortunately so far been ignored.

Despite the fact that the concept of 'soundscape' is originally rooted in music (Kang et al., 2016), as well as Schafer's assertion that "from art, particularly music, we will learn how man creates ideal soundscapes" (Schafer, 1977), the mechanisms connecting music and emotions are substantially different to the mechanisms at work in soundscape-elicited emotions. Music is (almost always) composed intentionally to arouse a wide range of emotions. Listeners usually consciously experience music, engage in decoding "intended" emotions and are aware of the manifold stylistic elements to "inspire" the audience. The effect of acoustic environments on emotions is more subtle and often goes unnoticed. Acoustic environments are rarely explicitly ordered or designed to induce emotions. Accordingly, Ma and Thompson observed that core acoustic attributes relevant for elicited emotions by music and speech are also relevant for the emotional character of environmental sounds, but the authors simultaneously explain that acoustic environments have other acoustic attributes with emotional significance (Ma and Thompson, 2015). It is evident that the findings of emotion theory regarding music cannot be directly mapped onto soundscape contexts.

Dimensional Models in Soundscape

Soundscape researchers searching for basic soundscape-related emotions and their underlying indicators have strongly based their concepts on common findings in environmental psychology with respect to the *dimensional* notion of emotion and affect. For example, Russell et al. (1981) explained that "[...] *exciting places are both pleasant and arousing. Peaceful and comfortable places are also pleasant but unarousing. Frightening and harsh places are unpleasant and high in arousing quality. Depressing places are unpleasant and unarousing.*" (Russell et al., 1981). These observations pave the way for similar understandings of the effects of acoustic environments on people. Bradley and Lang (2000) discovered that physiological responses elicited by visual stimuli appear to be organized fundamentally along dimensions of *pleasure* and *arousal*, implicating underlying motivational systems of appetite and defense and suggesting the likely intermodal generalizability of these dimensions. Consequently, Mehrabian and Russell (1974) believed in a common core of responses as an immediate result of stimulation to all types of stimuli regardless of the sense modality stimulated, a stance which has been influential for soundscape researchers looking for fundamental emotion dimensions elicited by acoustic environments.

Emotion theory holds that pleasant and unpleasant feelings form a bipolar continuum (Russell and Carroll, 1999), which dovetails with the fundamental soundscape concept that sound is a resource (Schafer, 1977). The soundscape approach focuses on sounds that are preferred by humans, as opposed to noise control's focus on sounds of discomfort – those causing sleep disturbance, annoyance, communication interference, or effects on cognitive processes (Brown and Lam, 2015). As emotion theory centers on the relationship between person and environment rather than on either environment or intrapersonal events alone (Lazarus, 1991), the current trend in soundscape research to study emotions is propitious. Moreover, because emotions seem to have evolutionary roots in preparing the organism for action, the meaning of emotions, their link to the acoustic environment, and evolutionary needs are understandably subject to discussion. The circumplex concept as an approximation of fundamental emotions is a convenient and heuristic affect model in this case. It is not surprising that soundscape-related emotion researchers have adopted this notion of elicited emotions and that the affective concepts of Mehrabian and Russell attributed to environments frequently serve as a starting point.

Indeed, Bjork (1985) replicated Mehrabian and Russell's dimensions *pleasantness* and *arousal* in the context of elicited emotions by natural sounds. Later, Axelsson et al. (2010) intensively studied the affective qualities attributed to acoustic environments and proposed a few basic dimensions of affective qualities for soundscapes that reflect the main features of the circumplex model (Russell et al., 1981). **Figure 2** presents a side-by-side comparison of Russell's research and Axelsson et al. (2010)'s recent application of the dimensional model in soundscape contexts:

In their work, Axelsson et al. (2010) discovered the basic dimensions *pleasantness*, *eventfulness* and *familiarity* in the context of soundscape. However, Axelsson et al. (2010) point out that the small variation in familiarity of soundscapes results means that the familiarity component is considered to be of limited importance for applied work, though it may at least be relevant to basic research. In an interesting convergence, the underlying dimensions of affect that were detected for acoustic stimuli are similar to those determined for affective image processing (Bradley and Lang, 2007a; Axelsson, 2011). The first two independent dimensions, *pleasantness* and *eventfulness*, might reflect evolutionary needs across sensory domains, promoting survival by preferring certain environments

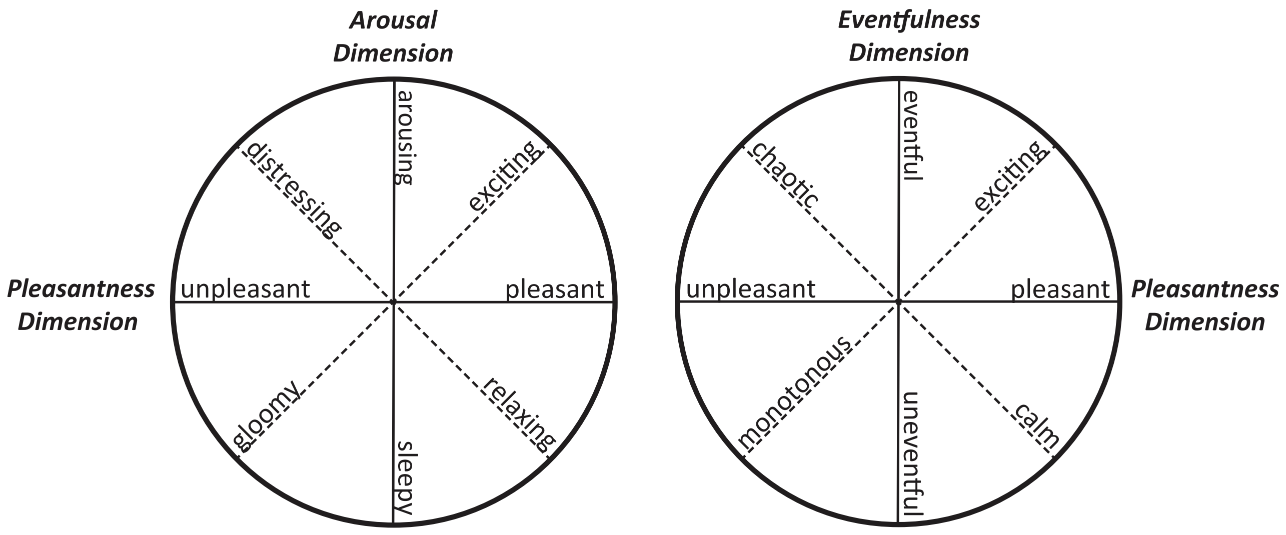


FIGURE 2 | Two-dimensional representation of the affective quality attributed to physical environments generally (left, adapted from Russell et al., 1981) and to acoustic environments in particular (right, adapted from Axelsson et al., 2010).

and avoiding others (van den Bosch et al., 2018). In 1984, Truax had already conjectured the main soundscape-related dimensions *variety* and *coherence*, which seem to be close to Axelsson's proposed dimensions *pleasantness*, *eventfulness* and *familiarity*. *Eventfulness* can be considered as a semantic dimension of (auditory) *order* and *variation*. For example, a busy flea market with bustling activities or a popular, overcrowded urban city park are commonly perceived as eventful.

The Diversity of Soundscape (Emotion) Dimensions

Beyond the typical dimensions related to *hedonic valence* and *arousal*, sometimes soundscape investigations explore other or additional dimensions through statistical analysis to reduce the number of observed variables to a few fundamental ones. The dimensions proposed as appropriate to soundscape have expanded significantly in recent years. Aletta et al. (2016) suggested *appropriateness* as a third soundscape dimension for consideration. An encountered situation is usually matched against existing cognitive schemes, i.e., personal expectations; thus *appropriateness,* the level of match between expectation and real-world situation, can influence an individual's positive affective responses to a situation. In contrast, inappropriate matches lead to negative affective responses (van den Bosch et al., 2018), again harkening to survival origins. Tarlao et al. (2019) determined basic dimensions that they labeled *appreciation*, *dynamism,* and *monotony* as separate factors. Cain et al. (2013) and Davies et al. (2013) observed *calmness* and *vibrancy* as principal dimensions of emotional responses to soundscapes, which appear to be similar to the rotated circumplex model of Axelsson et al. (2010). Aletta and Kang (2018) investigated descriptors predicting *vibrancy* and surprisingly did not observe a significant correlation with pleasantness. This may indicate an independent dimension or, as the authors suggest, an accidental measurement of *eventfulness* being obtained through the research (Aletta and Kang, 2018). Andringa and van den Bosch (2013) referred to the main dimensions *pleasure* and *activation* in their work. Welch et al. (2019) observed the soundscape dimensions *calming*, *protecting*, *hectic*, *belonging* and *stability*. Yu et al. (2016) extracted the major factors of soundscape perception to be *preference*, *loudness*, *communication*, *playfulness,* and *richness* in the context of urban shopping streets. Sudarsono et al. (2019) derived the dimensions *privacy, disturbance*, *dynamic*, *fear*, and *satisfaction* in crowded third-class hospital wards. Zhang and Kang (2020) tried to distinguish between felt and perceived emotions induced by soundscapes and identified in their factor analysis dimensions labeled *comfort*, *enjoyment, excitement, desolation, tension,* or *familiarity* indicating a mixture of *hedonic valence* and *activation* dimensions.

Table 1 lists the detected soundscape dimensions in selected publications. As shown in **Table 1**, most of the listed studies are based on controlled laboratory experiments. Field surveys were only rarely conducted to determine fundamental dimensions of emotions in soundscape.

Universality of Dimensions

For Jeon et al. (2018), the components *pleasantness* and *eventfulness* commonly identified in several studies from different countries appear to be universal across languages, cultures, and environments. The ISO/TS 12913-2 expresses the general appreciation for this model in proposing a questionnaire consisting of response scales related to different affective attributes (International Organization for Standardization, 2018). The use of multiple ratings across sets of scales in the circumplex allows for reliable assessments of core affects including main emotional dimensions as recommended in the ISO/TS 12913-2 and ISO/TS 12913-3.

TABLE 1 | Soundscape descriptors as emotion dimensions.

Authors	Detected dimensions	Applied method
Truax, 1984	Coherence, variety**	Theoretical deduction
Bjork, 1985	Pleasantness*, arousal**	Semantic differential method, principal components analysis (L)
Västfjäll et al., 2003b	Valence*, activation**	Multiple rating scales, sum of scales (L)
Axelsson et al., 2010	Pleasantness*, eventfulness**, familiarity	Semantic differential method, principal components analysis (L)
Cain et al., 2013	Calmness*, vibrancy**	Semantic differential method,[2] principal components analysis (L)
Andringa and van den Bosch, 2013	Pleasure*, activation**	Based on literature
Yu et al., 2016	Preference*, loudness, communication, playfulness richness**	Semantic differential method
International Organization for Standardization, 2019	Pleasantness*, eventfulness**	Defined based on literature
Tarlao et al., 2019	Appreciation*, dynamism**, monotony	Semantic differential method, principal components analysis (F)
Sudarsono et al., 2019	Privacy, disturbance*, dynamic**, fear, satisfaction*	Semantic differential method, principal components analysis (F)
Welch et al., 2019	Calming**, protecting*, hectic, belonging, stability	Qualitative method analyzing written text, semantic differential method, principal components analysis (L)
Zhang and Kang, 2020	Enjoyment*, excitement**, desolation, tension, familiarity (*related to felt emotions*)	Semantic differential method, factor analysis (L)
Zhang and Kang, 2020	Comfortable*, festive**, desolate, familiar, attractive*, nostalgic (*related to perceived emotions*)	Semantic differential method, factor analysis (L)

Dimensions are marked with asterisks if they resemble the Mehrabian and Russell's pleasantness (*) and arousal (**) dimensions. The letters L (laboratory experiment) and F (field study) indicate the general type of the study. [2]The reference to the "semantic differential method" in this table refers to the use of multiple category rating scales to be judged by participants that vary in format and design.

However, although emotions are understood to be essentially universal, cultural differences in emotions are frequently reported, suggesting a social component within the elicitation of emotion. Choi et al. discovered inconsistencies regarding the relationship between categorical emotions and dimensional emotions, which may reflect cultural differences (Choi et al., 2015), whereas Jeon et al. (2018) attributed differences in reported emotions to different connotative meanings and semantics rather than the emotions themselves. It seems likely that, the universal character of emotion applies to the human set of emotions, whereas a cultural impact takes place more on the emotion regulation stage (cf. Mesquita and Frijda, 1992).

Moreover, fundamental differences between studies lie in their instructive process – that is to say, whether the participants were requested to report "*how the sound makes you feel*" (Cain et al., 2013) vs. "*how the sound environment is*" (ISO/TS 12913-2). Accordingly, Axelsson (2011) defined affective quality as a property of the stimulus that refers to its capacity to change our emotional responses thereby capturing the notion of *perceived* emotions. Kallinen and Ravaja (2006) observed in the context of music-induced emotions that, even though the *perceived* and *felt emotions* were more or less the same, they also demonstrated differences. Thus, it is likely that differences in the determination of emotion dimensions are due to the missing distinction of *perceived* emotions (*assigned* intrinsic property of the stimulus) and *felt* emotions (*elicited* emotions within the individual).

Overall, it appears that the first two dimensions discussed, *calmness/pleasantness* and *activity/eventfulness,* frequently emerge in numerous investigations, as indicated in **Table 1**. These could be regarded as a preliminary standard model for the perceptual dimensions of soundscapes (cf. Davies et al., 2014). Nevertheless, the search for additional dimensions to complement the widely established standard model appears to be ongoing, as current studies are still producing results that cannot yet be generalized across all contexts.

WHAT DETERMINES EMOTIONAL RESPONSES TO ACOUSTIC ENVIRONMENTS?

In daily life, the various types of external stimuli that humans receive across different modalities have powerful effects on evoked emotions, influencing decision-making and subsequent behavior (Yang et al., 2018). However, the link between external stimuli and elicited emotions is still subject to extensive research. If the intrinsic properties of soundscapes leading to certain basic emotions are well understood, soundscape designers could intentionally create emotional soundscape compositions to evoke a target mood (Fan et al., 2016).

Axelsson (2011) highlighted the importance of *information load*, which drives one's affective responses to stimuli. Aesthetic appreciation is grounded in the relationship between the amount of information of stimuli and people's capacity to process this information, which leads to emotional responses. According to Axelsson (2011), the amount of information of a stimulus is absolute while the degree of information load is relative, depending on the individual's processing capacity.

In this approach it appears obvious that emotional responses are not solely dependent on the stimulus but are also a part of the perceiving individual. The notion of information load agrees with findings of Mehrabian and Russell, who used the concept of information rate related to meaningfulness, familiar events versus novel, and unexpected, surprising events (Mehrabian and Russell, 1974).

van den Bosch et al. (2018) related affective qualities to the indicators *affordance* and *complexity* and thereby advanced the establishment of *audible safety* as a driving force of appraisal. Affordance can be understood as cues from the environment that immediately allow the detection of function, and these cues in turn furnish behavior (Gibson, 1979). Using an evolutionary perspective, *audible safety* is an important cue in environments for warning humans of potential danger. Auditory environments that lack considerable *audible safety* require people to become vigilant and alert, resulting in stress and appraised unpleasantness. This perspective leads to the assumption that observed affective quality dimensions reflect old evolutionary motives of surviving in, coping with, and flourishing in an environment. The concept of *audible safety* resembles the semantic dimension of *control/power* observed by Gehm and Scherer (1988). Human beings appraise their soundscapes based on the level of safety they attribute to them, which guide emotional response and behavior (van den Bosch, 2015). This notion implies that soundscapes are not only appraised through emotional-based factors, but also by the extent of safety attributed to them. van den Bosch (2015) argued that the understanding of the acoustical properties of a place is far less important than understanding how that place influences a person emotionally. In the context of pictures as affective stimuli, (Bradley and Lang, 2007b) claimed that no obvious physical parameters can be used to organize emotional stimuli and to predict emotion. As the general concept of psychophysics postulates a measurable relationship between physical stimuli and the perceptions they produce, the search for the causes of emotion laying outside the human mind appears consequential. According to Frijda (1986), the links between stimulus and response are prewired, innate stimulus-response connections.

When it comes to sound stimuli, studies have shown that the affective quality of sounds, including acute physiological reactions, do not depend solely on the intensity of sounds (Bradley and Lang, 2000). This observation is fully in line with the soundscape theory, which assumes that soundscape exists through perception of the acoustic environment influenced by a multitude of factors (International Organization for Standardization, 2014). Accordingly, Davies et al. (2013) pointed to physiological experiments demonstrating that the body and brain respond to emotional content as well as simple noise levels. Bradley and Lang (2007a) reported that about 14% of the arousal variance concerning the set of International Affective Digitized Sounds IADS could be attributed to sound intensity variations. The IADS database consists of 167 natural sounds of 6 s duration that are common in daily life, which elicit different responses on the affective dimensions of *valence*, *arousal*, and *dominance* (Choi et al., 2015). Yang et al. (2018) confirmed the findings of Bradley and Lang and observed that the relationship between a physical intensity of sound and *valence* looked more complex and that classical level indicators explained only a few percentage points of the total variance.

Figure 3 proposes a conceptual framework for understanding the process of emotional responses triggered by a soundscape, drawing from the various outcomes of previous research on emotions induced by acoustic environments. The diagram builds directly on the conceptual framework for a soundscape laid out in ISO 12913-1, which describes the process of perceiving an acoustic environment in context (International Organization for Standardization, 2014). The factor *context* continues to stand for the interactions between an individual and their (*acoustic*) *environment* (sound sources and their specific configuration), including all interrelationships in space and time between person, activity and place (International Organization for Standardization, 2014). Context here also includes elements such as the personal history, life experiences, and cultural background of the individual. The new conceptual framework introduced above, which squarely integrates facets of emotion in its structure and organization, stands apart from the known framework in the feedback loop anchored by appraisal. Here, the initial affective appraisal of a soundscape influences first short-term behavioral responses (such as moving away from the area), which in turn influence longer-term outcomes (such as habits or health effects). The resulting shifts in mood, attitude, and knowledge held by an individual may then modify prospective appraisals, leading to modified responses and so on. The conceptual framework emphasizes the importance of the frequently unconsciously elicited (basic) emotions by a soundscape, which exert influence on individuals' behavior, well-being, and health without one being aware of it.

It appears that most soundscape research dealing with emotions does not differentiate basic emotions and appraisal. It is necessary to understand the nature of emotion and emotion processing as they are increasingly applied in studies that map soundscape (emotion) descriptors to physical indicators. The challenge is that most indicators do not consider context and meaning. Research has shown that emotional responses to sounds allegedly devoid of meaning seem to imply physical characteristics that induce affect (Västfjäll, 2012). But without considering the meaning attributed to sounds and situations, a process that always occurs in the real perceptual world, acoustical indicators do not allow comprehensive prediction of the (emotional) responses. Accordingly, van den Bosch (2015) explained that the acoustical properties of a place are far less important than understanding how the direct experience of that place influences a person emotionally. The pursuit of identifying the determinants of emotions beyond physical indicators appears justified.

It appears that the underlying mechanism to explain *pleasure* and *arousal* is related to the degree of *order* and *variation*, and these terms point the way for identifying appropriate (acoustic) indicators. The different endeavors to determine valid indicators with large amounts of explained variance illustrate

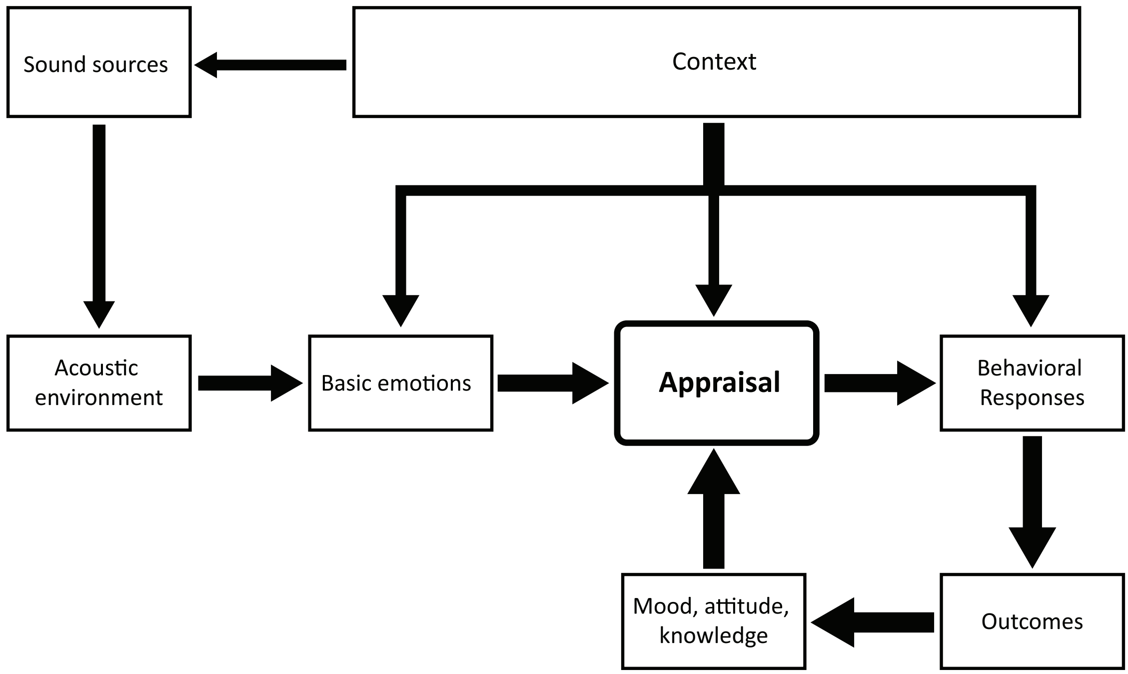


FIGURE 3 | Conceptual framework of the emotional construct of soundscape [adapted from ISO 12913-1 (International Organization for Standardization, 2014)].

that non-acoustic indicators must be considered. Västfjäll et al. (2003a) asserted that multimodal affective perception of an environment differs from unimodal perception. Consequently soundscape, as a multi-dimensional perception of an (acoustic) environment, requires the consideration of multimodal affective perception. This indicates the necessity in predictive models to integrate different sensory modalities. However, emotional reactions to short sound situations observed in experiments, which represent only brief glimpses, cannot simply be attributed to the operation of different underlying 'motivational states' in real life (Hall et al., 2013).

It seems that only one conclusion can be drawn from the hunt for the underlying indicators so far: before it is possible to establish predictive models of soundscape, it is necessary to fully agree upon the necessary descriptors to be predicted (Aletta et al., 2016). Although there is a growing body of knowledge regarding the predictability of emotion-related soundscape descriptors by means of acoustic and non-acoustic indicators, the comprehensive mixture of models, equations, and formulas using a wide variety of different indicators shows the general lack of consensus between researchers regarding the roots and causes of soundscape emotion and appraisal.

CONCLUSION

In the 1970s, the soundscape pioneer Schafer demanded that the soundscape analyst must begin by discovering the significant features of the soundscape (Schafer, 1977). According to the latest soundscape research, elicited emotions are significant soundscape features aside the component sounds themselves. The explicit incorporation of emotions into soundscape research appears to be highly justified. Emotions elicited by soundscapes do not merely affect how we experience the sounds around us; they also color other information we process, such as the interpretations of people and events (cf. Ma and Thompson, 2015). It seems that emotion is a simple reaction to a soundscape as well as a fleeting source for several major, less evanescent phenomena. However, the exact role of emotion in the context of soundscape has not yet been clarified.

Emotion and affect can be measured in terms of physiological (re)activity, (overt) behavior, and affective self-reports. So far, soundscape research has turned its attention mostly to the measurement of verbal reporting on emotions and affect (Kuppens et al., 2012). However, it seems that a methodological distinction is rarely made between requesting related reports to intrinsically or extrinsically triggered emotions. Although the impact of the missing distinction between empirical outcomes might be minor, it may be possible that a stimulus can elicit a felt emotion differently than the emotional quality perceived by the listener (Kallinen and Ravaja, 2006; Zhang and Kang, 2020).

The inclusion of emotion-related elements into the common conceptual framework of ISO/TS 12913-1 opens the door to a progressive integration of emotion theory within soundscape and promises to guide future research substantially. In contrast to the ISO framework, the modified conceptual framework introduced in this article includes the loop of solidified emotions transformed into mood and attitudes entering future appraisals. The distinction between the different stages of emotion and appraisal including long-term effects must guide further research.

By currently accepting *hedonic valence* (pleasantness) and *arousal* or *activation* (eventfulness) as the main affective descriptors of soundscape appraisal among soundscape researchers (Davies et al., 2014), the field of soundscape study has initiated the hunt for underlying indicators (van den Bosch et al., 2018). It seems that those descriptors of soundscape appraisal can be substituted with common descriptors such as *annoyance* or *quality* (Aletta et al., 2016). The *Pleasure* and *arousal* dimensions that underlie affective judgments represent appetitive and defensive motivation, leading to responses and outcomes as described in the ISO 12913-1 (International Organization for Standardization, 2014).
According to Kang et al. (2016), the commonly identified dimensions put emphasis on emotion linked to the appraisal of soundscapes and therefore need to be addressed in soundscape research. However, the emotion-stimulating potential of acoustic environments on human beings is still not comprehensively understood. "*We often experience emotions as happening to us, not as chosen by us. We do not simply decide when to have a particular emotion*" (Ekman, 1994). Therefore, a better understanding of emotions' causes and effects is essential for any design of soundscapes. Unfortunately, emotions promoted by vibrant and lively soundscapes such as those in public urban areas still lack deeper investigations that incorporate emotion theory (Carvalho et al., 2019). However, studies have shown that emotional responses to soundscapes largely resemble emotions otherwise induced by the other senses (cf. Axelsson, 2011). This feeds the hope of developing a universal concept referring to the link between stimulus and elicited emotion independent of the sensory domain. More research will be necessary to determine possible interactions between various sensory responses to emotion.

The recent progress made within soundscape research of establishing emotion-related categories and dimensions as a core principle in soundscape research offers new options in characterizing acoustic environments from the perspective of perception. It marks significant advancement compared to the simplified, singular focus on annoyance from noise research that preceded soundscape inquiry. Beyond the almost established dimensions, it seems necessary to continue work on context-related descriptors like *affordance, coherence,* or *congruence*. Supported by emotions, perception always encompasses the conversion from sensory input to something coherent and meaningful. These categories are particularly important because pleasantness is not the only design motif employed in creating preferred soundscapes. As Davies et al. (2014) observed, participants designed soundscapes based on what was expected or appropriate rather than simply on what they liked. However, aspects like expectation or appropriateness involve cognitive processing and go beyond automatic emotions elicited by the very moment.

Research on emotion in soundscape opens exciting new research pathways. By understanding the emotional responses in different soundscapes, the knowledge of the acoustic environment might help to approach the management of urban sound as a resource for design practice (Carvalho et al., 2019).

FUTURE RESEARCH TASKS

It is beyond doubt that a deeper understanding of emotions elicited by soundscapes and their measurability would be a significant step forward for soundscape research. It would allow for improving perception-related assessment of actual soundscapes as well as promoting advanced design techniques. However, significant questions remain:

(1) What are the limits on the reportability of emotion experience? Can we exclusively rely on self-reported emotional experiences, assuming that the most important affective qualities are accessible by consciousness? To what extent do studies reveal information about the nature of emotion and not only about the nature of semantic concepts underlying the used attributes and scales? (Russell, 1980; Gehm and Scherer, 1988; Nielsen and Kaszniak, 2007).
(2) To what extent does the very act of reporting alter the emotional response itself? (Nielsen and Kaszniak, 2007; Rottenberg et al., 2007).
(3) How is emotion temporally structured? What is the time window for measuring experiential, behavior and physiological responses? In what way are long-lasting emotional states composed of single fleeting, evanescent emotions? Do human beings use heuristics when reporting their emotions over short episodic versus longer time frames? What is the relationship between retrospective measures and aggregated instantaneous measures? (Feldmann Barrett, 1997; Robinson and Clore, 2002; Nielsen and Kaszniak, 2007; Rottenberg et al., 2007; Gärling et al., 2020).
(4) What sorts of reporting schemes are best suited to the different emotion dimensions and affective qualities, and are these schemes culturally invariant? (Mesquita and Frijda, 1992; Nielsen and Kaszniak, 2007).
(5) As human emotion is relational and individual, is it worthwhile separating the intrinsic emotional potential of the environment from the different *appraisal histories people have* and different *affect intensities as an individual magnitude of emotional responsiveness* which influence emotions? (Larsen and Diener, 1987; Lazarus and Smith, 1988).

It seems that the lively hunt for underlying indicators of the established fundamental dimensions of emotion might obstruct the necessary view on fundamental but still unanswered theoretical issues. The measurement of emotion for soundscape studies is only of additional value if researchers work on the fundamental theoretical questions before driving headlong into more field-based research initiatives.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Aletta, F., and Kang, J. (2018). Towards an urban vibrancy model: a soundscape approach. *Int. J. Environ. Res. Public Health* 2018:1712. doi: 10.3390/ijerph15081712

Aletta, F., Kang, J., and Axelsson, Ö. (2016). Soundscape descriptors and a conceptual framework for developing predictive soundscape models. *Landsc. Urban Plan.* 149, 65–74. doi: 10.1016/j.landurbplan.2016.02.001

Andringa, T. C., and van den Bosch, K. A. M. (2013). "Core affect and soundscape assessment: fore- and background soundscape design for quality of life," in *Proceedings of Internoise 2013*, Innsbruck.

Axelsson, Ö. (2011). *Aesthetic Appreciation Explicated*. Doctoral thesis, Stockholm University, Stockholm.

Axelsson, Ö., Nilsson, M. E., and Berglund, B. (2010). A principal components model of soundscape perception. *J. Acoust. Soc. Am.* 128, 2836–2846. doi: 10.1121/1.3493436

Bakker, I., van der Voordt, T., Vink, P., and de Boon, J. (2014). Pleasure, arousal, dominance: Mehrabian and Russell revisited. *Curr. Psychol.* 33, 405–421. doi: 10.1007/s12144-014-9219-4

Bjork, E. A. (1985). The perceived quality of natural sound. *Acustica* 57, 185–188.

Bradley, M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9

Bradley, M. M., and Lang, P. J. (1999). *International Affective Digitized Sounds (IADS-1): Stimuli, Instruction Manual, and Affective Ratings*. Technical Report No. B-2. Gainesville, FL: University of Florida.

Bradley, M. M., and Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology* 37, 204–215. doi: 10.1111/1469-8986.3720204

Bradley, M. M., and Lang, P. J. (2007a). "The international affective picture system (IAPS) in the study of emotion and attention," in *Handbook of Emotion Elicitation and Assessment*, eds J. A. Coan and J. J. B. Allen (New York, NY: Oxford University Press).

Bradley, M. M., and Lang, P. J. (2007b). *The International Affective Digitized Sounds: Affective Ratings of Sounds and Instruction Manual*. Technical Report No. B-3. Gainesville, FL: University of Florida.

Brosch, T., Pourtois, G., and Sander, D. (2010). The perception and categorisation of emotional stimuli: a review. *Cogn. Emot.* 24, 377–400. doi: 10.1080/02699930902975754

Brown, A. L., and Lam, K. C. (2015). Soundscape planning as a complement to environmental noise control. *Noise News Int.* 23, 14–21.

Cain, R., Jenning, P., and Poxon, J. (2013). The development and application of the emotional dimensions of a soundscape. *Appl. Acoust.* 74, 232–239. doi: 10.1016/j.apacoust.2011.11.006

Carvalho, M. L., Davies, W. J., and Fazenda, B. (2019). "Investigation of emotional states in different urban soundscapes through laboratory reproductions of 3D audiovisual samples," in *Proceedings of the 14th International Postgraduate Research Conference in the Built Environment: Contemporary and Future Directions in the Built Environment*, Salford.

Choi, Y., Lee, S., Choi, I. M., Jung, S., Park, Y. K., and Kim, C. (2015). International affective digitized sounds in Korea: a cross-cultural adaptation and validation study. *Acta Acust. United Acust.* 101, 134–144. doi: 10.3813/AAA.918811

Clore, G. L. (1994). "Why emotions are never unconscious," in *The Nature of Emotions. Fundamental Questions*, eds P. Ekman and R. J. Davidson (New York, NY: Oxford University Press), 285–290.

Coan, J. A., and Allen, J. J. B. (2007). "Introduction. Organizing the tools and methods of affective science," in *Handbook of Emotion Elicitation and Assessment*, eds J. A. Coan and J. J. B. Allen (Oxford: Oxford University Press).

Davidson, R. J. (1994). "On emotion, mood and related affective constructs," in *The Nature of Emotions. Fundamental Questions*, eds P. Ekman and R. J. Davidson (New York, NY: Oxford University Press), 15–19.

Davies, W. J., Adams, M. D., Bruce, N. S., Cain, R., Carlyle, A., Cusack, J., et al. (2013). Perception of soundscapes: an interdisciplinary approach. *Appl. Acoust.* 74, 224–231. doi: 10.1016/j.apacoust.2012.05.010

Davies, W. J., Bruce, N. S., and Murphy, J. E. (2014). Soundscape reproduction and synthesis'. *Acta Acust. United Acust.* 100, 285–292.

Ekman, P. (1992). An argument for basic emotions. *Cogn. Emot.* 6, 169–200. doi: 10.1080/02699939208411068

Ekman, P. (1994). "All emotions are basic," in *The Nature of Emotions. Fundamental Questions*, eds P. Ekman and R. J. Davidson (New York, NY: Oxford University Press), 15–19.

Fan, J., Thorogood, M., and Pasquier, P. (2016). Automatic soundscape affect recognition using a dimensional approach. *J. Audio Eng. Soc.* 64, 646–653. doi: 10.17743/jaes.2016.0044

Feldmann Barrett, L. (1997). The relationship among momentary emotion experiences, personality descriptions and retrospective ratings of emotion. *Pers. Soc. Psychol. Bull.* 23, 1100–1110. doi: 10.1177/01461672972310010

Feldmann Barrett, L., and Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *J. Pers. Soc. Psychol.* 74, 967–984. doi: 10.1037/0022-3514.74.4.967

Fredrickson, B. L., and Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *J. Pers. Soc. Psychol.* 65, 45–55. doi: 10.1037/0022-3514.65.1.45

Frijda, N. H. (1986). *The Emotions. Studies in Emotion and Social Interaction*. Cambridge: Cambridge University Press.

Gärling, T., Ettema, D., Connolly, F. F., Friman, M., and Olsson, L. E. (2020). Review and assessment of self-reports of travel-related emotional wellbeing. *J. Transp. Health* 17:100843. doi: 10.1016/j.jth.2020.100843

Gehm, T. L., and Scherer, K. R. (1988). "Factors determining the dimensions of subjunctive emotional space," in *Facets of Emotion. Recent Research*, ed. K. R. Scherer (Hillsdale, MI: Lawrence Erlbaum Associates).

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.

Gray, E. K., and Watson, D. (2007). "Assessing positive and negative affect via self-report," in *Handbook of Emotion Elicitation and Assessment*, eds J. A. Coan and J. J. B. Allen (New York, NY: Oxford University Press).

Gross, J. J. (2010). The future's so bright, I gotta wear shades. *Emot. Rev.* 2, 212–216. doi: 10.1177/1754073910361982

Hall, D. A., Irwin, A., Edmondson-Jones, M., Phillips, S., and Poxon, J. E. W. (2013). An exploratory evaluation of perceptual, psychoacoustic and acoustical properties of urban soundscapes. *Appl. Acoust.* 74, 248–254. doi: 10.1016/j.apacoust.2011.03.006

Hume, K., and Ahtamad, M. (2013). Physiological responses to and subjective estimates of soundscape elements. *Appl. Acoust.* 74, 275–281. doi: 10.1016/j.apacoust.2011.10.009

International Organization for Standardization (2014). *ISO 12913-1:2014 Acoustics-Soundscape-Part 1: Definition and Conceptual Framework*. Geneva: International Organization for Standardization.

International Organization for Standardization (2018). *ISO/TS 12913-2:2018 Acoustics-Soundscape-Part 2: Data Collection and Reporting Requirements*. Geneva: International Organization for Standardization.

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

James, W. (1884). What is an emotion? *Mind* 9, 188–205.

Jeon, J. Y., Hong, J. Y., Lavandier, C., Lafon, J., Axelsson, Ö., and Hurtig, M. (2018). A cross-national comparison in assessment of urban park soundscapes in France, Korea, and Sweden through laboratory experiments. *Appl. Acoust.* 133, 107–117. doi: 10.1016/j.apacoust.2017.12.016

Jordan, P. (2019). Historic approaches to sonic encounter at the Berlin Wall Memorial. *Acoustics* 1, 517–537. doi: 10.3390/acoustics1030029

Juslin, P. N. (2013a). From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Phys. Life Rev.* 10, 235–266. doi: 10.1016/j.plrev.2013.05.008

Juslin, P. N. (2013b). What does music express? Basic emotions and beyond. *Front. Psychol.* 4:596. doi: 10.3389/fpsyg.2013.00596

Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 559–621. doi: 10.1017/S0140525X08005293

Kallinen, K., and Ravaja, N. (2006). Emotion perceived and emotion felt: same and different. *Music. Sci.* 5, 123–147. doi: 10.1177/102986490601000203

Kamp, I. V., Klaboe, R., Brown, A. L., and Lercher, P. (2016). "Soundscapes, human restoration and quality of life," in *Soundscape and the Built Environment*, eds J. Kang and B. Schulte-Fortkamp (Boca Raton, FL: CRC Press).

Kang, J., Aletta, F., Gjestland, T. T., Brown, L. A., Botteldooren, D., Schulte-Fortkamp, B., et al. (2016). Ten questions on the soundscapes of the built environment. *Build. Environ.* 108, 284–294. doi: 10.1016/j.buildenv.2016.08.011

Kuppens, P., Champagne, D., and Tuerlinckx, F. (2012). The dynamic interplay between appraisal and core affect in daily life. *Front. Psychol.* 3:380. doi: 10.3389/fpsyg.2012.00380

Lang, P. J., Greenwald, M. K., Bradley, M. M., and Hamm, A. O. (1993). Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology* 30, 261–273. doi: 10.1111/j.1469-8986.1993.tb03352.x

Larsen, R. J., and Diener, E. (1987). Affect intensity as an individual difference characteristic: a review. *J. Res. Pers.* 21, 1–39. doi: 10.1016/0092-6566(87)90023-7

Lazarus, R., and Smith, C. (1988). Knowledge and appraisal in the cognition — emotion relationship. *Cogn. Emot.* 2, 281–300. doi: 10.1080/02699938808412701

Lazarus, R. S. (1991). *Emotion & Adaptation*. New York, NY: Oxford University Press.

Ma, W., and Thompson, W. F. (2015). Human emotions track changes in the acoustic environment. *Proc. Natl. Acad. Sci. U.S.A.* 112, 14563–14568. doi: 10.1073/pnas.1515087112

Mehrabian, A., and Russell, J. A. (1974). *An Approach to Environmental Psychology*. Cambridge, MA: MIT Press.

Mesquita, B., and Frijda, N. H. (1992). Cultural variations in emotions: a review. *Psychol. Bull.* 112, 179–204. doi: 10.1037/0033-2909.112.2.179

Müller, A. W., and Reisenzein, R. (2013). *Emotionen – Natur und Funktion. Philosophie und Psychologie im Dialog*. Göttingen: Vandenhoeck & Ruprecht.

Nielsen, L., and Kaszniak, A. W. (2007). "Conceptual, theoretical, and methodological issues in inferring subjective emotion experience. Recommendations for researchers," in *Handbook of Emotion Elicitation and Assessment*, eds J. A. Coan and J. J. B. Allen (New York, NY: Oxford University Press).

Osgood, C. E., Suci, G. E., and Tannenbaum, P. H. (1975). *The Measurement of Meaning*. Chicago, IL: University of Illinois Press.

Paeschke, A. (2003). *Prosodische Analyse Emotionaler Sprechweise*. Berlin: Logos Verlag.

Robinson, M. D., and Clore, G. (2002). Episodic and semantic knowledge in emotional self-report: evidence for two judgment processes. *J. Pers. Soc. Psychol.* 83, 198–215. doi: 10.1037/0022-3514.83.1.198

Rottenberg, J., Ray, R. D., and Gross, J. J. (2007). "Emotion elicitation using films," in *Handbook of Emotion Elicitation and Assessment*, eds J. A. Coan and J. J. B. Allen (New York, NY: Oxford University Press).

Russel, J. A., and Pratt, G. (1980). A description of the affective quality attributed to environments. *J. Pers. Soc. Psychol.* 38, 311–322. doi: 10.1037/0022-3514.38.2.311

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev.* 110, 145–172. doi: 10.1037/0033-295X.110.1.145

Russell, J. A., and Carroll, J. M. (1999). The phoenix of bipolarity: reply to Watson and Tellegen (1999). *Psychol. Bull.* 125, 611–617. doi: 10.1037/0033-2909.125.5.611

Russell, J. A., Ward, L. M., and Pratt, G. (1981). Affective quality attributed to environments: a factor analytic study. *Environ. Behav.* 13, 259–288. doi: 10.1177/0013916581133001

Schafer, R. M. (1977). *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester, NY: Destiny Books.

Schafer, R. M. (2012). Soundscape studies: the early days and the future. *Soundscape J. Acoust. Ecol.* 12, 6–8.

Schulte-Fortkamp, B., and Fiebig, A. (2016). "Impact of soundscape in terms of perception," in *Soundscape and the Built Environment*, eds J. Kang and B. Schulte-Fortkamp (Boca Raton, FL: CRC Press), 69–88. doi: 10.1201/b19145-5

Sudarsono, A. S., Sarwono, S. J., Shabrina, A., and Tamrin, L. N. (2019). "The soundscape dimensions of third-class hospital ward in Indonesia," in *Proceedings of International Congress on Acoustics 2019*, Aachen.

Tarlao, C., Steele, D., and Guastavino, C. (2019). "Investigating factors influencing soundscape evaluations," in *Proceedings of Internoise 2019*, Madrid.

Truax, B. (1984). *Acoustic Communication*. Norwood, NJ: Ablex Publishing Corporation.

Truax, B. (2016). Environmental sound and its relation to human emotion. *Can. Acoust.* 44:3.

van den Bosch, K. A. M. (2015). *Safe and Sound: Soundscape Research in Special Needs Care*. Doctoral thesis, University of Groningen, Groningen.

van den Bosch, K. A. M., Welch, D., and Andringa, T. C. (2018). The evolution of soundscape appraisal through enactive cognition. *Front. Psychol.* 9:1129. doi: 10.3389/fpsyg.2018.01129

Västfjäll, D. (2012). Emotional reactions to sounds without meaning. *Psychology* 3, 606–609. doi: 10.4236/psych.2012.38091

Västfjäll, D., Friman, M., Gärling, T., and Kleiner, M. (2002). The measurement of core affect: a Swedish self-report measure. *Scand. J. Psychol.* 43, 19–31. doi: 10.1111/1467-9450.00265

Västfjäll, D., Kleiner, M., and Gärling, T. (2003a). Affective reactions to and preference for combinations of interior aircraft sound and vibration. *Int. J. Aviat. Psychol.* 13, 33–47. doi: 10.1207/S15327108IJAP1301_3

Västfjäll, D., Kleiner, M., and Gärling, T. (2003b). Affective reactions to interior aircraft sound. *Acta Acust. United Acust.* 89, 693–701.

Watson, D., and Tellegen, A. (1985). Toward a consensual structure of mood. *Psychol. Bull.* 98, 219–235. doi: 10.1037//0033-2909.98.2.219

Watson, D., Wiese, D., Vaidya, J., and Tellegen, A. (1999). The two general activation systems of affect: structural findings, evolutionary considerations, and psychobiological evidence. *J. Pers. Soc. Psychol.* 76, 820–838. doi: 10.1037/0022-3514.76.5.820

Welch, D., Shepherd, D., Dirks, K., Tan, M. Y., and Coad, G. (2019). Use of creative writing to develop a semantic differential tool for assessing soundscapes. *Front. Psychol.* 9:2698. doi: 10.3389/fpsyg.2018.02698

Wundt, W. (1906). *Vorlesungen Über die Menschen- und Tierseele*. Hamburg: Voss Verlag.

Yang, W., Makita, K., Nakao, T., Kanayama, N., Machizawa, M. G., Sasaoka, T., et al. (2018). Affective auditory stimulus database: an expanded version of the international affective digitized sounds (IADS-E). *Behav. Res. Methods* 50, 1415–1429. doi: 10.3758/s13428-018-1027-6

Yu, B., Kang, J., and Ma, H. (2016). Development of indicators for the soundscape in urban shopping streets. *Acta Acust. United Acust.* 102, 462–473. doi: 10.3813/AAA.918965

Zhang, Y., and Kang, J. (2020). The development of emotional indicators for the soundscape in urban open public spaces. *IOP Conf. Ser. Mater. Sci. Eng.* 780:052006. doi: 10.1088/1757-899X/780/5/052006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 November 2020
doi: 10.3389/fpsyg.2020.570797

A Tale of Three Misters: The Effect of Water Features on Soundscape Assessments in a Montreal Public Space

Christopher Trudeau[1,2]*, *Daniel Steele*[1,2] *and Catherine Guastavino*[1,2]

[1] School of Information Studies, McGill University, Montréal, QC, Canada, [2] Center for Interdisciplinary Research in Music Media and Technology, McGill University, Montréal, QC, Canada

The acoustic environments of small, central urban parks are often dominated by traffic sounds. Water sounds can be used to mitigate the negative impacts of unwanted sounds through masking. Studies comparing the effects of different water sounds are typically conducted using recordings in laboratory settings where ecological validity is limited. An urban redesign project in Montreal took the innovative approach of trying three sequential temporary designs of a new public square, each of which included a distinct water feature that produced a lightly-audible mist. Here we report on a field experiment evaluating the effect of the water feature in each of the three designs. Respondents ($n = 274$) evaluated their experience with the three different designs using questionnaires including soundscape (SSQP) and restorativeness scales, and perceived loudness. The results indicate a significant interaction effect between the water feature and the design of the space, particularly on ratings of chaotic and loud. While two water feature designs had an overall "positive" effect (i.e., less loud and chaotic) on soundscape assessment, the third water feature design produced the opposite effect. These findings hold even after accounting for ambient temperature. This opportunity to test multiple water features in the same space revealed that water features do not automatically improve soundscape assessments. The visual design, function of the space and environmental conditions should be carefully considered and calls for more field studies. We discuss consequences and considerations for the use of water features in public spaces as well as the implications in terms of ecological validity of soundscape studies.

Keywords: urban soundscape, pocket park, restoration, urban design, field experiment, water feature

Edited by:
PerMagnus Lindborg,
City University of Hong Kong,
Hong Kong

Reviewed by:
Jooyoung Hong,
Singapore University of Technology
and Design, Singapore
Gunnar Cerwén,
Swedish University of Agricultural
Sciences, Sweden

****Correspondence:***
Christopher Trudeau
christopher.trudeau@mail.mcgill.ca

Specialty section:
This article was submitted to
Environmental Psychology,
a section of the journal
Frontiers in Psychology

Received: *09 June 2020*
Accepted: *27 October 2020*
Published: *25 November 2020*

Citation:
Trudeau C, Steele D and Guastavino C (2020) A Tale of Three Misters: The Effect of Water Features on Soundscape Assessments in a Montreal Public Space. Front. Psychol. 11:570797. doi: 10.3389/fpsyg.2020.570797

1. INTRODUCTION

Water features, as a broad category, have wide ranging uses in urban spaces of all sizes as visual landmarks, as gathering spots and as means to escape heat. They also have a range of uses and impacts on the sound environment, including the masking of unwanted sounds, such as traffic noise (Galbrun and Ali, 2013; Ekman et al., 2015). Yet water features come in a variety of shapes and sizes, with different combinations serving different purposes (Galbrun and Ali, 2013). Moreover, the users' sonic experience in a space depends not only on the sound environment, but also the listening context (Schulte-Fortkamp et al., 2007).

Addressing this relationship between sound environment and experience, a body of work on urban soundscape, defined by the ISO as the "acoustic environment as perceived or experienced and/or understood by a person or people, in context" (International Standards Organization, 2014) has focused on human perception. This ISO definition provides a potential framework to study the sound environments of pocket parks, wherein context "includes the interrelationships between person and activity and place" (International Standards Organization, 2014). Soundscape research considers multidisciplinary and mixed methods approaches in characterizing acoustic environments, with an emphasis on human perception, rather than the physical measurements used in traditional noise control approaches (e.g., decibel levels) (Dubois et al., 2004; Brown, 2010). This translates into a shift from the idea of sound as a pollutant to the potential of using sound as a resource (Schulte-Fortkamp et al., 2007).

An emerging question of interest in soundscape studies is on the use of sound as resource in an environment to provide restoration to its users. Restorative environments enable users to recover from the negative effects of noise exposure, including drained cognitive resources, and to reflect upon daily or life issues (Kaplan, 1995). Originally focused on visual environments, the concept of restoration has been extended to include soundscapes (Payne, 2008). As such, the acoustic environment also affords all of the facets of traditional restorative environments.

This study is conducted in the context of this body of soundscape research that emphasizes the context in which sounds are heard, in particular the context in which water features are deployed in the design of a new space. The aim is to balance the experimental control afforded by laboratory studies, where controlling many conditions is relatively easy, with the ecological validity of *in-situ* studies, where context is inherent in the research design. This study reports on the findings of an *in-situ* soundscape questionnaire deployed in a single public space as it underwent three temporary designs, including a misting water feature in each.

2. LITERATURE REVIEW

Large urban parks dominated by greenery have been shown to provide psychological restoration for their users (Jansson and Persson, 2010; Nilsson et al., 2010; Refshauge et al., 2012). Small urban public parks, referred to as "pocket parks" (Nordh and Østby, 2013) are often as busy as the surrounding city. The extent to which pocket parks afford restoration remains understudied. Attention Restoration Theory (ART) suggests that high-quality public spaces can have a positive impact on mental well-being as measured through four components: fascination, being-away, compatibility and extent (Kaplan and Kaplan, 1989). Additionally, there is a possible association between the use of restorative spaces and longer-term (i.e., lingering) attention restoration (e.g., Berto, 2005). This implies that a user's visit to high-quality spaces can have lasting effects on learning and work performance.

A laboratory-based study using visual assessments of pocket parks showed that they have the potential to afford recovery and restoration-related activities (Nordh and Østby, 2013). In particular, the potential for socializing activities was found to be an important element in restoration (Peschardt and Stigsdotter, 2013). To our knowledge, the sonic dimension has only more recently been touched upon in a systematic manner (e.g., Payne and Guastavino, 2018; Steele et al., 2019; Senese et al., 2020). In general, positively-perceived soundscapes are associated with positive effects on well-being (Aletta et al., 2016). Nature sounds appear to effect a faster recovery than other types which could be explained by positive emotions associated with nature (Alvarsson et al., 2010).

In the last decade, a number of soundscape scales have been developed and refined to measure human perceptions of acoustic environments and explore variations on what "sound as resource" could mean in practice (Axelsson et al., 2010; Tarlao et al., 2016; Engel et al., 2018). Axelsson and his team (Axelsson et al., 2010) created and validated the Swedish Soundscape Quality Protocol (SSQP), comprised of eight unidimensional scales (pleasant, unpleasant, eventful, uneventful, calm, monotonous, vibrant, and chaotic). The restorativeness of a sound environment has been operationalized using the Perceived Restorativeness Soundscape Scale (PRSS) (Payne and Guastavino, 2018). Developed from the Perceived Restorativeness Scale (PRS) (Hartig et al., 1997), the PRSS addresses each of the four components of ART in relation to its sound environment rather than the physical place (Payne and Guastavino, 2018).

Laboratory studies show that some water features can improve soundscape ratings of parks in urban areas dominated by road traffic noise (Jeon et al., 2012; Galbrun and Ali, 2013; Skoda et al., 2014; Ekman et al., 2015; Hong et al., 2020a; Senese et al., 2020). Small- and medium-sized features increased ratings of pleasantness (Ekman et al., 2015), peacefulness and tranquility (Galbrun and Ali, 2013), as well as restorativeness ratings (for the fascination and being-away components) (Senese et al., 2020). Moreover, adding desirable sounds, including water features, also reduces perceived loudness, though water features do not always increase pleasantness ratings (De Coensel et al., 2011; Hong et al., 2020a,b).

Due to the difficulty of creating control conditions, *in-situ* studies on the effects of water features on soundscape ratings in pocket parks are rare. To the knowledge of the authors, only two such studies exist. An *in-situ* study of a large fountain in an important Stockholm park found no statistically significant direct effects on soundscape ratings attributable to the fountain (Axelsson et al., 2014). The second study in the courtyard of a German building similarly found no statistically significant results from a small, functioning water feature (Skoda et al., 2014). The same study also evaluated the effect of water sounds over headphones on soundscape ratings in the same courtyard and found significant results. The authors argue that the headphones focused the participants' attention to the water sounds and, in the absence of alternate sound sources, tended toward a central response for each rating. Given the importance of the interrelationship between person, activity and place in soundscape assessment, more *in-situ* studies should be completed to complement lab-based studies on water features in urban parks of all sizes.

Studies have also evaluated the mechanisms through which the water features affect soundscape assessment of urban parks. In particular, two types of masking have been introduced to explain how this might occur: energetic masking and informational masking. In energetic masking, a secondary sound source disrupts the processing of the primary signal in the inner ear (Moore, 2012). Informational masking results when a masking sound that varies unpredictably or that is acoustically very similar to the primary signal produces more masking than would otherwise be expected from energetic masking alone (Moore, 2012). Natural streams and fountains using upward jets were more effective than waterfalls at improving ratings of peacefulness and tranquility, suggesting that energetic masking road traffic noise is not the primary mechanism mediating those particular ratings (Galbrun and Ali, 2013). Unlike waterfalls, smaller water features do not produce the same low-frequency content that is produced by road traffic, and energetic masking is therefore not likely to occur. No known research exists on the *in-situ* effects of lightly audible water features in urban public spaces.

The present study addresses these research gaps by focusing on two research questions:

RQ1—Can small water features that are lightly audible in a outdoor urban public space have an observable effect on soundscape *in-situ* assessments?

RQ2—Can the measured effects of a small mister change if it is deployed in different configurations within the same outdoor urban public space?

3. METHODS

During the summer of 2018, Montreal's Plateau-Mont-Royal borough embarked on the design of a new pocket park from a previously empty space. The ∼900 square meter space is open to public streets on the west, north and east sides, while the south side is bordered by an alleyway and row houses that are inaccessible to vehicular traffic. The street to the north is a busy commercial and transportation artery with a lot of vehicular traffic and many pedestrians. The east and west streets are both residential.

The borough collaborated with designers and facilitators from private firms to determine the needs of the local community through a series of public consultations (see the archived web page for more information[1]). Responding to the community input, three designs were created, highlighting these different needs (i.e., conversation, relaxation, and cultural entertainment). **Table 1** provides a full description of the design themes. The three designs used temporary amenities to inform options and democratize the process of creating the final design of the space (slated for 2021).

Each design included benches, planters, platforms, tables, and chairs in different arrangements. As well, each design featured a lightly-audible water feature that sprayed a fine, upward mist of water (hereafter, these features are referred to as misters). Emerging from small holes in the ground, the thin jets of water were sustained and would reach a height between 1.5 and 2 m, depending on the amount of wind. The misters produced a constant sound that was higher in frequency as compared to the human voice. While the same modular components were used for each iteration of the mister, the placement of these differed between designs. Across all designs, the misters were active during roughly 60% of the data collection sessions, allowing for a quasi-experimental approach featuring six conditions (see **Table 4** for a more detailed breakdown). In the context of this paper, Design refers to the collection of modular components, as well as their arrangement. While Design includes the misters, this component is given specific attention as a variable of interest because it was not always on.

In Design 1, the misters were configured as part of a central island where users could sit and view the feature, not unlike a traditional fountain. In Design 2, the misters were located mixed in with planters and vegetation in the quietest section of the park. In Design 3, the misters were located in more open spaces with a less clear function, and when the misters were off these spaces were often used by pedestrians.

3.1. Sound Level Measurements

Baseline sound level measurements were taken with a B&K 2250 Sound Level Meter, calibrated before use. A total of 24 $L_{eq,10min}$ measurements were taken when the misters were off. Recordings covered the weekday (Monday and Friday), weekend (Saturday) and evening/night (Friday and Saturday) periods at three different locations in the space (M1—the northwestern corner adjacent to the commercial artery; M2—the center; M3—the southeastern corner adjacent to the residential buildings; see **Figure 1**). There were no major events (e.g., construction projects, festivals) in the vicinity of the space during questionnaire taking that would strongly influence the sound level. Measurements taken at Location M1 ranged from 61.9 to 66.5 dBA, while those at Locations M2 and M3 were lower (M2: 57.9–61.7 dBA; M3: 57.3–61.4 dBA). These measurements justified the division of the space into the "noisy side," on the northern half, and the "quiet side," on the southern half.

3.2. Questionnaire Instrument

During each design, users of the space were asked to complete a questionnaire about the soundscape. The questionnaires included scales measured using 5-point Likert-scales that were drawn from the SSQP and PRSS scales. From the SSQP, we included the pleasant, monotonous, vibrant, chaotic, calm and eventful scales. The standard SSQP also includes two additional factors that were not included in our questionnaire: (1) uneventful, as it does not have an adequate translation in French; and (2) unpleasant, as it is so heavily correlated with pleasant (Tarlao et al., 2016). One scale was used from the PRSS: taking a break from the daily routine. In addition, two scales were added to the questionnaire to measure appropriateness for the respondent's activity and the perceived loudness of the space. **Table 2** provides more detailed information about the questionnaire construction, including the scales used (the full

[1] https://web.archive.org/web/20190301152402mp/https://www.realisonsmtl.ca/962mont-royal

TABLE 1 | Description of the designs and the locations of each mister within the space.

Design	Title	Theme	# of Misters	Mister location	Dates
1	"La place dans la place"	This design promotes gathering and meeting around an elevated platform.	1	In the center of the elevated platform.	31 May–24 June
2	"Une nouvelle promenade"	This design was bisected by a walkway. The northern section, adjacent to the busy road, was a waiting space. The southern section, adjacent to the quieter alleyway, was a garden space designed for relaxation.	2	Connected with the gardens near the south side.	02 July–12 August
3	"Un amphithéatre ouvert sur l'avenue"	This design promotes an animated space, featuring a stage in the south-west corner of the space.	2	Two locations: one in the center of the space and the other in a corner of the noisy-side.	20 August–30 October

Dates provided are the start and end dates for the specific design.

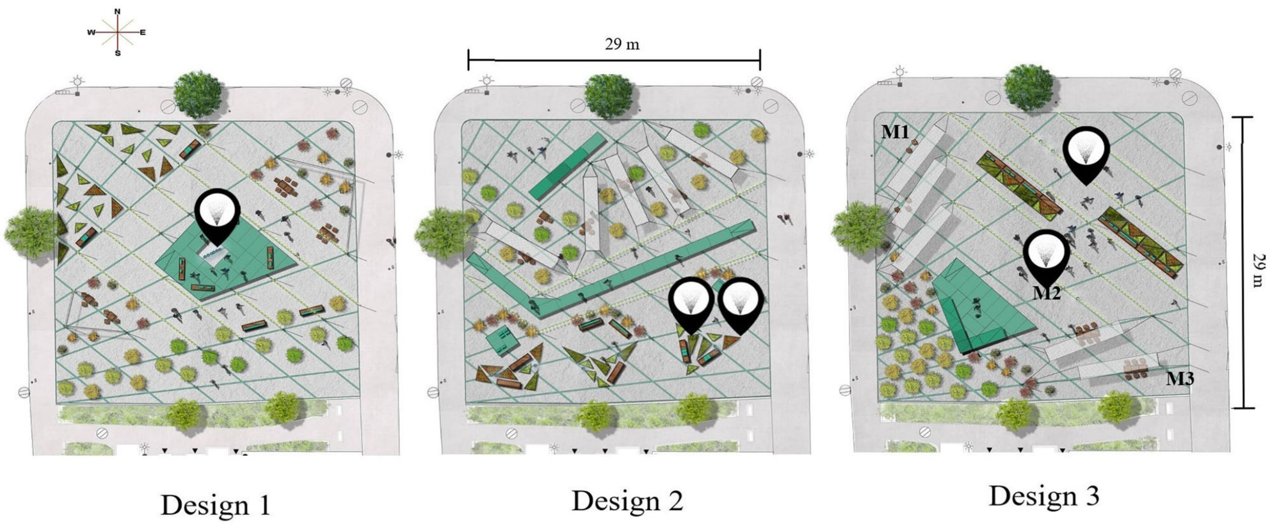


FIGURE 1 | The pins indicate the locations of the misters for each design. In Design 3, there were two misters that were roughly 10 meters apart, one in the center of the space and the other in the northeast corner. The locations M1, M2, and M3 in Design 3 indicate the approximate locations where the sound measurements were made. Design layouts provided by design firm Castor et Pollux and used with permission. Copyright to Cynthia Tarlao for this photo.

versions of both the English and French questionnaires can be found in the **Supplementary Material**). In the interest of clarity, these nine ratings will be collectively referred to as soundscape scales.

Respondents were also asked to list the sounds that they heard in the space and provide demographic information at the end of the questionnaire.

3.3. Procedure

Questionnaires were collected during 11 sessions, each of which lasted between 1 and 3 h, covering the hours of 11:00 a.m. to 9:00 p.m. The sessions varied in length due to weather, temperature and respondent availability, and were carried out on both weekdays ($n = 7$) and weekends ($n = 4$). Respondents were approached about taking the paper-based questionnaire after having already been in the space for at least 2–3 min. Some of the respondents were alone ($n = 129$) and others were in groups ($n = 145$). Respondents were able to complete the questionnaire using pen and paper in the language of their choice, English or French, and we collapsed the data across languages. Tarlao et al. (2019) provides a description of the differences between French and English responses.

In addition to questionnaire data, we tracked for each respondent the design (1, 2, or 3) and whether the mister was on or off (0 = off; 1 = on). **Figure 2** shows the effect and jet of the misters for design 2, though designs 1 and 3 were similar. For designs 2 and 3, we noted the respondents' location within the space. Based on the sound levels of the space and the conceptual designs of the space, these locations were then collapsed into only two areas: a noisy and a quiet side. Respondents were distributed almost evenly between the two halves of the space during the second and third designs. Additionally, Montreal temperature data was scraped from the website for Historical Climate Data from the Government of Canada.[2] The temperatures ranged

[2]https://climate.weather.gc.ca/

TABLE 2 | Detailed listing of the questions respondents answered that are relevant to the analysis in this paper.

Section	Question	Type	Scale
Activity	*What brings you here today?*	Open-ended	
Sound sources	*Please list below the sounds/noises that you are hearing around you.*	Open-ended	
Soundscape evaluation	I find this soundscape to be:		
	Pleasant	Likert-scale	SSQP
	Appropriate for my activity	Likert-scale	–
	Monotonous	Likert-scale	SSQP
	Vibrant	Likert-scale	SSQP
	Chaotic	Likert-scale	SSQP
	Calm	Likert-scale	SSQP
	Eventful	Likert-scale	SSQP
	Spending time in this soundscape gives me a break from my day-to-day routine	Likert-scale	PRSS
	I find the sound level here to be loud	Likert-scale	–
Demographic information	*I am*	Multiple choice *(Man/Woman/Other)*	
	I was born in the year	Open-ended	

The order of the questions in the table matches that of the questionnaire. The full English and French versions of the questionnaires can be found in the ***Supplementary Material****.*

FIGURE 2 | Picture of the mister from Design 2. The effect and jet of the misters for each design was similar. Copyright to Cynthia Tarlao for this photo.

between 16.5 and 32.7°C, with a small positive correlation between the temperature and the design ($r = 0.27$, $p = 0.00007$), suggesting a small but significant increase in temperature from Designs 1 to 3.

3.4. Open-Ended Question Categorization

The sounds listed by respondents were grouped into one of six categories: water, human, traffic, mechanical, nature, music and other. The categorization was mutually-exclusive, so that each source was only placed in one of the six categories. **Table 3** provides the definitions of each category and some examples of sources mentioned by respondents. Explicit mentions of water sounds (e.g., "fountain," "mister," "water") were categorized from this list, and a dummy variable was created to represent whether the respondent mentioned water as a sound source (0 = No, 1 = Yes).

3.5. Profile of the Respondents

In all, there were 274 respondents aged 18–84 (mean of 38 and a standard deviation of 15). Nine respondents did not provide their age. There were more women ($n = 154$) respondents than men ($n = 111$) respondents, though it does not represent a significant imbalance (χ^2, $df = 2$, p-value $= 0.19$). Nine respondents indicated "other" or "prefer not to answer" as gender. **Table 4** shows the demographic breakdown of the respondents per Design.

3.6. Statistical Analyses

The Likert-scales were converted to numbers to derive descriptive statistics (Strongly Disagree = 1; Disagree = 2; Neutral = 3; Agree = 4; Strongly Agree = 5). Missing values were replaced with the mean value of that scale, collapsed over all conditions. The number of missing values depended on the scale, but ranged between 1 (0.4%) and 18 (7%).

In order to investigate the effect of the mister and the design on the soundscape scales, we fit a 3 (design) × 2 (mister status) factorial MANCOVA as independent variables and temperature as a covariate. A MANCOVA test extends a standard MANOVA to include a covariate that cannot be accounted for through experimental design, as is the case for temperature in this study. Given the imbalance in the number of respondents for each of the six conditions, we ran a Levene test which was not significant, suggesting non-homogeneity of variance. Thus, we used Pillai's trace for our MANCOVA test to account for the non-homogeneity of variance. Significant MANCOVA results were further investigated using factorial ANCOVA tests for each scale separately. Finally, *post-hoc* tests were performed to determine which conditions account for changes in the scales. All p-values

TABLE 3 | Sounds mentioned by respondents were categorized into one of six groups.

Term	Definition	Examples
Human	Any source where the sound is the direct result of human activity or speech. Does not include music.	"Peoples' voices," "skateboards"[3]
Mechanical	Any non-vehicular mechanical source, especially construction and HVAC equipment.	"Supermarket AC"[4]
Music	Any sound created by a musical instrument, whether amplified or not.	"Guitar," "soft percussion," "tibetan bong," "bagpipes"[5]
Nature	Any sound produced by nature. Excludes water-related sources.	"Birds," "cicada," "wind"[6]
Traffic	Any sound produced by a part or all of a vehicle propelled by motor.	"Cars," "traffic," "planes"[7]
Water	Any sound produced by the movement of water, regardless of the actual source of the sound.	"Mist sound," "sprinkler," "the mist"[8]
Other	All remaining responses	"Ambient sounds"[9]

Definitions are provided in this table, along with examples taken from the questionnaires completed by respondents. Some of the examples are originally in French for which translations are provided here. The originals can be found in the endnotes.

TABLE 4 | Number of questionnaires completed by respondents for each design showing the mister status and gender and age distribution.

Design	Mister		Gender		Age			
	Off	On	Women	Men	Min	Max	Mean	Median
1	25	75	64	32	18	77	39	34
2	15	61	33	39	18	73	35	30
3	70	28	57	40	19	84	38	33
Total	110	164	154	111	–	–	–	–

were adjusted using the Holm correction. This analysis was performed separately using two binary variables. In the first analysis, the variable was the status of the mister. In the second analysis, the variable was whether the respondent mentioned hearing water sounds (e.g., "water fountain").

In order to better understand the effect of the misters on the soundscape scales, we performed analyses using the data on respondent location and the sound sources they mentioned. As the mister is only lightly audible, we first investigated the possibility that the Design 2 misters located in the quiet side of the space, did not have the same significant effect on soundscape scales as in the noisy side of the space. To test this hypothesis, we performed ANCOVA tests on the chaotic and loudness scales using the location of the respondent and the status of the misters as independent variables, with temperature as a covariate. Given that only responses from Design 2 were used, Design was not included as an independent variable in these tests.

We also wanted to determine whether the mister had the effect of displacing traffic sounds (e.g., cars, buses). We performed a logistic regression of mentions of traffic sounds, using the design and status of the misters as independent variables. Logistic regression indicates the log odds of the occurrence of a binary outcome, as is the case of the mention (or not) of traffic sounds by questionnaire respondents.

All statistical analyses were performed using R version 3.6.2 (R Core Team, 2020).

4. RESULTS

In general, respondents rated the sound environments of all three designs as pleasant and appropriate for their activity, and found that these environments provided an opportunity for a break. Regardless of the condition, the average response for all three scales was above the mid-point (i.e., agree/strongly agree). As well, they rated monotonous and chaotic below the mid-point (i.e., disagree/strongly disagree). The respondents were more divided on ratings of vibrant, calm, eventful and loudness. **Figure 3** shows the complete distribution of the soundscape ratings along the Likert-scale (i.e., strongly disagree, disagree, neutral, agree, strongly agree) provided by respondents, broken down by design. Moreover, the differences between the soundscape ratings for each design were relatively small (all <0.45 on a 5-point scale), none of which reached statistical significance.

[3]"voix des personnes," "skateboards."
[4]"Air climatisé du supermarché."
[5]"Guitare," "percussions douces," "bong tibetain," "cornemuses."
[6]"Oiseaux," "cigale," "vent."
[7]"Avions."
[8]"La bruine."
[9]"Bruits ambiants."

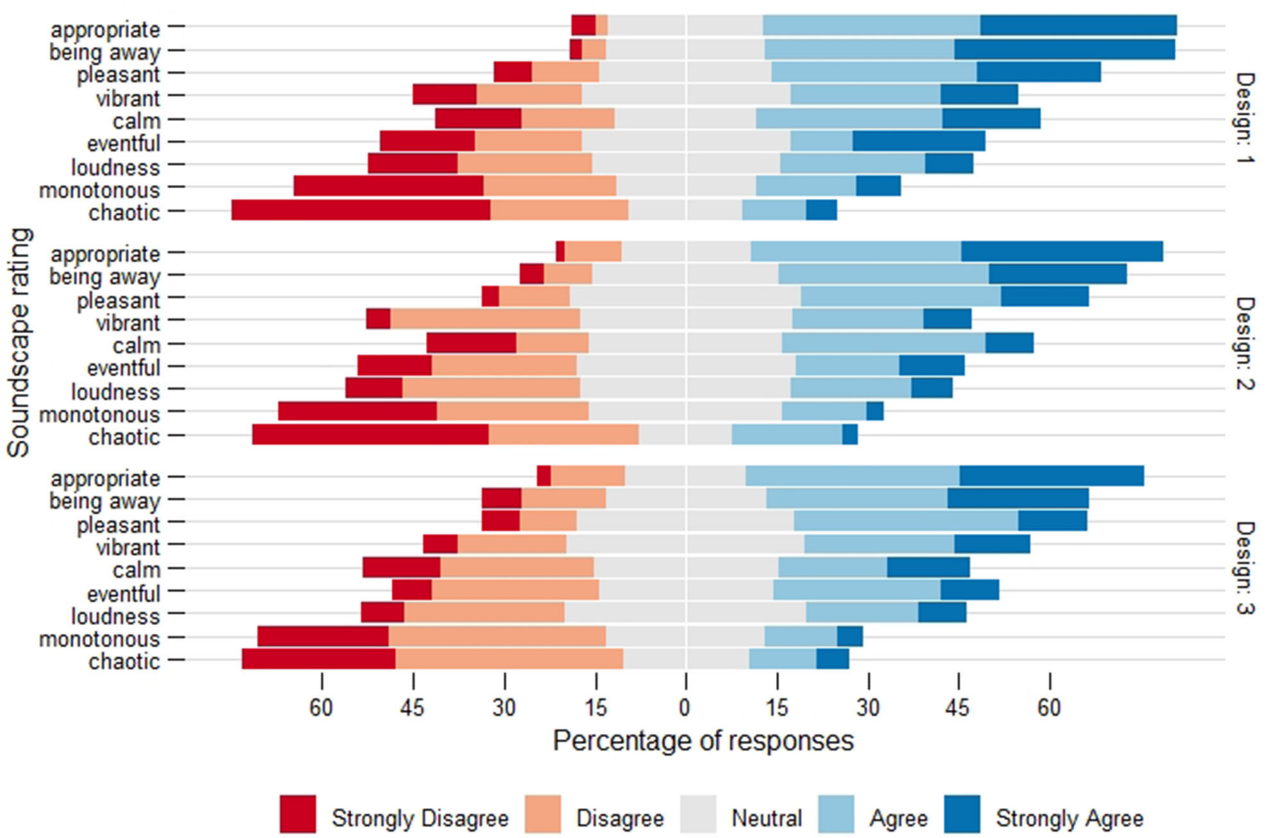


FIGURE 3 | Distribution of the Likert-scale soundscape ratings for each design. Respondents generally found the sound environment was appropriate for activity, provided an opportunity for a break and was pleasant. As well, they disagreed that the sound environment was monotonous and chaotic. No significant differences were found across designs.

TABLE 5 | Results of the MANCOVA test including the status of the water mister.

Variable	Df	Pillai	Approx F	Num Df	Den Df	Pr (>F)	
Temperature	1	0.0804	2.517	9	259	0.009	**
Design	2	0.061	0.908	18	520	0.569	
Mister	1	0.013	0.373	9	259	0.947	
Design × Mister	2	0.121	1.861	18	520	0.019	*
Residuals	267	NA	NA	NA	NA	NA	

Shows that the interaction effect of design and mister are statistically significant at the $p < 0.05$ level.
$^{**}p < 0.01$; $^{*}p < 0.05$.

4.1. Misting Water Feature

The interaction of the mister and the design had an overall effect on respondents' soundscape evaluations according to the MANCOVA of all scales against the six conditions (see **Table 5** for full results). The test also indicated that temperature alone had a significant effect on the soundscape ratings. However, temperature was included as a covariate in the analysis because it could not be controlled through experimental design, and therefore significant results from this variable are not discussed further. No other direct effect was found to be statistically significant, either from the mister or the design. Thus, after controlling for the effect of the daily temperature, the mister had a significant effect on soundscape evaluations that changed depending on the design of the space.

Looking at the individual soundscape ratings, we observed a significant interaction effect of mister status and design on the chaotic and loudness scales using a factorial ANCOVA. With the exception of temperature, no other significant effects were found (**Table 6** shows the results for the chaotic and loudness scales only).

TABLE 6 | Results from the factorial ANCOVA for the chaotic and loudness scales.

Variable	Term	DF	Sum Sq.	Mean Sq.	*F*-Stat	Pr (>F)	Adj. Pr (>F)	
Chaotic	Temperature	1	10.94	10.94	8.70	0.003	0.03	*
	Design	2	2.65	1.33	1.05	0.35	1.00	
	Mister	1	0.49	0.49	0.39	0.53	1.00	
	Design × Mister	2	17.18	8.59	6.83	0.001	0.01	*
	Residuals	267	335.75	1.26	NA	NA	NA	
Loudness	Temperature	1	2.76	2.76	2.44	0.12	0.72	
	Design	2	1.03	0.514	0.45	0.64	1.00	
	Mister	1	2.60	2.60	2.29	0.13	1.00	
	Design × Mister	2	15.38	7.69	6.79	0.001	0.01	*
	Residuals	267	302.16	1.13	NA	NA	NA	

There was a significant interaction effect for design and mister.
**p < 0.05.*

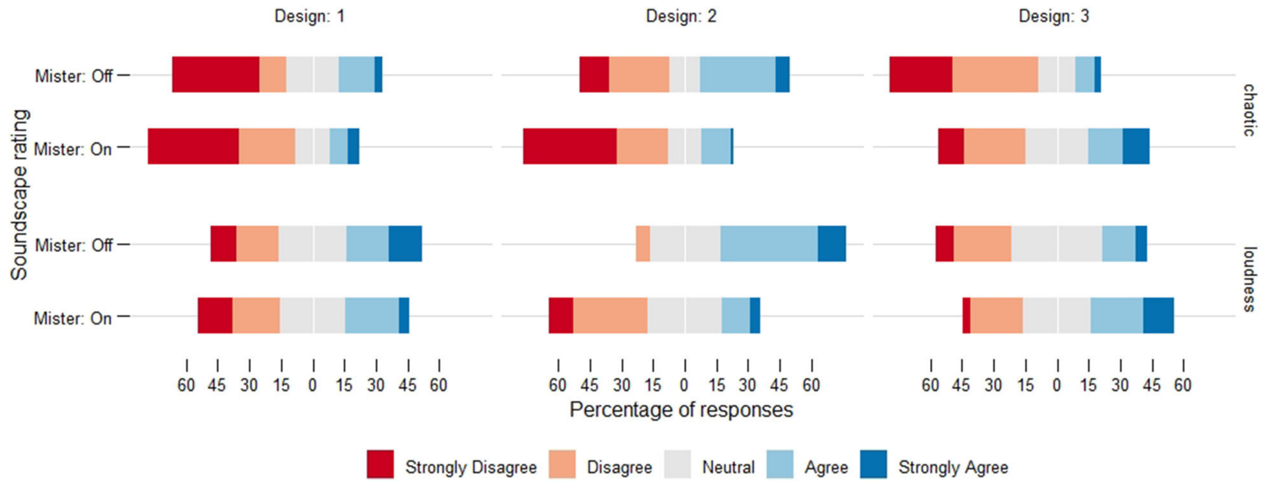


FIGURE 4 | Distribution of the Likert-scale responses for chaotic and loudness only. Each subplot represents a single condition (design and status of the mister).

During Designs 1 and 2, the mean values for both chaotic and loudness ratings were lower when the mister was on, with a larger percentage of respondents disagreed with both ratings (i.e., responded disagree/strongly disagree). The exact opposite occurred during Design 3: the mean values for chaotic and loudness were higher when the mister was on and a higher percentage agreed with both ratings (i.e., responded agree/strongly agree) (see **Figure 4** for the distribution of responses and **Figure 5** for the mean values).

4.2. Water as Sound Source

In all, water was mentioned by 66 respondents (representing 40% of all respondents when the fountain was on). All mentions of water sounds occurred when the mister was active. Across the three Designs, respondents mentioned water sounds more frequently during Designs 1 and 2 than in Design 3 (see **Table 7**). A chi-squared test confirms that this difference is statistically significant ($\chi^2 = 13.29$, $df = 2$, p-value $= 0.001$).

Further analysis of the mentions of different sound sources suggests that the addition of the mister did not remove negative sounds (e.g., traffic sounds) through energetic masking. The average number of sound sources mentioned by respondents increased by 0.8 when one of those sounds was water, as indicated in **Table 7**, which shows the average number of sounds mentioned for those who did and for those who did not mention water sounds.

Respondents who mentioned water sounds rated the soundscape scales differently than those respondents who did not mention water sounds, even after accounting for temperature. This was confirmed by a 3 (design) × 2 (water mentioned) factorial MANCOVA investigating all nine soundscape scales as dependent variables. This was a main effect and was therefore independent of design of the space (see **Table 8** for the complete results of the MANCOVA test).

Analysing each of the soundscape scales individually through a factorial ANCOVA test found a marginally significant result for the calmness scale (**df** $= 1$, $F = 9.91$, $p = 0.08$). Across all designs, the sound environment was rated as calmer when respondents mentioned water sounds than when they did not mention water sounds (see **Figure 6** for the mean values).

4.3. Effect of Mister as a Function of Location

The misters contribute only a lightly audible misting sound to the space. Given the arrangement of the misters on the quiet side of the space for Design 2, the misting sound was difficult, if not always impossible, to hear on the noisy side. This is confirmed by the proportion of those who mentioned water sounds in the noisy side of the space (1 out of 45 respondents) vs. those in the quiet side of the space (15 out of 31 respondents). Despite only one respondent making a specific mention of water sounds, there was no significant effect of location on the chaotic and loudness ratings, either as main or interaction effects. This was confirmed by the 2 (location: noisy vs. quiet) × 2 (mister status) ANCOVA tests using the chaotic and loudness scales as dependent variables.

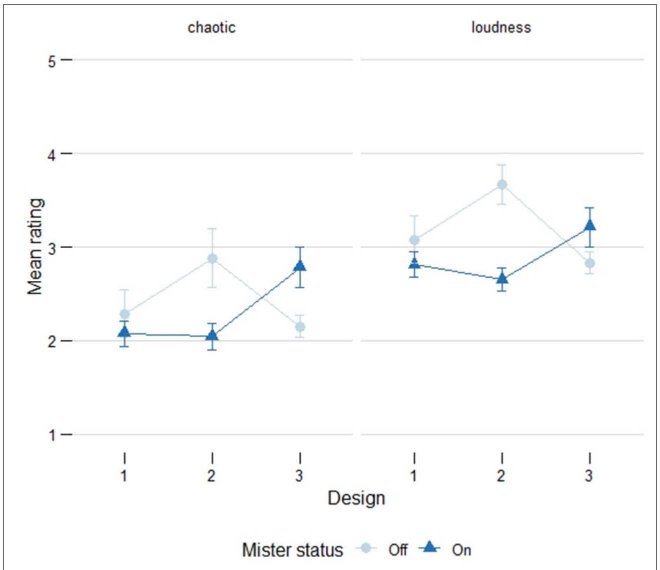


FIGURE 5 | Mean values for the chaotic and loudness ratings. The x-axis indicates the design and the lines compare the status of the mister (off/on). Errorbars indicate the standard error. The mean rating for chaotic and loudness scales was lower when the mister was on during Designs 1 and 2, but they were higher for Design 3, indicating that the mister did not have a consistent effect across designs.

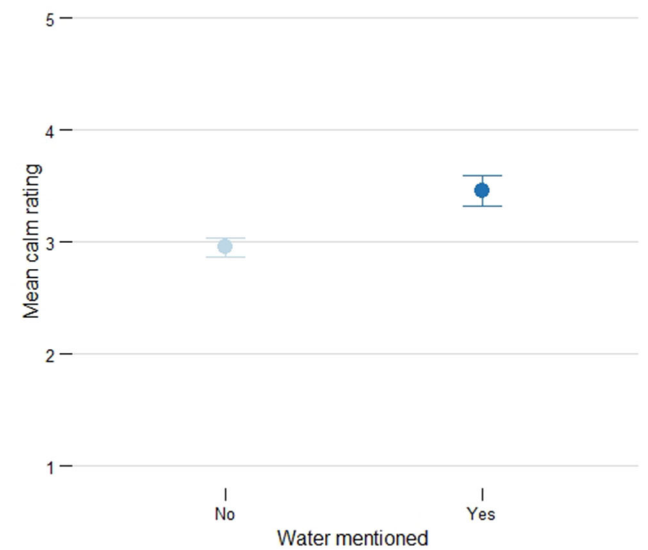


FIGURE 6 | The mean calmness rating increases when the respondent mentions water sounds.

TABLE 7 | Comparison of the mean number of sources mentioned when water is one of them and when it is not (standard deviation in italics).

Design	Count of respondents		Mean # of sources *(SD)*	
	Water mentioned		Water mentioned	
	No	Yes	No	Yes
1	64	36	3.1 *(1.7)*	3.9 *(1.2)*
2	60	16	3.6 *(1.7)*	4.0 *(1.2)*
3	84	14	3.4 *(1.9)*	4.3 *(1.3)*

Most respondents did not mention water sounds, but the proportion was higher in Design 3.

TABLE 8 | Results of the MANCOVA test including a factor for whether water was mentioned by the respondent.

	Df	Pillai	Approx F	Num Df	Den Df	Pr (>F)	
Temperature	1	0.080	2.495	9	259	0.009	**
Design	2	0.060	0.900	18	520	0.579	
Water mentioned	1	0.063	1.934	9	259	0.048	*
Design × Water mentioned	2	0.089	1.348	18	520	0.153	
Residuals	267	NA	NA	NA	NA	NA	

When respondents mentioned sources of water sounds, their responses were significantly different.
$^{**}p < 0.01$; $^{*}p < 0.05$.

4.4. Logistic Regression of Traffic Mentions Against Mister

The likelihood of a respondent mentioning traffic sounds was not significantly affected by either the status of the mister or the design of the space according to a logistic regression test. There was also no interaction effect between the two variables (mister status and design). Indeed, across all six conditions (3 design $\times$ 2 mister) 86% or more of the respondents indicated hearing traffic sounds.

5. DISCUSSION

Despite the heavy volume of vehicular traffic that was audible throughout the space, the sound environment provided an opportunity for restoration to users. Support for this argument can be seen in the extent to which respondents agreed that the sound environment was appropriate for their activity, was pleasant and provided a break across all three Design conditions. The apparent paradox of a potentially restorative space dominated by traffic noise reinforces the importance of considering more than just decibel levels. Indeed, it supports the argument that sound can be a resource as put forward by the soundscape approach (Schulte-Fortkamp et al., 2007). In particular, the sound environment of this urban space was a resource that afforded respondents a break in their daily routine, given that, on average, they rated this scale above the mid-point across all design conditions. This extends other studies that have looked at restoration along the sonic dimension (Payne and Guastavino, 2018; Steele et al., 2019; Senese et al., 2020).

5.1. Effectiveness of a Mister in a Small Urban Public Space

RQ1: Can small water features that are lightly audible in a outdoor urban public space have an observable effect on soundscape in-situ assessments?

The mister had no direct significant effect on any of the SSQP scales we used. Though they were studying a larger fountain, Axelsson et al. (2010) similarly found that the water feature did not significantly affect soundscape ratings. This contradicts laboratory-based studies where significant effects were found (Jeon et al., 2012; Galbrun and Ali, 2013; Ekman et al., 2015; Hong et al., 2020a; Senese et al., 2020) and suggests that the context of the urban public space plays an important role in soundscape assessments. The use of headphones and other laboratory equipment could focus the respondent's attention toward the added water feature, making it artificially more effective (Skoda et al., 2014). However, when in the context of an urban public space, the respondent's attention is more scattered, which could reduce the effectiveness of a water feature on soundscape assessments (Skoda et al., 2014). Moreover, as Axelsson et al. (2010) suggest, improving soundscape quality is not as simple as adding a water feature. This appears to also be true for small misting jets.

As well, the mister did not significantly affect the way respondents rated the being-away scale. While this suggests that the mister does not add to the affordance for restoration offered by a sound environment under these conditions, this may be related to the use of a single scale to represent restoration. This contrasts with previous research that reports that water features support psychological restoration along the being-away and fascination scales (Senese et al., 2020). The use of a single scale for restoration in our study is an important limitation that will be addressed in future iterations of our questionnaire. A significant effect may have been found if other ART scales were used to reflect fascination, extent and compatibility. For example, the mister in Design 1 was a focal point for the space that draws in the attention of its users, suggesting that fascination is an appropriate scale to use.

5.2. Configuration of the Mister Within the Space

RQ2: Can the measured effects of a small mister change if it is deployed in different configurations within the same outdoor urban public space?

When we consider the effect of the fountain in the context of a specific Design, we note that there is a significant effect. Thus, while a mister on its own is not universally effective at improving soundscape assessments, its integration can be beneficial or it can have the opposite of the intended effect.

We found that the mister had a significant lowering effect on the perceived chaoticness and loudness of the space in Designs 1 and 2. These two findings are in agreement with the laboratory findings cited in the literature review that smaller water features can have desirable effects on soundscape ratings in a pocket park that is dominated by road traffic (Galbrun and Ali, 2013; Ekman et al., 2015). In particular, it aligns with the findings of lower perceived loudness (see De Coensel et al., 2011; Hong et al., 2020a). It is also consistent with findings that upward jets are effective at improving soundscape ratings (Galbrun and Ali, 2013). However, in Design 3, the mister is associated with increased ratings for chaotic and loudness.

5.3. Broader Discussion

It is unclear why the mister decreased chaotic and loudness ratings in Designs 1 and 2, while increasing them in Design 3. The literature has often promoted energetic masking as a strategy to deal with unwanted road traffic noise, though this is not possible with smaller water features because they do not generate the necessary low-frequency content (Galbrun and Ali, 2013), and we assume were not sufficiently loud for the task of energetic masking either. We know from the sound sources mentioned by respondents that traffic noise was heard (and named) even when the mister was on, effectively ruling out the possibility that the mister reduced loudness through an energetic masking of traffic noise.

Another possibility is that the sound of the mister is generally considered to be pleasant and desirable and adding such desirable sounds can positively impact soundscape ratings (De Coensel et al., 2011). Moreover, this is consistent with the finding that natural streams and upward pointing jets are preferred over other types of fountains (Galbrun and Ali, 2013). In this scenario, the audible informational content provided by the misters in all 3 designs acts on the perceptions of the users of the space.

Informational masking could play a role in reducing chaotic and loudness ratings, though further research is required to validate this possibility. Moreover, this does not account for the change in ratings on the noisy side of the space during the second design where the mister was mostly not audible and was mentioned only once (during a weekend when there is reduced road traffic). While it is possible that respondents heard the mister but did not mention it as a sound source, that is unlikely given how few respondents mentioned the water sounds.

The soundscape approach emphasizes that expectation and perception play an important role in the person's context when they are assessing their sound environment (Dubois et al., 2004; Schulte-Fortkamp et al., 2007). Given that the space is small and that the mister is visible from every angle, it is possible that the mister affected respondents' expectations of the space which, in turn, affected the soundscape ratings. In Designs 1 and 2, the misters were configured in the space so that they would complement other activities (e.g., reading, eating). As such, the presence of the misters attracted people to the space, shifting the emphasis of the soundscape from traffic noise toward human activity. This possible mechanism could also explain why the same misting jets in the same urban space, but with a different configuration, actually increased chaotic and loudness ratings: the water spewed out onto gravel intended for human activity, making the area muddy and unusable. In the context of Design 3, however, the soundscape became less filled with dynamic human activity, and instead more chaotic and traffic-related. As such, this study provides tentative evidence that soundscape ratings are affected by the expectations about the activities that can take place in a space.

The varying number of mister locations within the space is a limitation of this study. There are challenges imposed on *in-situ* research design in spaces where the configuration itself changes. In our case, Design 1 featured a single mister in the center of the space, while Design 2 had two misters roughly 2 m apart and Design 3 had two misters that were 10 m apart. Furthermore, the Design 1 mister was the prominent focal point of the space, which further contrasts with the misters in Designs 2 and 3. Finally, the misters were not off for exactly 50% of each Design condition, introducing a potential source of bias. It is unclear what impact this had on the soundscape assessments made by the respondents, and there is no discernible trend based on either the number of misters or the distance between them.

A further limitation of this study is the difficulty in tracking the activity of each respondent and whether different types of activities would be more suited toward engaging with water features. The questionnaire contained an open-ended question asking respondents "What brings you here today?" The question was received ambiguously by respondents who either stated what they were doing in the space (e.g., "relaxing," "on my lunch break") or why they were in the vicinity of the space (e.g., "tourism," "getting an ice cream"). Thus, we could not fully explore the role activity played in creating a rich context as laid out by the ISO definition of a soundscape, in which person, place, and activity are interrelated (International Standards Organization, 2014). This is an avenue for future research, as more studies are needed on the effect of activity on soundscape assessments, especially in the case of pocket parks. As well, future contributions are necessary to establish a standardized methodology for collecting activity-related data during soundscape evaluations.

That said, this study confirms that questionnaires can be effectively used in a quasi-experimental research design to evaluate the impact of a water feature on soundscape assessments. Through a combination of SSQP (Axelsson et al., 2010), restorativeness (Payne and Guastavino, 2018) and loudness scales, we were able to capture soundscape evaluations that largely agree with existing literature though add nuance (and some further questions) to this growing body of knowledge. Standardized questionnaires should include more scales from the PRSS to represent different components of psychological restoration.

6. CONCLUSION

This study extends the *in-situ* research on the effect of water features to include lightly-audible misters in outdoor, urban environments. Our analysis shows that in some, but not all conditions, adding a mister can enhance the soundscape. In the context of a small urban space located next to a busy street with heavy pedestrian and vehicular traffic, these changes to the scales can be considered as positive and desirable effects. Showing that a small mister can enhance the soundscape is an important finding given that large fountains are often not possible in small pocket parks. First, the size of a large fountain may crowd out other activities. Second, its cost is often prohibitively high.

Further research is required on the ecological validity of lab-based research involving misters and small water features in small public spaces, given the conflicting findings between lab and *in-situ* research. Laboratory settings can cause the participant to focus on the sound environment in a way that is not possible in multimodal environments. Moreover, in a public space, the user is engaged in an activity that contributes to their soundscape assessment. This is not to say that water features have no effect if users of a space cannot hear them. Instead, the misters might provide visual and experiential appeal even without the auditory component, which affects respondents expectations of the space. It does suggest that the mechanism by which misters affect soundscape ratings in context is not straightforward and needs to be better understood.

This study suggests potential design considerations when using a mister in a public space. First, the dimensions of both the mister and the misting water should be chosen so that it attracts users into the space. While it is not possible to rule out the role of masking, designers can use misters to provide sounds that have semantic meaning and positive associations for the users of the space. Second, the mister should be configured within the space in such a way that does not preclude the likely activities. Therefore, it is important to have a good sense of what users want from the space. The design of the space and the configuration of the mister within it together should clearly indicate to the user the activities afforded by the space. These considerations inform the context that is highlighted by the soundscape approach (i.e., the relationship between person, place and activity) and in turn can be used to improve the soundscape of a small public space.

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available. Requests to access these datasets should be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by McGill's Research Ethics Board (REB-2), McGill University (REB-55-0615). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

CT and DS collected and analyzed the data under the guidance of CG. All authors designed the experiment and contributed to the writing of the manuscript.

FUNDING

This research was supported by a grant from Canada's Social Sciences and Humanities Research Council (#430-2016-01198), awarded to CG, for the research partnership Sounds in the City (www.sounds-in-the-city.org) and for which ethics approval was obtained.

ACKNOWLEDGMENTS

The authors thank Cynthia Tarlao, Mariana Mejia-Ahrens, and Sarah Bogdanovitch for their assistance with data collection and acoustic measurements. CIRMMT provided the equipment used to take the necessary acoustic measurements. We thank our collaborators at the Ville de Montréal and the design firm Castor et Pollux for the collaboration.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.570797/full#supplementary-material

REFERENCES

Aletta, F., Oberman, T., and Kang, J. (2016). Associations between positive health-related effects and soundscapes perceptual constructs: a systematic review. *Int. J. Environ. Res. Public Health* 15:2392. doi: 10.3390/ijerph15112392

Alvarsson, J. J., Wiens, S., and Nilsson, M. E. (2010). Stress recovery during exposure to nature sound and environmental noise. *Int. J. Environ. Res. Public Health* 7, 1036–1046. doi: 10.3390/ijerph7031036

Axelsson, Ö., Nilsson, M. E., and Berglund, B. (2010). A principal components model of soundscape perception. *J. Acoust. Soc. Am.* 128, 2836–2846. doi: 10.1121/1.3493436

Axelsson, Ö., Nilsson, M. E., Hellström, B., and Lundén, P. (2014). A field experiment on the impact of sounds from a jet-and-basin fountain on soundscape quality in an urban park. *Landsc. Urban Plan.* 123, 49–60. doi: 10.1016/j.landurbplan.2013.12.005

Berto, R. (2005). Exposure to restorative environments helps restore attentional capacity. *J. Environ. Psychol.* 25:2005. doi: 10.1016/j.jenvp.2005.07.001

Brown, A. L. (2010). Soundscapes and environmental noise management. *Noise Control Eng. J.* 58, 493–500. doi: 10.1515/noise-2017-0007

De Coensel, B., Vanwetswinkel, S., and Botteldooren, D. (2011). Effects of natural sounds on the perception of road traffic noise. *J. Acoust. Soc. Am.* 129, EL148–EL153. doi: 10.1121/1.3567073

Dubois, D., Guastavino, C., Maffiolo, V., and Raimbault, M. (2004). A cognitive approach to soundscape research. *J. Acoust. Soc. Am.* 115:2495. doi: 10.1121/1.426115

Ekman, R., Lundén, P., and Nilsson, M. (2015). Similarity and pleasantness assessments of water-fountain sounds recorded in urban public spaces. *J. Acoust. Soc. Am.* 138, 3043–3052. doi: 10.1121/1.4934956

Engel, M. S., Fiebig, A., Pfaffenbach, C., and Fels, J. (2018). A review of socio-acoustic surveys for soundscape studies. *Curr. Pollut. Rep.* 4, 220–239. doi: 10.1007/s40726-018-0094-8

Galbrun, L., and Ali, T. T. (2013). Acoustical and perceptual assessment of water sounds and their use over road traffic noise. *J. Acoust. Soc. Am.* 133, 227–237. doi: 10.1121/1.4770242

Hartig, T., Korpela, K. M., Evans, G. W., and Gärling, T. (1997). A measure of restorative quality in environments. *Scand. Housing Plan. Res.* 14, 175–194. doi: 10.1080/02815739708730435

Hong, J. Y., Lam, B., Ong, Z.-T., Ooi, K., Gan, W.-S., Kang, J., et al. (2020a). The effects of spatial separations between water sound and traffic noise sources on soundscape assessment. *Build. Environ.* 167:106423. doi: 10.1016/j.buildenv.2019.106423

Hong, J. Y., Ong, Z.-T., Lam, B., Ooi, K., Gan, W.-S., Kang, J., et al. (2020b). Effects of adding natural sounds to urban noises on the perceived loudness of noise and soundscape quality. *Sci. Total Environ.* 711:134571. doi: 10.1016/j.scitotenv.2019.134571

International Standards Organization (2014). *ISO 12913-1:2014 Acoustics—Soundscape—Part 1: Definition and Conceptual Framework*. Available online at: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/05/21/52161.html (accessed March 1, 2020).

Jansson, M., and Persson, B. (2010). Playground planning and management: an evaluation of standard-influenced provision through user needs. *Urban For. Urban Green.* 9, 33–42. doi: 10.1016/j.ufug.2009.10.003

Jeon, J. Y., Lee, P. J., You, J., and Kang, J. (2012). Acoustical characteristics of water sounds for soundscape enhancement in urban open spaces. *J. Acoust. Soc. Am.* 131, 2101–2109. doi: 10.1121/1.3681938

Kaplan, R., and Kaplan, S. (1989). *The Experience of Nature: A Psychological Perspective*. Cambridge: Cambridge University Press.

Kaplan, S. (1995). The restorative benefits of nature: toward an integrative framework. *J. Environ. Psychol.* 15, 169–182.

Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing, 6th Edn.* Bingley: Emerald.

Nilsson, K., Sangster, M., Gallis, C., Hartig, T., De Vries, S., Seeland, K., and Schipperijn, J. (2010). *Forests, Trees and Human Health*. New York, NY: Springer Science & Business Media.

Nordh, H., and Østby, K. (2013). Pocket parks for people–a study of park design and use. *Urban For. Urban Green.* 12, 12–17. doi: 10.1016/j.ufug.2012.11.003

Payne, S. R. (2008). Are perceived soundscapes within urban parks restorative. *J. Acoust. Soc. Am.* 123, 3809–3809. doi: 10.1121/1.2935525

Payne, S. R., and Guastavino, C. (2018). Exploring the validity of the perceived restorativeness soundscape scale: a psycholinguistic approach. *Front. Psychol.* 9:2224. doi: 10.3389/fpsyg.2018.02224

Peschardt, K. K., and Stigsdotter, U. K. (2013). Associations between park characteristics and perceived restorativeness of small public urban green spaces. *Landsc. Urban Plan.* 112, 26–39. doi: 10.1016/j.landurbplan.2012.12.013

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

Refshauge, A. D., Stigsdotter, U. K., and Cosco, N. G. (2012). Adults' motivation for bringing their children to park playgrounds. *Urban For. Urban Green.* 11, 396–405. doi: 10.1016/j.ufug.2012.06.002

Schulte-Fortkamp, B., Brooks, B. M., and Bray, W. R. (2007). Soundscape: an approach to rely on human perception and expertise in the post-modern community noise era. *Acoust. Today* 3, 7–15. doi: 10.1121/1.2961148

Senese, V. P., Pascale, A., Maffei, L., Cioffi, F., Sergi, I., Gnisci, A., et al. (2020). "The influence of personality traits on the measure of restorativeness in an urban park: a multisensory immersive virtual reality study," in *Neural Approaches to Dynamics of Signal Exchanges Smart Innovation, Systems and Technologies*, eds A. Esposito, M. Faundez-Zanuy, F. C. Morabito, and E. Pasero (Singapore: Springer), 347–357. doi: 10.1007/978-981-13-8950-4_31

Skoda, S., Steffens, J., and Becker-Schweitzer, J. (2014). "Road traffic noise annoyance in domestic environments can be reduced by water sounds," in *Proceedings of Forum Acusticum* (Krakow).

Steele, D., Bild, E., Tarlao, C., and Guastavino, C. (2019). Soundtracking the public space: outcomes of the Musikiosk soundscape intervention. *Int. J. Environ. Res. Public Health* 16:1865. doi: 10.3390/ijerph16101865

Tarlao, C., Fernandez, P., and Guastavino, C. (2016). "Comparing soundscape evaluations in French and English across three studies in Montreal," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (Hamburg: Institute of Noise Control Engineering), 4701–4707.

Tarlao, C., Steele, D., and Guastavino, C. (2019). "Investigating factors influencing soundscape evaluations across multiple urban spaces in Montreal," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (Madrid: Institute of Noise Control Engineering), 6888–6899.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 22 January 2021
doi: 10.3389/fpsyg.2020.602831

Introducing a Method for Intervals Correction on Multiple Likert Scales: A Case Study on an Urban Soundscape Data Collection Instrument

*Matteo Lionello, Francesco Aletta, Andrew Mitchell and Jian Kang**

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

OPEN ACCESS

Edited by:
Kimihiro Sakagami,
Kobe University, Japan

Reviewed by:
Gunnar Cerwén,
Swedish University of Agricultural Sciences, Sweden
Jooyoung Hong,
Singapore University of Technology and Design, Singapore

****Correspondence:***
Jian Kang
j.kang@ucl.ac.uk

Specialty section:
This article was submitted to Environmental Psychology, a section of the journal Frontiers in Psychology

Received: *04 September 2020*
Accepted: *17 December 2020*
Published: *22 January 2021*

Citation:
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. Front. Psychol. 11:602831. doi: 10.3389/fpsyg.2020.602831

Likert scales are useful for collecting data on attitudes and perceptions from large samples of people. In particular, they have become a well-established tool in soundscape studies for conducting *in situ* surveys to determine how people experience urban public spaces. However, it is still unclear whether the metrics of the scales are consistently interpreted during a typical assessment task. The current work aims at identifying some general trends in the interpretation of Likert scale metrics and introducing a procedure for the derivation of metric corrections by analyzing a case study dataset of 984 soundscape assessments across 11 urban locations in London. According to ISO/TS 12913-2:2018, soundscapes can be assessed through the scaling of 8 dimensions: pleasant, annoying, vibrant, monotonous, eventful, uneventful, calm, and chaotic. The hypothesis underlying this study is that a link exists between correlations across the percentage of assessments falling in each Likert scale category and a dilation/compression factor affecting the interpretation of the scales metric. The outcome of this metric correction value derivation is introduced for soundscape, and a new projection of the London soundscapes according to the corrected circumplex space is compared with the initial ISO circumplex space. The overall results show a general non-equidistant interpretation of the scales, particularly on the vibrant-monotonous direction. The implications of this correction have been demonstrated through a Linear Ridge Classifier task for predicting the London soundscape responses using objective acoustic parameters, which shows significant improvement when applied to the corrected data. The results suggest that the corrected values account for the non-equidistant interpretation of the Likert metrics, thereby allowing mathematical operations to be viable when applied to the data.

Keywords: multiple likert scales, ordinal against interval scales, likert scale correction, likert equidistance, urban soundscape, soundscape modeling, soundscape indices

1. INTRODUCTION

Likert scales (Likert, 1932) are commonly used in social sciences for the collection of attitudes and opinions. A Likert scale is composed of an odd number (typically 5 or 7) of ordered categories ranging from "strongly disagree" to "strongly agree" (or vice versa) with a "neutral" assessment being the midpoint category. Each point in the scale represents the degree to which the respondent agrees or disagrees with regard to a specific statement or construct, which is then typically associated with a value. There has been a long debate (Jamieson, 2005; Pell, 2005; Carifio and Perla, 2008) around whether or not the categories of a Likert scales can be interpreted by people as being equidistant. In their original conception, Likert scales are a sorted sequence of ranked categories where only nonparametric tools can be used. Inferring that the scales have an equidistant property between their categories allows the use of more powerful and precise parametric tools, and potentially mathematical operations (Adroher et al., 2018). This assumption is called an interval interpretation of the scales. It is indeed common practice for researchers to assume that participants will interpret the categories in the scales as equidistant (Lionello et al., 2020). The case of multiple scales mapped to a low-dimensional space expands the challenge of validating the equidistance property, both within each scale and between different scales of the same instrument, as in the case of the soundscape data collection protocol considered in this study.

Performing a scaling task with a Likert instrument essentially means mapping a perceptual space. Thus, trying to validate the equidistance property with a separate experiment would be challenging as it would imply mapping a different space through a potentially nonidentical task (see section 2.2). For this reason, any attempt at validating equidistance of Likert categories should be sought within datasets originating from the same scaling task (Lantz, 2013).

Soundscape, which is defined as the perceived sound environment by individuals and people in context (ISO, 2014), is going through a standardization process, especially for data collection instruments and corresponding analysis techniques. Assessment scales (e.g., Likert scales, Visual Analog scales, etc.) play an important role in the development of methods and tools for soundscape analysis (Fiebig and Herweg, 2017; Aletta et al., 2019; Lionello et al., 2020). One of the procedures currently used for soundscape assessments is the "Method A" described in the ISO/TS 12913-2:2018 (see section 2.1). This method makes use of Likert scales as its primary tool; while originally defined for "soundwalks" (i.e., assisted listening exercises on site) that are typically designed for 10–20 participants, the Method A could also be used for large-scale soundscape surveys on site, enabling the collection of data from, potentially, hundreds of public spaces users in a relatively short period of time. Several adjectives, which the model by Axelsson et al. (2010) assumes to be laying onto a vector space where their correlation is known (see section 2.1 and **Figure 1**), are presented to participants for them to indicate their degree of agreement or disagreement on whether each adjective is suitable to describe the soundscape they experience. These adjectives, which in the context of this study are referred to as "perceptual attributes," are defined to represent the dimensional components describing the decision process occurring in the quality evaluation of the soundscape experience by listeners. By assigning each category to a given number, certain assumptions introduced in section 2.1 allow researchers to mathematically collapse several scales into one or more values to describe the average assessment of the soundscape. The current study aims to understand the limits within which these operations can take place, and where correction factors can be placed in order to best report the abstract representation of urban soundscapes in the listener's mind.

In a previous study (Lionello et al., 2019), soundscape datasets were collected at different sites. Strong dependencies were found between the percentages of scores falling in three groups of Likert categories, i.e., "agreement," "disagreement," and "neutral," across different locations for different perceptual attributes. Nonetheless, some soundscapes datasets were found to show an asymmetric distribution across the mean and variances of their perceptual attributes and their average was observed to fall on a larger interval compared to the average of the corresponding opposite attribute. These findings encouraged a larger and more systematic investigation of the interpretation of the metric scales, showing that the typical inference of an equidistant property may occasionally be violated. In the current study, the previous analysis is extended to a larger dataset: the general goal is determining a procedure that is potentially applicable to other soundscape studies for the introduction of scale correction values. To the best of the authors' knowledge, this study represents the first attempt in soundscape literature to apply psychometric correction factors to soundscape assessment scales.

The correction values found are bound to the paradigm defined by the data collection framework; although the overall methodology would remain valid, a change in (even small) aspects of the assessment task (as detailed in section 5.6) would likely render the derived correction values themselves invalid. In the current case, the model is bound to the following points: five-point Likert scales where each point is labeled; *in situ* data collection; and the scale selection to be assessed and the target of the scales assessments, while the conditions for which the current methodology can be applied are as follows: multiple Likert scales laying on a known vector space; multiple sets of situations where surveys are collected; and a consistent sample of surveys ($N \approx 100$) assessed for each of these sets. At this stage, under certain hypothesis discussed in section 5.2, the correction factors are assumed to be invariant with respect to the sample of soundscapes currently reported. The study aims to address two main research questions: Are the relationships between Likert categories in each scale and between scales coherently understood by participants as expected from the equidistant property of Likert scales? If not, what corrections can be introduced to adjust them and to project soundscapes assessment that are more consistent with the participant's interpretation?

In section 2, a theoretical background related to soundscape data collection, analysis standards, and Likert scaling task will be presented and the main issues are identified. In sections 3 and 4, the protocol followed for the data collection is introduced and the results of the method applied to soundscapes are reported. In section 5, the correction factors for the investigated metrics and the limitations of the proposed framework are discussed.

2. THEORETICAL BACKGROUND FOR APPLICATION TO SOUNDSCAPE STUDIES

2.1. Introduction to the ISO 12913 Series on Soundscape

The International Organization for Standardization (ISO) has been working during the past decade on the ISO 12913 (Acoustics—Soundscape) series. This currently includes three parts: Part 1—definition and conceptual framework (ISO, 2014); Part 2—data collection and reporting requirements (ISO/TS, 2018); and Part 3—data analysis (ISO/TS, 2019). Part 1 is published as a full standard document, while the other two parts are published as technical specifications. Part 1 defines the "soundscape" as a perceptual construct, as opposed to the "acoustic environment," which is the physical phenomenon. Parts 2 and 3 are the "operational" documents where the instruments for data collection and analysis procedures are described. Part 2 provides three different options for gathering data on how people experience(d) acoustic environments (i.e., soundscape data), which include questionnaires to be used on-site (Method A or Method B), and narrative interviews protocols to be used off-site (Method C). In this study, we focus on Method A; this is adapted from the previously established Swedish Soundscape Quality Protocol (SSQP), emerging from the work on urban soundscapes by Axelsson and colleagues at the University of Stockholm during the years 2005–2010 (Axelsson et al., 2010). The perceptual attributes used for the soundscape assessment were defined in the context of a laboratory experiment; they were selected from the analysis of the principal components across 116 adjectives scaled by 100 students on 50 audio-recordings of urban locations in Sweden (Axelsson et al., 2010). The experiment validated the circumplex model of affect (Russell, 1980; Posner et al., 2005) for soundscape assessment tasks and the following 8 perceptual attributes were identified: "pleasant," "annoying," "vibrant," "monotonous," "calm," "chaotic," "eventful," and "uneventful." The dimensions corresponding to the perceptual attributes lay onto a bidimensional space described in **Figure 1** where pleasant and annoying are parallel to each other and orthogonal to eventful and uneventful, which are also parallel to each other. The other four perceptual attributes lay along the bisectors of the plan. Perceptual attributes that are parallel to each other can be gathered in four pairs: "pleasant-annoying," "calm-chaotic," "vibrant-monotonous," and "eventful-uneventful." The scaling of these eight perceptual attributes was included in the Method A of the ISO/TS 12913-2:2018, where the protocol requires the participants to listen to a given acoustic environment and then proposes the following task: "For each of the eight scales below, to what extent do you agree that the present surrounding sound environment is...," followed by eight perceptual attributes, each associated to a Likert scale. This instrument had effectively been used in soundscape studies for several years before it was included in the ISO/TS 12913-2:2018, and until the ISO/TS 12913-3:2019 document was published, there was no clear indication on how the data collected through this protocol should have been analyzed or indeed "represented." Simply plotting the mean scores of the participants' sample as individual values on the circumplex model was often considered a pragmatic approach (Aletta et al., 2017; Kang et al., 2018): using a spider plot would allow to visualize an average "soundscape profile" for a given acoustic environment (see **Figure 1**). However, this is not a particularly comprehensive representation, nor one that allows for easy and meaningful comparisons between soundscapes. Therefore, Part 3 of the ISO series offers further guidance; it provides that the 8 attributes should be projected onto the bidimensional circumplex model by computing an orthonormal projection (Kogan et al., 2016; Lindburg and Friberg, 2016; ISO/TS, 2019) onto the two main dimensions of the circumplex model, which from now on we will identify as "ISO Pleasantness" and "ISO Eventfulness" to distinguish them from the simple perceptual attributes. This process is schematized (**Figure 1**): it shows how the assessment of a soundscape derived from the eight attributes scored independently can be reported to a point (x = ISO Pleasant; y = ISO Eventful) on the ISO circumplex model. Such an orthonormal projection assumes the participants to interpret the categories of the single Likert scales as being equidistant, and the eight perceptual attributes to be related as per the circumplex model. By following this assumption, it is possible to match, for instance, disagreement of annoying with agreement of pleasant, and a neutral score of pleasant with a neutral score of annoying, and so on for all the paired perceptual attributes. Having a final pair of coordinates allows the soundscape to be pinned in the circumplex model in order to cluster agglomerations of soundscapes, to classify the soundscapes according to the perceptual attribute dictating the bisector of the quarter where they fall, and to calculate the distances between soundscapes and the distances from them and the axes. The introduction of redundancy in the scaling of all the eight perceptual attributes is supported by the idea that, during the scaling task, participants may focus their attention to different categories of sounds according to the valence of the sound source (Berglund et al., 2007). The scaling of all the eight perceptual attributes also introduces a higher resolution in the final projection and it could also be used as an exclusion criterion for those participants whose assessments fall too far from the overall statistics.

Although one could imagine the kind of soundscapes falling along the edge regions of each bisector (e.g., distant traffic noise for monotonous, sounds of urban parks for calm, festive alleys atmosphere for vibrant, street affected by loud traffic noise for chaotic, etc.), and so to gradually shift from one of these to another one, a potential problem is to understand what kind of soundscape location could be represented in the center of the circumplex model and what is the meaning of the distance from one point to the center. A second challenge is whether the model should be inscribed within a circle, as currently described in Part 3 of the ISO (ISO/TS, 2019), or rather inside a square and making so, for instance, the agreement of vibrant match the agreement of pleasant. Moreover, it is not possible to know if the dimensions maintain exact overlapping intervals and ranges between each other.

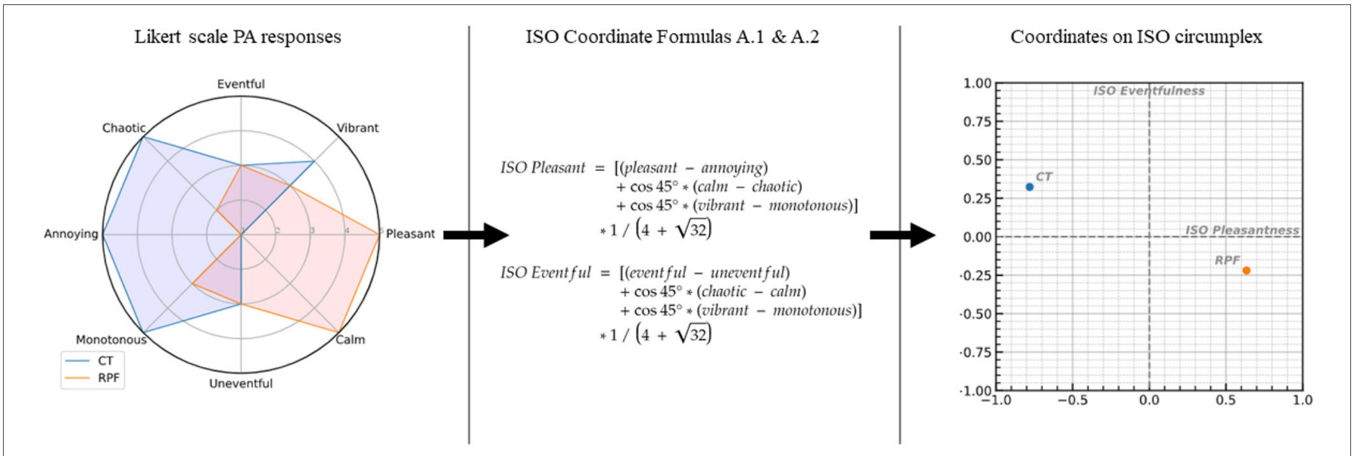


FIGURE 1 | The procedure for projecting the 8 perceptual attribute dimensions **(left)** onto the bidimensional circumplex model **(right)** according to the formulas given in ISO12913:3-2019 **(center)**. The data used are example single surveys drawn from Camden Town (CT) and Regents Park Fields (RPF). Each line of the radar plot shown on the left represents one perceptual attribute laid onto the bidimensional space, with paired attributes in opposing directions. The transformation is performed according to formulas A.1 and A.2 given in ISO12913:3-2019. The $*1/(4+\sqrt{32})$ term is included in order to scale the resulting coordinate values between −1 and +1.

2.2. Scaling Task and Equidistance of Likert Scale Categories

Some factors inherent in an *in situ* survey (such as ecological validity, participants' psychological state and attention, behavioral and routine context, and variety of population) introduce deviations in the *in situ* scaling compared to a laboratory setup (Rickards et al., 2012). The introduction of these deviations means the *in situ* scaling task does not represent an endomorphism within the space originally found from the principal component analysis performed in the laboratory experiments. The mapping of the N-dimensional abstract representation of the soundscape in the participant's mind to the 8-dimensional space potentially affects the interpretation of the Likert scales without maintaining the assumed equidistance property between the points of the original field (Maffiolo et al., 1999). In this way, the scaling task may result in the negative and positive poles of the Likert scale being unbalanced, collapsing one edge, dilating the distance between points, or omitting middle points. Furthermore, the assumption of endomorphism would not justify the need for scaling both negative and positive poles of each dimension (e.g., annoying and pleasant; vibrant and monotonous, etc.). Where individual participant behavior might rely on several demographic, social, psychological, and affective variables and not be an easy problem to solve, systematic common behaviors across the population are easier to detect by analyzing general trends over the scores.

2.3. A Note on Terminology

Throughout this paper, the following terms are used in order to describe the survey data as collected and after the scaling method is applied:

Likert scale: The assessment scale relative to one perceptual attribute submitted to participants, comprising ordered categories that range from "strongly disagree" on one pole to "strongly agree" on the other.

Likert categories: The labels applied to the ordered categories (i.e., "strongly disagree," "disagree," "neither agree, nor disagree," "agree," "strongly agree").

Likert scale metric: The geometrical function dictating the distances between the Likert points on one scale, ranging equal intervals between its points.

Likert value(s): The numerical value applied to each category (1–5, when considered as equidistant).

Rescaled metric: The new geometrical function dictating the distances between the Likert points of one scale, built to range perceptually equidistant intervals along different directions on the circumplex space.

Corrected value(s): The newly derived numerical values, based on the rescaled metric, to be applied to each Likert scale category.

ISO Pleasantness/Eventfulness or coordinates: The coordinate pair of values (x: Pleasant, y: Eventful) to place the response value on the bidimensional circumplex model.

Corrected ISO Pleasantness/Eventfulness or coordinates: The coordinate pair of response values on the circumplex model (x: Corrected Pleasant, y: Corrected Eventful) calculated through the rescaled metrics.

It has been noted that the term "metric" is inconsistently used and understood across fields and studies, where it may be interpreted as a statistic or index, as a synonym of "scale," or to distinguish between ordinal (non-metric) and interval (metric) data (Adroher et al., 2018). In this paper, its use is intended as its mathematical definition, as a distance function.

3. METHODOLOGY

3.1. Soundscape Data Collection Method

The data collection, currently used in some studies (Lionello et al., 2019; Aletta et al., 2020; Mitchell et al., 2020), followed the Soundscape Indices (SSID) Protocol (Mitchell et al., 2020), collecting *in situ* responses (soundscape assessment data) from users of public spaces at 11 different locations in London (UK) (**Figure 2**). For the soundscape-related questions, the SSID protocol is in turn based on Method A of the (ISO/TS, 2018). At each site, approximately 100 participants were asked to fill a questionnaire including the scaling of the eight perceptual attributes—i.e., pleasant, calm, uneventful, monotonous, annoying, chaotic, eventful, and vibrant—on a five-point Likert scale ranging from "strongly agree" (5) to "strongly disagree" (1) (see **Figure 2**). In order to reach the required amount of participants, surveys were collected during multiple sessions at the same location, trying to meet the same general context (e.g., time of the day, weather conditions, and social presence).

3.2. Participants

In addition to the soundscape-related questions, the SSID protocol also collected basic demographics about the participants. There is some evidence to suggest that personal characteristics such as age, gender, and educational level can influence a person's assessment of the soundscape (Yang and Kang, 2005; Xiao and Hilton, 2019) to a limited degree, however these factors have not been considered within this study. In total, the data collection included $N = 984$ respondents, comprising 52.9% female, 45.6% male, and 0.14% nonconforming or prefer-not-to-say, with a mean age of 34.7 years. Participants were required to be at least 18 years of age, but no maximum age limit was applied. The majority of the sample (57.5%) were full-time employed, with 3% unemployed, 7% retired, 32% student, and 6% other or rather-not-say. A plurality of respondents (36%) are university graduates, 1.9% have some high school, 15.7% are high school graduates, 12.9% have some college, 5% have some postgraduate work, and 23.2% have a postgraduate degree. According to data from Eurostat (Eurostat, 2020), the proportion of the Inner London working-age population who have completed university level or higher education was 66.8%, compared to 64.2% within this dataset, indicating a reasonable sampling of the local population. The self-identified ethnic composition was 70.3% white, 14.2% Asian/Asian British, 5.3% mixed/multiple ethnic groups, 2.7% Black/African/Caribbean/Black British, 1.9% Middle Eastern, 2% other ethnic group, and 2.6% rather not say, with 28.5% identifying as local, 27.8% tourist, 10.9% other, and 32.7% rather-not-say.

3.3. Data Collection Sites

Eleven urban public spaces were considered for data collection, with surveys occurring between March and October 2019 and including 30 total sessions. For each site (see **Supplementary Figure 1**), the initial goal was to collect 100 responses, meaning the data collection for a single site was often split over multiple sessions on successive days. A minimum number of 15 responses per session had been fixed to ensure consistency among the responses within a session. In most cases, due to incomplete questionnaires, restricted site access, and limited time, <100 responses per location were successfully selected for the final analysis. The total number of responses per site was as follows: Camden Town (CamdenTown: 94); Euston Tap (EustonTap: 98); Marchmont Community Garden (MarchmontGarden: 88); St Pancras Lock (PancrasLock: 90); Regent's Park Broadwalk (RegentsParkFields: 114); Regent's Park Japanese Garden (RegentsParkJapan: 90); Russell Square (RussellSq: 86); St Paul's Churchyard (StPaulsCross: 64); St Paul's Paternoster Row (StPaulsRow: 64); Tate Modern (TateModern: 100); Torrington Square (TorringtonSq: 96). More details about the sites can be found in Aletta et al. (2020).

The initial selection of investigated sites was driven by the need to include a reasonably varied sample of urban settings and contextual factors, including (but not limited to) urban morphology, architectural typology, dominant sound sources, amount of greenness, cultural/historical significance, and crowdedness. Due to the practicalities of performing large-scale *in situ* surveys (the most obvious of which is a minimum presence of members of the public to approach and invite for the survey), it was not possible to achieve a full spectrum of representative urban spaces types (e.g., surveying "semi-desert" public spaces is not possible if there are no people to approach). Consequently, the selected locations skew toward crowded urban squares, but do include a wide variety of greenness levels, visual openness, historical significance, and sound sources profiles. The resulting set of soundscape assessments therefore does not fully cover the soundscape circumplex space as defined by Axelsson et al. (2010), instead clustering toward the vibrant (i.e., positive pleasantness and positive eventfulness) quadrant. To some extent, this reflects an inherent challenge with conducting *in situ* data collection, as the accessible sites are limited by practical realities, a limitation which may only be possible to address in the future with further laboratory studies.

3.4. Data Processing

An overall flowchart of how the data have been processed and used across the whole study is shown in **Figure 3**. Despite a dataset amounting to ($N = 984$) records, because of the lack of a homogeneous distribution across the five Likert categories and the relatively small number of total locations, weak correlations were initially found between single response categories. Thus, in order to investigate the interval properties of the Likert scale metrics, the scores of the Likert scales were collapsed into three grouped categories: "agreement" that included "strongly agree" and "somewhat agree" (1–2); "disagreement" that included "somewhat disagree" and "strongly disagree" (4–5); and "neutral" that corresponded to "neither agree nor disagree" (3) scores. This grouping choice was motivated by the need for leveling the distribution of the original categories, and for augmenting the precision of both correlation and slope regression analysis (which is introduced in section 3.4.2). This approach has also been adopted in previous studies on soundscape modeling (Giannakopoulos et al., 2019; Lionello

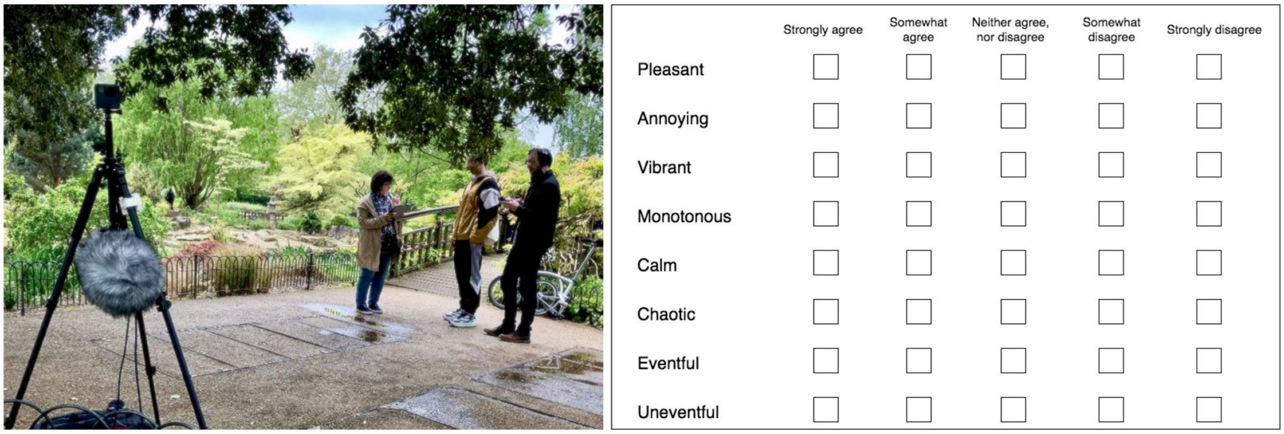


FIGURE 2 | (Left) Data collection on site, public space user fills the questionnaire while extra visual and acoustics measurements are taken. **(Right)** Perceptual attributes scaling part of the questionnaire used during the data collection.

et al., 2019). For each of the 11 locations, the percentage of scores (in terms of occurrences) falling in each group of these three new categories (agreement/neutral/disagreement) was calculated. Thus, 24 variables (3 categories * 8 perceptual attributes) for the 11 locations were considered.

3.4.1. Slope Coefficients to Introduce Correction Factors

In this section, the correlation of percentage of responses of grouped categories, found between the perceptual attributes across the 11 locations, were used to analyze a systematic behavior hidden in the way participants scaled their responses. The results of the analysis of these behaviors are then used in section 3.4.2 to calculate the new coefficients, which will be used in place of the original Likert scale values. By plotting the 11 soundscapes with respect to the ratio of scores falling between pairs of grouped categories, the soundscapes can range inside a triangular region bounded by $y = -x+1$ (in the boundary case of exact reciprocal proportions between the percentage of responses in the two grouped categories), $y = 0$, and $x = 0$ (in the boundary cases of no responses falling in one of the two categories examined, see **Figure 4**). Ideally, points would be expected to be randomly distributed within this region as the percentage of negative, positive, and neutral answers within each perceptual attribute are not expected to be correlated across the different soundscapes. Where these percentages are correlated, the points are not randomly distributed, and a regression slope coefficient can be derived from the pattern of points, as shown in **Figure 4**.

The procedure is based on the hypothesis that a dependency exists between the dilation or compression of the interval between two Likert categories and the regression slopes of percentage of answers falling in the respective grouped categories, which may either belong to the same scale or to two different scales. In this hypothesis, the interpretation of the dilation or compression of the Likert intervals is taken to be commonly shared across the participants, as it will be demonstrated in section 4.2.

To demonstrate this relationship, let us consider the following boundary cases in a scatter plot of soundscapes with respect to their percentages of agreements (dependent variable) and disagreements (independent variable) scores of an arbitrary perceptual attribute as seen in **Figure 4**. If the soundscapes lay onto a line with proportional coefficient equal to 0, the metric of the corresponding perceptual attribute would collapse across the disagreement poles ranging only between agreement and neutral values. In the case where the soundscapes lay onto a line with regression slope of -1, the percentage of agreement would be exactly reciprocal to the percentage of disagreement letting the number of the remaining neutral category scores be null. In this last case (case b in **Figure 4**), the neutral middle point score would be removed by making the whole scale range over 4 points instead of 5 points and by dilating the distances within both disagreement and agreement points.

By considering the previous examples, it is possible to advance the hypothesis that the angle β_j of the jth slope is linearly dependent on a dilation coefficient of the metric scale between the considered grouped Likert categories (see **Figure 4**). This allows us to introduce the following analysis of dilation across the intervals of the scale metrics.

3.4.2. Metric Scale Dilation Between Perceptual Attributes

By examining the correlation between perceptual attributes, it is possible to obtain a description of the dilation and compression ratios between the metric scales belonging to the respective perceptual attributes. The slope regression between score categories from different perceptual attributes is used to set the proportions of the intervals across all the metric scales. However, one more assumption must be taken regarding the centering of the metrics. The current study can identify dilation and compression of the intervals, but this procedure still cannot properly identify an eventual shift between the scales.

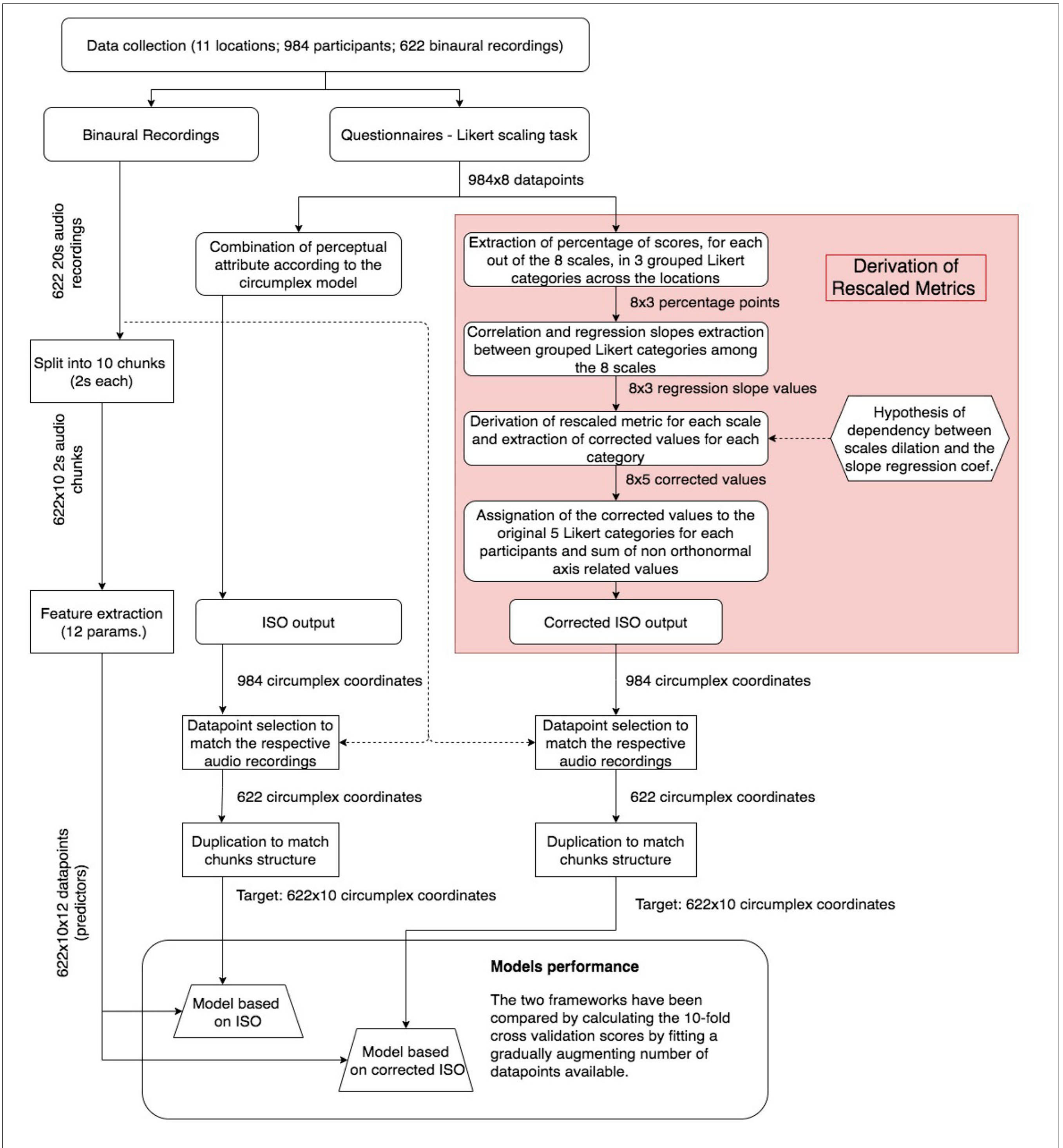


FIGURE 3 | Flow chart of the methodology and data processing presented in this study.

Following the relationship between regression slopes and dilation factors introduced in section 3.4.1, the regression slopes were selected to link all the grouped categories across all the perceptual attributes by combining those coefficients associated with the largest correlation across all their possible combinations. Once the system of equations relating all the grouped categories to each other across all the perceptual attributes is obtained, it is then possible to calculate the proportion values between the grouped categories with respect to one of them, which is arbitrarily fixed. Once the three values for $a_{i,0}$ (agree),

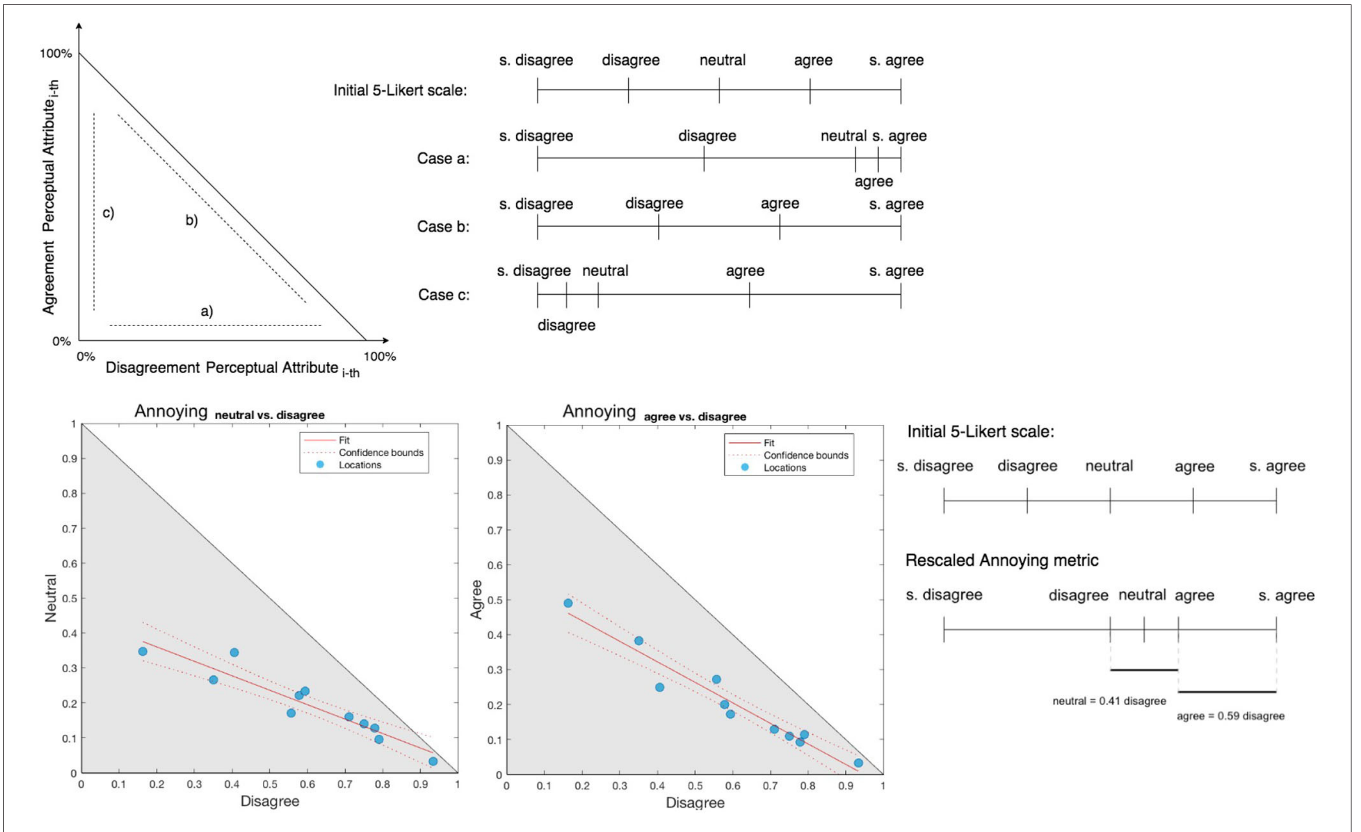


FIGURE 4 | (Top) Boundary cases of regression slopes of percentage of agreement and disagreement grouped categories for a given perceptual attribute across multiple soundscapes. **(Bottom)** Instance of application of the rescaled on the annoying metric. The two regression slopes are then used to rescale coherently the metrics. The rescaled version of annoying metric, reported in this example, is different from the final version presented later in **Figure 5** as, in this case, the metric is not enriched with information from the other scales.

$d_{i,0}$ (disagree), and $n_{i,0}$ (neutral) intervals for each perceptual attribute ($i = 1, 2, \ldots, 8$) are obtained, the barycenter of each rescaled metric is set as $b_i = \frac{a_{i,0}+d_{i,0}}{2}$. Then in the new metric scale the "strongly agree" point is set to $aa_i = a_{i,0} - b_i$ and the "strongly disagree" to $dd_i = d_{i,0} - b_i$. The neutral is modeled as $n_i = z_i b_i$, where $z_i \in \{-1, 1\}$ takes the sign according to the corresponding correlation between neutral and the other two grouped categories. The middle points "somewhat disagree" and "somewhat agree" are, respectively, found as $d_i = n_i - n_{i,0}/2$ and $a_i = n_i + n_{i,0}/2$. In three cases, namely vibrant, monotonous, and chaotic, no information was retrieved for $n_{i,0}$ as the neutral score percentages did not score relevant correlation (either $r < 0.7$ or $p > 0.05$) with any other score category across the perceptual attributes. For these, an equal range interval was assumed between neutral point and each edge: $d_i = \frac{n_i+dd_i}{2}$ and $a_i = \frac{n_i+aa_i}{2}$. The Likert scale categories can now be assigned to the new values $(aa_i, a_i, n_i, d_i, dd_i|_{i=1,\ldots,8})$ obtained for each category and each perceptual attribute. To calculate the new valence and arousal projection of one participant' assessments, all the new values assignation save for eventful and uneventful are summed together to calculate the valence, while annoying and pleasant are omitted and chaotic and calm changed of sign for the calculation of the arousal.

3.5. Application of the Rescaled Metrics

In order to test the usefulness of the correction factors, a classification model was built based on objective (psycho)acoustic metrics derived from 20-s binaural recording conducted while participants were responding to the survey. In order to compare the predictability of the two frameworks, the classification task was performed on both the ISO coordinates responses and with their corrected version. The models were designed to predict the individual assessments calculated as reported in section 3.4.2 and assigned to five categories (bins) defined by five equidistant intervals along the continuum of output values (number of samples falling in each bin in the corrected ISO coordinates pleasantness: [27, 144, 249, 134, 69], eventfulness: [33, 156, 304, 98, 32]). The same classification task was performed on the orthonormal projection by using the same predictors and samples (number of samples falling in each bin in the ISO coordinates pleasantness: [17, 107, 149, 240, 110]; eventfulness: [15, 139, 288, 145, 36]). The predictors, namely A-weighted sound level (LAeq), psychoacoustic loudness, sharpness, roughness, tonality, and speech interference level, were selected partially according to the results obtained across the soundscape modeling literature (Lionello et al., 2020) and calculated with the ArtemiS Suite software (v. 11.5, HEAD

TABLE 1 | Correlation and regression slopes between average scaling values of paired perceptual attributes across the 11 sites (e.g., annoying = – 0.87 × pleasant + 5.51).

	Correlation	Intercept	Slope
Annoying vs. pleasant	– 0.99	5.51	– 0.87
Monotonous vs. vibrant	– 0.53	4.81	– 0.71
Chaotic vs. calm	– 0.99	5.28	– 0.81
Uneventful vs. eventful	– 0.79	4.67	– 0.65

*See **Supplementary Figure 2** for the plotting.*

acoustics GmbH) (see **Supplementary Material**). The dataset used for this part partially overlaps what used to compute the correction values. Note that 622 binaural recordings taken during each filling of the questionnaires were cut to 20 s and split into 10 chunks 2 s long each. For each chunk, the mean and standard deviation of the previous listed acoustic parameters (see **Supplementary Table 2**) were calculated and used as input for the model. The models were fit for each of the four targets (ISO circumplex pleasantness and eventfulness and their corrected versions), multiple times with an increasing number of samples at each time to identify the convergence between training and validation data in the two systems of coordinates. Validation and training sets were composed of a total 622 × 10 datapoints by keeping all the chunks of one corresponding binaural recording on the same set. A 10-fold cross-validation algorithm performing Ridge Classification with Scikit-learn library for Python was performed on the progressively increasing number of samples passed to the model.

4. RESULTS

4.1. Dependencies Within Paired Perceptual Attributes

Table 1 reports the correlations and regression slopes of the scores averaged for each location between only opposite perceptual attributes. Correlation between nonparallel perceptual attributes was not investigated as the correlation would follow the distribution of soundscapes across the circumplex model.

Annoying–pleasant and chaotic–calm pairs show similar results. In both cases, the range of mean values across the locations occupies a moderately large portion of the Likert scales (see **Supplementary Figure 2** and **Supplementary Table 1**). In both cases, the linear dependency between the scores in the two pairs are characterized with large correlations (r = 0.99) and with slope coefficients with absolute values slightly lower than 1 (−0.87 for annoying–pleasant pair and −0.81 for chaotic–calm pair, where a −1 slope along with a +5 intercept value identifies a perfect overlap of two scales). The corresponding regression slopes show a larger agreement in the positive attribute (pleasant and calm) than disagreement for the negative attribute (annoying and chaotic). In the calm–chaotic pair, which scores are larger spread along the disagreement pole of calm and agreement pole of chaotic, it is also seen that a larger disagreement in calm corresponds to a smaller agreement in chaotic. The monotonous–vibrant pair shows a more random behavior ($r = -0.53$) with all their respective average scores falling in a small region close to neutral score (see also **Supplementary Table 1**) between neutral and somewhat agree for vibrant and between neutral and somewhat disagree for monotonous. Within the uneventful–eventful pair, despite a similar small range of values falling between neutral and somewhat agree for eventful (save for Marchmont Garden and Regents Park Fields locations, see **Supplementary Table 1**) and between neutral and somewhat disagree for uneventful, a moderate correlation ($r = -0.79$) indicates that participants are more likely to disagree with uneventful rather than agree with eventful.

4.2. Extraction of the Correction Values

For each location, the percentage of assessments falling in each of the three grouped categories (see section 3.4) was calculated in each observed location. Their correlation and p-values across all the locations were calculated between all the perceptual attributes and reported in **Table 2**. Regression slopes are shown in **Table 3**. In **Table 2**, it can be noticed that for each perceptual attribute the percentage of neutral scores are negatively correlated with agreement across positive perceptual attributes (pleasant, vibrant, calm, eventful), and negatively correlated with disagreement across negative perceptual attributes.

The correction factors method introduced in section 3.4.2 has been applied to extract these values from **Figure 3** and these are reported in numerical and visual format in **Figure 5**. The values, reported in the same figure, are not normalized as the method introduced provides a relative proportional information between the scales. The results obtained in the table are given by setting the pleasant disagree value $a_0 = -1$, following the formulas introduced in section 3.4.2.

In-depth discussions of these results for each perceptual attribute have been included in **Supplementary Material** (section S.1).

4.3. Application on Soundscape Modeling

The results of the Ridge Classifier prediction models for predicting both the ISO targets and the rescaled metric targets are shown in **Figure 6**. For both the circumplex coordinates, the results show a higher accuracy on training and test sets for the model predicting the categories of re-scaled items compared to the one predicting the categories computed from the ISO- (accuracy of pleasantness on the rightmost point of the curve: ISO 41.4%, rescaled 46.5%; accuracy of eventfulness on the rightmost point of the curve: ISO 46.2%, rescaled 48.3%). Nonetheless, it can be observed from the graphs a closer convergence of the learning curve for the model based on the rescaled metrics rather than the original ISO one. The training curves reported in **Figure 6** provides an upper limit under which the validation performance can improve. By augmenting the number of samples fit in the model, the training tests decrease their accuracy as they rely on a larger variance across the samples. At the same time, it is more likely that the statistics of samples in the test sets match the ones in the training sets, therefore increasing the performance of the test set. However, the distances between the training and test curves in each of

TABLE 2 | Correlation table among percentage of agreement, disagreement, and neutral scaling for each perceptual attribute across all the soundscapes.

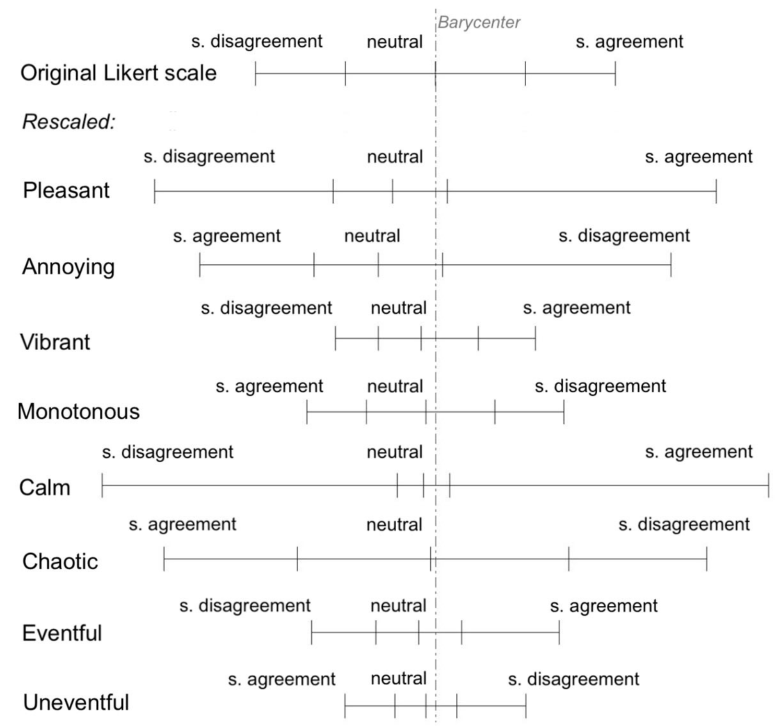
		Pleasant			Annoying			Vibrant			Monotonous			Calm			Chaotic			Eventful			Uneventful		
		Disagree	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**
Pleas.	Disagree	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Neutral	0.63*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Agree	−0.96***	−0.82**	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Ann.	Disagree	−0.92***	−0.85**	0.98***	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Neutral	0.77**	0.92***	−0.89***	−0.94***	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Agree	0.97***	0.73*	−0.97***	−0.97***	0.83**	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Vib.	Disagree	0.64*	0.42	−0.62*	−0.67*	0.61*	0.66*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Neutral	−0.10	−0.24	0.16	0.12	−0.27	0.00	−0.14	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Agree	−0.39	−0.11	0.33	0.39	−0.23	−0.48	−0.61*	−0.69*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Mon.	Disagree	−0.82**	−0.58	0.81**	0.82**	−0.76**	−0.8**	−0.55	0.00	0.40	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Neutral	0.08	0.03	−0.07	−0.12	0.10	0.12	0.13	0.54	−0.52	−0.52	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Agree	0.91***	0.66*	−0.91***	−0.89***	0.83**	0.86***	0.56	−0.31	−0.16	−0.86***	0.02	–	–	–	–	–	–	–	–	–	–	–	–	–
Calm	Disagree	0.88***	0.8**	−0.93***	−0.91***	0.87***	0.87***	0.43	−0.34	−0.03	−0.73*	−0.04	0.88***	–	–	–	–	–	–	–	–	–	–	–	–
	Neutral	0.10	0.73*	−0.33	−0.46	0.62*	0.30	0.24	0.03	−0.20	−0.13	0.03	0.13	0.31	–	–	–	–	–	–	–	–	–	–	–
	Agree	−0.84**	−0.87***	0.93***	0.93***	−0.92***	−0.87***	−0.44	0.32	0.06	0.71*	0.04	−0.85**	−0.99***	−0.44	–	–	–	–	–	–	–	–	–	–
Cha.	Disagree	−0.83**	−0.88***	0.92***	0.95***	−0.94***	−0.89***	−0.47	0.20	0.18	0.74**	−0.09	−0.82**	−0.95***	−0.54	0.98***	–	–	–	–	–	–	–	–	–
	Neutral	−0.18	0.24	0.05	−0.08	0.22	−0.02	−0.19	0.12	0.04	−0.07	0.27	−0.08	0.04	0.60*	−0.13	−0.27	–	–	–	–	–	–	–	–
	Agree	0.91***	0.84**	−0.97***	−0.96***	0.91***	0.93***	0.55	−0.25	−0.20	−0.75**	0.01	0.87***	0.98***	0.38	−0.98***	−0.96***	−0.02	–	–	–	–	–	–	–
Eve.	Disagree	−0.02	−0.03	0.03	−0.02	−0.07	0.08	0.33	0.67*	−0.78**	−0.15	0.56	−0.16	−0.36	0.16	0.32	0.19	0.06	−0.21	–	–	–	–	–	–
	Neutral	−0.67*	−0.70*	0.74**	0.70*	−0.64*	−0.70*	−0.54	0.19	0.24	0.43	−0.06	−0.47	−0.71*	−0.33	0.72*	0.70*	0.20	−0.79**	0.07	–	–	–	–	–
	Agree	0.41	0.44	−0.46	−0.41	0.43	0.36	0.07	−0.63*	0.45	−0.14	−0.39	0.4	0.7*	0.07	−0.68*	−0.56	−0.17	0.63*	−0.8**	−0.65*	–	–	–	–
Uneve.	Disagree	−0.07	0.08	0.02	0.06	0.05	−0.13	−0.32	−0.57	0.69*	0.09	−0.14	−0.02	0.30	−0.10	−0.27	−0.18	0.06	0.17	−0.81**	−0.30	0.80	–	–	–
	Neutral	−0.46	−0.61*	0.55	0.58	−0.70*	−0.45	−0.49	0.51	−0.06	0.49	−0.17	−0.47	−0.67*	−0.35	0.68*	0.66*	−0.07	−0.67*	0.34	0.70*	−0.68*	−0.57	–	–
	Agree	0.42	0.34	−0.43	−0.50	0.45	0.49	0.74**	0.33	−0.8**	−0.46	0.29	0.37	0.11	0.37	−0.16	−0.26	−0.03	0.27	0.74**	−0.15	−0.48	−0.8**	−0.03	–

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

TABLE 3 | Regression slopes for correlation coefficients $r > 0.7$ and p-values $p < 0.05$.

		Pleasant			Annoying			Vibrant			Monotonous			Calm			Chaotic			Eventful			Uneventful		
		Disagree	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**	**Disagree**	**Neutral**	**Agree**
Pleas.	Disagree	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Neutral	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Agree	−0.74	−0.36	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Ann.	Disagree	−0.81	−0.40	1.10	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Neutral	1.90	0.90	−2.5	−2.29	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Agree	1.33	0.68	−1.81	−1.65	0.74	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Vib.	Disagree	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Neutral	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Agree	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Mon.	Disagree	−1.72	–	2.32	2.14	−0.95	−1.30	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Neutral	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Agree	1.99	–	−2.69	−2.47	1.09	1.50	–	–	–	−1.18	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Calm	Disagree	0.74	0.36	−1.0	−0.92	0.4	0.55	–	–	–	−0.46	–	0.38	–	–	–	–	–	–	–	–	–	–	–	–
	Neutral	-	2.32	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Agree	−0.70	−0.34	0.94	0.86	−0.37	−0.52	–	–	–	0.44	–	−0.36	−0.94	–	–	–	–	–	–	–	–	–	–	–
Cha.	Disagree	−0.88	−0.42	1.17	1.07	−0.47	−0.65	–	–	–	0.53	–	−0.45	−1.18	–	1.24	–	–	–	–	–	–	–	–	–
	Neutral	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Agree	0.89	0.44	−1.21	−1.10	0.49	0.67	–	–	–	−0.55	–	0.46	1.22	–	−1.29	−1.04	–	–	–	–	–	–	–	–
Eve.	Disagree	–	–	–	–	–	–	–	–	−1.11	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Neutral	–	−1.39	3.86	3.56	–	−2.14	–	–	–	–	–	–	−3.95	–	4.05	–	–	−3.11	–	–	–	–	–	–
	Agree	–	–	–	–	–	–	–	–	–	–	–	–	2.34	–	–	–	–	–	−0.78	–	–	–	–	–
Uneve.	Disagree	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	−1.07	–	1.41	–	–	–
	Neutral	–	–	–	–	−2.13	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Agree	–	–	–	–	–	–	1.05	–	−1.41	–	–	–	–	–	–	–	–	–	1.33	–	–	−1.25	–	–

Independent variables are in the rows, and dependent variables are in the columns (e.g., $pleasant_{disagree} = -0.74 \times pleasant_{agree} + const.$*).*



Corrected values:	dd	d	n	a	aa
pleasant	-1.18	-0.43	-0.18	0.05	1.18
annoying	0.99	0.03	-0.24	-0.51	-0.99
vibrant	-0.42	-0.24	-0.06	0.18	0.42
monotonous	0.54	0.25	-0.04	-0.29	-0.54
calm	-1.40	-0.16	-0.05	0.06	1.40
chaotic	1.14	0.56	-0.02	-0.58	-1.14
eventful	-0.52	-0.25	-0.07	0.11	0.52
uneventful	0.38	0.09	-0.04	-0.17	-0.38

FIGURE 5 | (Left) Graphical representation of the dilation and compression across the dimensions. **(Right)** Corrected values of the Likert scales. dd, strongly disagree; d, somewhat disagree; n, neither agree nor disagree; a, somewhat agree; aa, strongly agree.

the four targets (ISO pleasantness, corrected pleasantness, ISO eventfulness, and corrected eventfulness) show that there is still a margin of improvement for the current models, which can be achieved by augmenting the data samples. Nevertheless, the distances between the training curves, in both graphs, show a systematically better performance, according to the model framework used in this study, of the corrected coordinates compared to the ISO ones.

5. DISCUSSION

5.1. Interpretation of the Correlations Within Pairs of Perceptual Attributes

The high correlation coefficients found in **Table 1** suggests a systematic unbalanced interpretation of the scales within pairs of perceptual attributes. By plotting the regression slopes found from **Table 1**, given in **Supplementary Figure 2**, the following conclusions can be made. A general trend across the soundscapes in our dataset shows that the average participant tends to assess a given soundscape as more pleasant than it is not annoying. This pattern continues to the other side of the pole, where a soundscape is rated as relatively less annoying than it is not pleasant, however this behavior is not symmetrical about neutral. The whole line demonstrating this behavior is shifted toward pleasant, such that a neutral pleasant rating (3) on average corresponds to a slightly lower than neutral (2.9) annoying rating. These trends are replicated similarly for the perceptual attribute pair calm–chaotic. Despite this slight unbalance between pleasant to annoying and calm to chaotic ratings, strong correlations ($r = -0.99$ for both pairs pleasant–annoying and calm–chaotic) are still present, as shown in **Table 1**.

A possible explanation for the unbalanced patterns observed in some pairs of attributes is that, when performing the scaling task, participants indeed do not recognize and/or interpret them as being paired, or else, semantically opposite as per the circumplex framework. While for some cases the pairing may be more obvious (e.g., eventful–uneventful), one cannot assume this is always the case (e.g., vibrant–monotonous). Even so, when the circumplex space is not presented visually as such, it is difficult to confirm whether participants are detecting paired items as they could be associating different meanings to the attributes. Without the visual representation, the framework relies on a common understanding of the specific terms used in order to achieve the dimensional relationships. Respondents are presented with eight apparently unrelated perceptual attributes to score, and this could lead to some inconsistencies while scoring corresponding attributes.

5.2. Correlation on Percentage of Agree, Disagree, and Neutral Scores

It must be noticed from **Figure 7** that the soundscapes are sampled from a narrow region parallel and transposed above the calm–chaotic bisector. A point which needs to be stressed is an eventual dependency between the distribution of the soundscapes

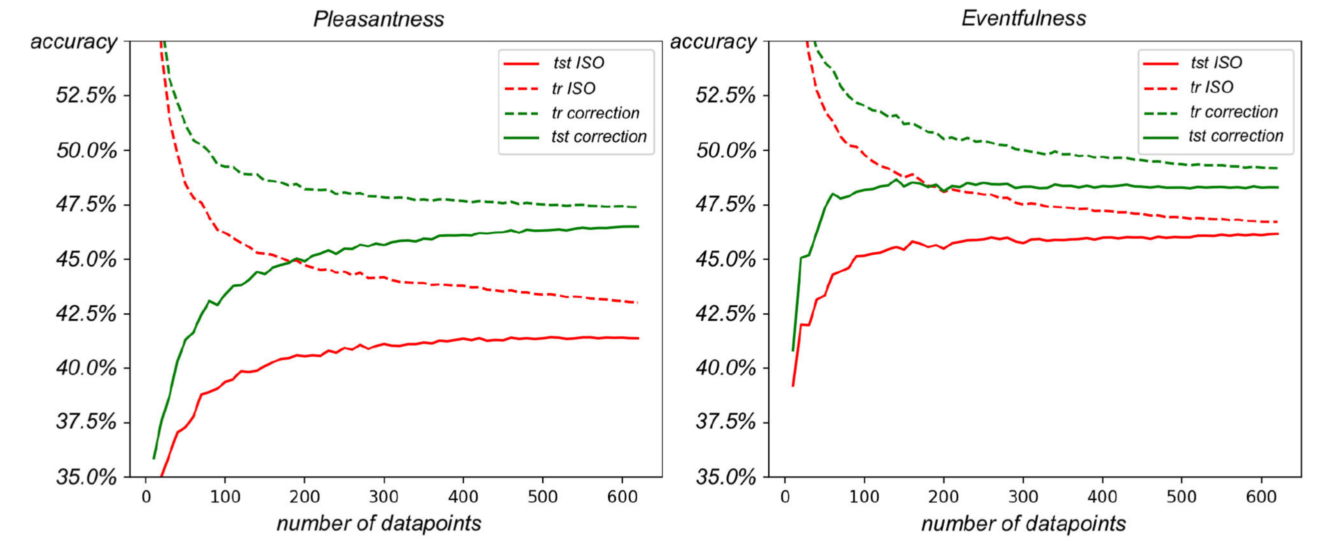


FIGURE 6 | Comparison of Ridge Classifiers fit on the corrected and original version of the ISO circumplex coordinates while increasing the number of datapoints. **(Left)** ISO and corrected ISO circumplex pleasantness. **(Right)** ISO and corrected ISO circumplex eventfulness.

onto the circumplex model and the results from **Table 2**. However, the low *p*-values in **Table 2** suggests the hypothesis that the slopes could enclose some universal properties of the soundscapes and are not dependent on specific locations.

The random behavior expected between percentage of assessments falling in each categories across the locations (see section 3.4.1) is shown to be not assessed in **Table 2**. The neutral answers show, especially across pleasant and annoying in **Table 3**, high correlations with the other two groups of Likert categories. Other strong correlations can be seen in multiple slopes in **Table 2**. This unexpected results show that there is some systematic behavior in the percentage classes and so a systematic biased interpretation of the Likert scaling.

5.3. Projection Onto the Corrected Circumplex Space

To demonstrate and visualize the scaling effects across the soundscape circumplex space, a density plot with randomly generated data is shown in **Figures 7A,B**. Note that 30,000 responses were simulated for each of the eight perceptual attributes, with a uniform distribution of integers from 1 to 5, representing raw Likert scale responses that uniformly cover their respective axes. These were then projected according to the recommendations of ISO/TS 12913-3:2019 (as shown in **Figure 1**), resulting in a normal distribution of responses in both the ISO Pleasantness and Eventfulness axes, and the distribution density is plotted on the bidimensional axis (**Figure 7A**). This dataset is then scaled according to the correction values shown in **Figure 5** and projected and normalized as described above. The resulting density distribution of the corrected circumplex space is plotted in **Figure 7B**. The change in the shape of the distribution density shown when moving from **Figures 7A,B** demonstrates the scaling of the original ISO space performed by the derived correction values.

Two main changes can be observed in the corrected density distribution: (1) an overall shift of the modal center of the space along the negative horizontal axis, and (2) a stretching along the chaotic–calm axis. In particular, this results in a compression of the vibrant-monotonous dimension—in practical terms, when this scaling is applied, soundscapes which may have fallen within the vibrant quadrant according to the standard method are likely to be resented as shifted toward the calm or chaotic quadrants. The compression found in **Figure 7B** shows a more likely representation of how the circumplex model is interpreted and experienced during the scaling task.

Moreover, this representation could be helpful in understanding distances between soundscapes and the actual impact in variation of coordinates when manipulating some elements in an existing or simulated soundscape, when visiting the same location under different contextual conditions, or when sampling assessments associated to participants with different perceptual sensitivity. Nonetheless, this change affecting the vibrant region of the model may also reflect a misunderstanding or disagreement about the meaning of vibrancy (as a perceptual construct) among respondents. This argument is partially supported by the results shown in **Tables 1**, **2**, where significant correlations with the vibrancy attribute are limited to eventful and uneventful. This would be consistent with a previous study where "vibrant" was found to be correlated with "eventful" but not with "pleasant" (Aletta and Kang, 2018), which is generally in contrast with the theory underpinning the circumplex model of affect. Previous literature shows that vibrant soundscapes are associated with simultaneous social presence (e.g., human sounds of chatter or laughter) and presence of musical sounds. Such features were not necessarily typical at the 11 sampled

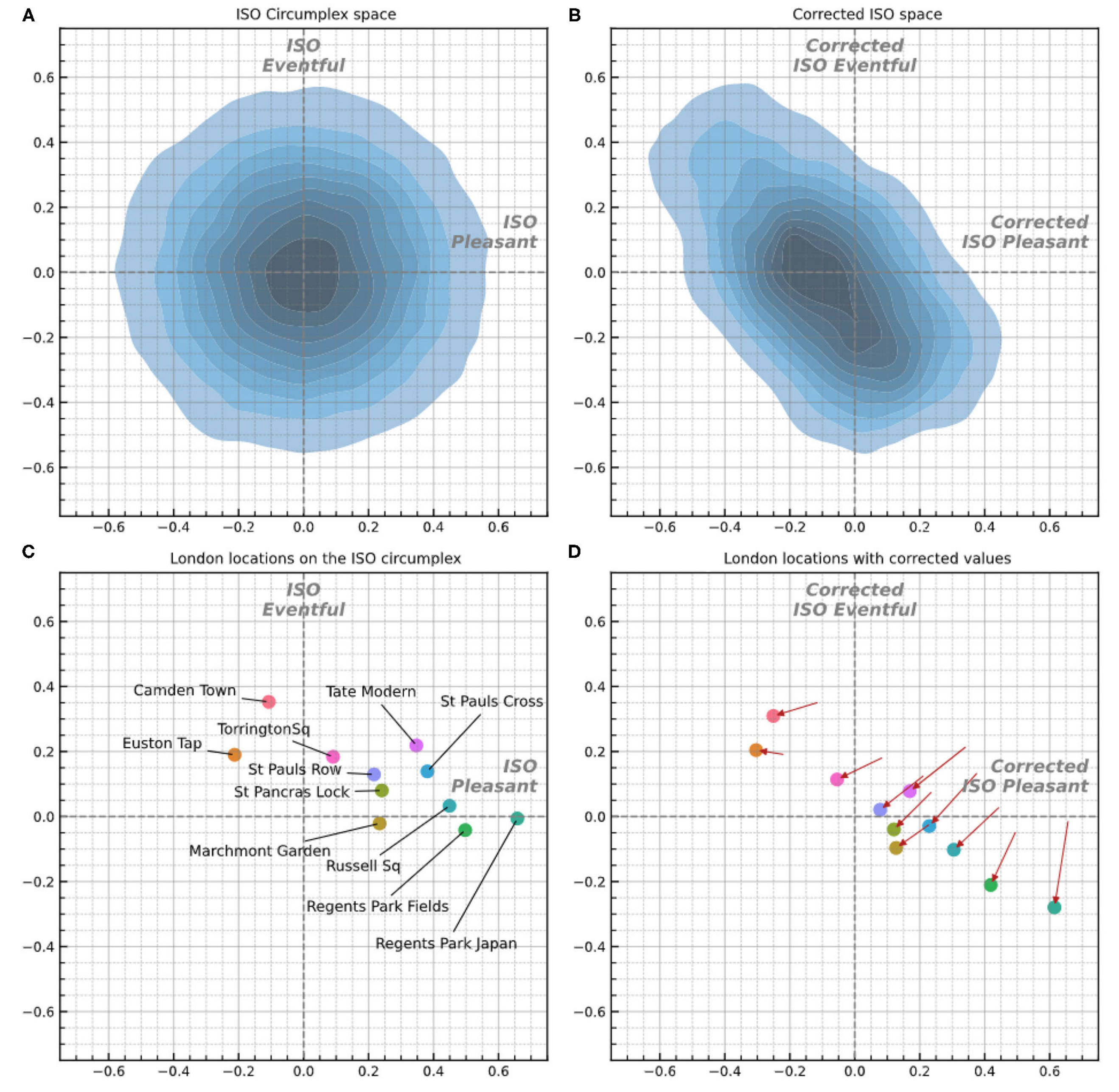


FIGURE 7 | (A) Density plot of 30,000 randomly generated Likert responses projected on the circumplex model of soundscape according to the original metrics given by ISO ISO/TS (2019). **(B)** The same randomly generated responses corrected according to the values derived in section 3.4.2 and projected onto the corrected circumplex model. **(C)** London soundscape locations projected onto the ISO circumplex. **(D)** London soundscape locations after correction—arrows indicate the travel from their initial ISO coordinates to the new corrected coordinates. The x- and y-axes in all diagrams are normalized to [−1, 1].

locations, so this could have resulted in more scattered responses around the vibrancy construct, inflating their representation in the un-corrected ISO model.

This analysis of uniformly simulated response data also reveals some fundamental concerns with the ISO circumplex framework, outside of the metric interpretation addressed by the correction values. The fact that random data, which uniformly cover the initial perceptual attribute space, are then transformed to a normal distribution in the projected ISO circumplex space indicates that, contrary to the common interpretation, the circumplex space bounded by $[-1, 1]$ is not uniformly available to be populated by soundscapes. This is a more fundamental question within the circumplex projection framework, which is independent of the Likert metric scaling caused by respondents' interpretations of the Likert scale, which is otherwise the focus of this study.

Taking the density distribution shown in **Figure 7A** as a probability density of where soundscapes can fall in the circumplex, it can be seen that soundscapes are much more likely to be placed toward the center of the circumplex. In this view, the effective limit for a soundscape composed of multiple responses (e.g., taken across a location) is in reality around [−0.6, 0.6], not [−1, 1]. Within the randomly generated data, <0.46% of pleasant values fell above 0.6. 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, etc.) coordinate is significantly less likely to occur than a neutral coordinate. The field of soundscape studies should therefore adjust our conception of the ISO circumplex space from ideally being equally populated by soundscapes across the full [−1, 1] and reframe our scaling of the value of the ideal "most pleasant" soundscape from [1, 0] to [0.6, 0]. Alternatively, a separate method of projecting and representing the pleasantness vs. eventfulness values, which does conform with the common understanding in the field, could be developed.

5.4. Correction of London Soundscape Coordinates

Applying the correction metrics to the actual London soundscapes data demonstrates how this compression and correction of the circumplex space affects the coordinates of real locations. **Figure 7C** shows the London soundscape locations projected into the ISO circumplex space, and in **Figure 7D**, the locations' corrected coordinates are plotted. The coordinates of each soundscape in the new circumplex model are determined by replacing the original scores of the assessments given by the participants—ranging from 1 "strongly disagree" to 5 "strongly agree"—with the new scores reported in **Figure 5**. The coordinates are then normalized ranging from $[-1, 1]$ by dividing them by the sum of the positive scores reachable in both the corrected pleasant and eventful dimensions.

The comparison of the new projection with the one done through the original metric values becomes a complex task as the new dimensions lose some information such as the slopes of the diagonal sub-dimensions and the neutral assessment regions. The general movement of the soundscape coordinates (as indicated by the arrows in **Figure 7D**) reflect the transformation of the circumplex indicated in **Figure 7B**. "Regents Park Japan" appears to be the calmest and one of the most pleasant soundscapes (whose value does not seem to be much affected by the new metrics) and least eventful soundscape in the new metric. "Camden Town" maintains the same high value of pleasantness, and "Euston Tap" remains the soundscape with the lowest pleasantness score. "Russell Sq," "St Pancras Lock," and "St Pauls Cross" are shifted from the vibrant to the calm quadrant and "Torrington Sq" is moved from the vibrant to chaotic quarter. The eventfulness distance between "Torrington Sq" and "Euston Tap" is significantly increased as well as the distance between "Tate Modern" and "Euston Tap." An overall trend appears to compress the distribution in a narrow region along the calm–chaotic bisector. Finally, it is possible to notice a compression along the pleasant dimension over those soundscapes falling in the positive pleasant side of the new projection plane, while the negative pleasant side of the plot preserves a similar spread as the original model.

5.5. Comparison of Linear Predictability Between ISO and Corrected ISO Targets

The better performance of Ridge Classifier introduced in section 4.3 in predicting the new metric compared to the ISO targets suggests a better linear mapping between acoustic information and the newly retrieved metrics compared to the raw orthonormal projection described in the standard. This performance improvement of the linear modeling task supports the idea that the corrected values create a better linear representation of the Likert scale, increasing the validity of applying mathematical operations that assume equidistant Likert categories. Specifically within soundscape studies, the improvement of the modeling results indicate that these correction values should be applied for the construction of future predictive soundscape models, which make use of the ISO circumplex framework. It should be noted that in this example the results are limited to a linear modeling case; it is unclear at this stage whether an eventual model that can incorporate nonlinear mapping would demonstrate the same improvement in performance using the rescaled metric values.

5.6. Limits of the Current Framework

As introduced earlier in section 1, the output of the correction scale model is bound by some constraints inherited by the design of the data collection. Here, follows a discussion upon these bonds trying to answer what it is expected to happen when some of these conditions change. It is first assumed that the output of the model is not affected by the particular distribution of the locations across the perceptual attributes space. This assumption is first needed under the consideration that in spite of the relatively narrow distribution of the locations across their perceptual attributes, as shown by the projection onto the circumplex model, the amount of locations in relation to the number of participants in each site makes the current dataset one of the largest projects in soundscape data collection that uses *in situ* surveys. The missing regions on the circumplex model, not covered in the current study, represent situations where the collection of the data represents a challenging task because of the reduced number of potential participants either because of low density of persons in the areas or because of less likely attitude to participating to the study. For the locations corresponding to these cases, the data collection is arranged to be performed through laboratory experiments. However, the biases introduced by the environmental validity would not fit the requirements of the current study. In those studies where extra laboratory experiments are required to fill missing regions across the score distributions of the scales, these are expected to be analyzed in comparison to the ground truth output of the model derived directly from the target collection procedure.

Our results and previous literature indicate that there is uncertainty around the concepts of vibrancy and monotonous within the ISO soundscape standard. This method has attempted to address some of this uncertainty internally, however it may also be possible to partially address this at the data collection stage by adjusting the semantic attributes used for these dimensions. This would then likely reduce the amount of internal correction needed. It is worthwhile to remark that in the original Swedish Soundscape Quality Protocol (SSQP) developed by Axelsson et al. (2010), the attribute used was "exciting," which was their translation from the Swedish version of the questionnaire. Starting with Cain et al. (2013), this was replaced with "vibrant," which has made its way into the ISO standard version. Future work in the space should investigate the differences between these and other versions of the attributes on the vibrant/monotonous dimension, as well as the usefulness of presenting multiple descriptions of the attributes to respondents.

6. CONCLUSIONS

When performing mathematical operations using Likert scaled survey data—whether that be calculating the mean of the scale values or performing a multidimensional projection—assumptions about the distance metric underlying the scale must be made. The typical assumption of equidistance between categories has been shown to not hold when examining multidimensional, paired Likert scales. By examining the correlations between response rates of the grouped Likert categories, and extracting commonly shared interpretations of the metric scaling, corrected Likert values are calculated. These corrected values account for the lack of correspondence between the equidistant Likert metrics and the participants' actual interpretation of the scaling task, thereby allowing mathematical operations to be valid when applied to the data. The implications of this scaling have been demonstrated through a Linear Ridge Classifier task, which shows significant improvement when applied to the corrected data.

This study was conceived and developed in the context of soundscape standardization processes about data collection methods and data analysis. The identity map that should match the interpretation of the scaling task for public space users with the formal model was questioned. Participants in the study used the scales differently from what would be expected based on the soundscape assessment theoretical framework. To address this, a correction factor matrix has been introduced for adjusting the Likert scale metrics and extracting corrected values applied to the categories for each Likert scale.

The findings indicate that (1) in soundscape studies, intervals are not necessarily interpreted to range equidistant spaces between Likert scale categories; (2) there is a matching between neutral and disagreement assessment for positive soundscape attributes and a correlation between agreement and neutral assessments across negative soundscape attributes; (3) intervals centered on neutral assessments are generally interpreted to be smaller than intervals placed on the extreme of the scales; and (4) the new metric is better described by (psycho)acoustic features compared to the original Likert scale metric, when used as indicators to predict how people experience urban soundscapes.

Moreover, from the results and comparison of the two projected spaces, the ISO and the corrected one, the following points have been found. The ISO circumplex model framework implies that a perceptual shift in the bidimensional space is direction independent. In other words, when the soundscape of a location changes due to dynamics of its contextual, physical, or other variables, the magnitude of perceptual differences should be equal regardless of the direction of shift or initial position in the bidimensional space. However, our data show that this is not the case in the original ISO space. The lack of this position- and direction-independence property in the perception of the ISO circumplex model along with the lack of overlapping match between Likert categories belonging to different perceptual attributes makes the circumplex projection, as described in the current ISO/TS 12913-3:2019, less effective in describing soundscapes by means of pleasantness and eventfulness coordinates. Particularly, the ISO space is found to be effected by a dilation along the vibrant–monotonous dimension, in terms of participants' scaling behaviors along that direction and in comparison to the same spatial shift in other directions. This exaggerated stretch is due to the artifacts of the misleading assumption of equally ranged Likert intervals, which is then passed to the ISO projections. Therefore, the vibrant dimension is overestimated in its length, compared to the other directions, due to the artifacts inherited from the unbalanced Likert scales belonging to different perceptual attributes.

It has also been shown that uniformly sampled Likert values, unaffected by the metric interpretation otherwise discussed here, are projected into the raw ISO space as a normal distribution, as opposed to a uniform distribution. This fact implies that soundscapes cannot be fairly distributed across the whole of the available range. This means that the original ISO mapping of the perceptual attributes into the circumplex model is neither a good representation of participants' interpretation of the projected space, nor a meaningfully spread representation of different soundscapes.

These findings suggest that the current ISO standard suffers from some inaccuracies of the standard metric as it is inherited from the raw Likert categories. By implementing the procedure described in this study, soundscape studies would benefit from a better representation in terms of how listeners experience soundscapes. In the proposed corrected circumplex projection, the space metric is intended to provide a perceptually equally spread space, along all the perceptual attribute directions, based on the scaling patterns retrieved from the participants' responses.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because, data used for the study is part of a larger dataset currently under development. Requests to access the datasets should be directed to: kang@ucl.ac.uk.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Departmental approval by the Ethics Lead at the UCL Institute for Environmental Design and Engineering (BSEER Research Ethics—Low Risk Application). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ML: conceptualization, methodology, validation, and investigation. ML, FA, and AM: formal analysis, visualization, resources, and writing. ML and AM: software and data curation. ML, FA, AM, and JK: discussion. FA and JK: supervision. JK: project administration and funding acquisition. All authors contributed to the article and approved the submitted version.

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]).

ACKNOWLEDGMENTS

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

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.602831/full#supplementary-material

REFERENCES

Adroher, N. D., Prodinger, B., Fellinghauer, C. S., and Tennant, A. (2018). All metrics are equal, but some metrics are more equal than others: a systematic search and review on the use of the term "metric." *PLoS ONE* 13:e193861. doi: 10.1371/journal.pone.0193861

Aletta, F., Brambilla, G., Maffei, L., and Masullo, M. (2017). Urban soundscapes: characterization of a pedestrian tourist route in Sorrento (Italy). *Urban Sci.* 1:4. doi: 10.3390/urbansci1010004

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. *Appl. Acoust.* 155, 190–203. doi: 10.1016/j.apacoust.2019.05.024

Aletta, F., and Kang, J. (2018). enTowards an urban vibrancy model: a soundscape approach. *Int. J. Environ. Res. Publ. Health* 15:1712. doi: 10.3390/ijerph15081712

Aletta, F., Oberman, T., Mitchell, A., Tong, H., and Kang, J. (2020). Assessing the changing urban sound environment during the COVID-19 lockdown period using short-term acoustic measurements. *Noise Mapp.* 7, 123–134. doi: 10.1515/noise-2020-0011

Axelsson, Ö., Nilsson, M., and Berglund, B. (2010). A principal components model of soundscape perception. *J. Acoust. Soc. Am.* 128, 2836–2846. doi: 10.1121/1.3493436

Berglund, B., Nilsson, M., and Axelsson, Ö. (2007). "Soundscape psychophysics in place," in *Proceeding of InterNoise, Vol. 6*, 3704–3711.

Cain, R., Jennings, P., and Poxon, J. (2013). The development and application of the emotional dimensions of a soundscape. *Appl. Acoust.* 74, 232–239. doi: 10.1016/j.apacoust.2011.11.006

Carifio, J., and Perla, R. (2008). Resolving the 50-year debate around using and misusing likert scales. *Med. Educ.* 42, 1150–1152. doi: 10.1111/j.1365-2923.2008.03172.x

[Dataset] Eurostat (2020). *Tertiary Educational Attainment, Age Group 25-64 by Sex and Nuts 2 Regions.* Available online at: https://ec.europa.eu/eurostat/tgm/table.do?tab=table (Retrieved September 02, 2020).

Fiebig, A., and Herweg, A. (2017). "The measurement of soundscapes -a study of methods and their implications," in *Internoise* (Hong Kong).

Giannakopoulos, T., Orfanidi, M., and Perantonis, S. (2019). "Athens urban soundscape (athus): a dataset for urban soundscape quality recognition," in *International Conference on Multimedia Modeling* (Thessaloniki).

ISO 12913-1:2014 (2014). *Acoustics "Soundscape" Part 1: Definition and Conceptual Framework*. Geneva: Standard, International Organization for Standardization.

ISO/TS 12913-2:2018 (2018). *Acoustics "Soundscape" Part 2: Data Collection and Reporting Requirements*. Geneva: Standard, International Organization for Standardization.

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

Jamieson, S. (2005). Likert scales: How to (ab) use them. *Med. Educ.* 38, 1217–1218. doi: 10.1111/j.1365-2929.2004.02012.x

Kang, J., Aletta, F., Margaritis, E., and Yang, M. (2018). A model for implementing soundscape maps in smart cities. *Noise Mapp.* 5, 46–59. doi: 10.1515/noise-2018-0004

Kogan, P., Turra, B., Arenas, J., Zeballos, F., Hinalaf, M., and Perez Villalobo, J. (2016). "Application of the swedish soundscape-quality protocol in one european and three latin-american cities," in *The International Conference on Acoustics* (Buenos Aires).

Lantz, B. (2013). Equidistance of likert-type scales and validation of inferential methods using experiments and simulations. *Electron. J. Bus. Res. Methods* 11, 16–28.

Likert, R. (1932). A technique for the measurement of attitudes. *Arch. Psychol.* 140, 1–55.

Lindburg, P., and Friberg, A. (2016). Personality traits bias the perceived quality of sonic environments. *Appl. Sci.* 6:405. doi: 10.3390/app6120405

Lionello, M., Aletta, F., and Kang, J. (2019). "On the dimension and scaling analysis of soundscape assessment tools: a casestudy about the "method a" of iso/ts 12913-2:2018," in *The 23rd International Congress on Acoustics* (Aachen). doi: 10.18154/RWTH-CONV-239311

Lionello, M., Aletta, F., and Kang, J. (2020). A systematic review of prediction models for the experience of urban soundscapes. *Appl. Acoust.* 170:107479. doi: 10.1016/j.apacoust.2020.107479

Maffiolo, V., Castellengo, M., and Dubois, D. (1999). Is pleasantness for soundscapes dimensional or categorical? *J. Acoust. Soc. Am.* 105:943. doi: 10.1121/1.425718

Mitchell, A., Oberman, T., Aletta, F., Erfanian, M., Kachlicka, M., Lionello, M., et al. (2020). The soundscape indices (SSID) protocol: a method for urban soundscape surveys-questionnaires with acoustical and contextual information. *Appl. Sci.* 10:2397. doi: 10.3390/app10072397

Pell, G. (2005). Use and misuse of likert scales. *Med. Educ.* 39, 970–970. doi: 10.1111/j.1365-2929.2005.02237.x

Posner, J., Russell, J., and Peterson, B. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* 17, 715–734. doi: 10.1017/S0954579405050340

Rickards, G., Magee, C., and Artino, A. (2012). You can't fix by analysis what you've spoiled by design: developing survey instruments and collecting validity evidence. *J. Grad. Med. Educ.* 4, 407–410. doi: 10.4300/JGME-D-12-00239.1

Russell, J. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178.

Xiao, J., and Hilton, A. (2019). An investigation of soundscape factors influencing perceptions of square dancing in urban streets: a case study in a county level city in China. *Int. J. Environ. Res. Publ. Health* 16:840. doi: 10.3390/ijerph16050840

Yang, W., and Kang, J. (2005). Soundscape and sound preferences in urban squares: a case study in sheffield. *J. Urban Design* 10, 61–80. doi: 10.1080/13574800500062395

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 March 2021
doi: 10.3389/fpsyg.2021.570741

Check for updates

Soundscape in Times of Change: Case Study of a City Neighbourhood During the COVID-19 Lockdown

Sara Lenzi[1], Juan Sádaba[2] and PerMagnus Lindborg[3]**

[1] DensityDesign Lab, Department of Design, Politecnico di Milano, Milan, Italy, [2] Department of Architecture, Universidad del País Vasco, San Sebastián, Spain, [3] School of Creative Media, City University of Hong Kong, Kowloon, Hong Kong

OPEN ACCESS

Edited by:
*Jose Q. Pinheiro,
Federal University of Rio Grande do Norte, Brazil*

Reviewed by:
*Gustavo Martineli Massola,
University of São Paulo, Brazil
Gleice V. M. A. Elali,
Federal University of Rio Grande do Norte, Brazil*

***Correspondence:**
*Sara Lenzi
sara.lenzi@polimi.it
PerMagnus Lindborg
pm.lindborg@cityu.edu.hk*

Specialty section:
This article was submitted to Environmental Psychology, a section of the journal Frontiers in Psychology

Received: 08 June 2020
Accepted: 29 January 2021
Published: 24 March 2021

Citation:
*Lenzi S, Sádaba J and Lindborg P (2021) Soundscape in Times of Change: Case Study of a City Neighbourhood During the COVID-19 Lockdown.
Front. Psychol. 12:570741.
doi: 10.3389/fpsyg.2021.570741*

The coronavirus disease 2019 (COVID-19) lockdown meant a greatly reduced social and economic activity. Sound is of major importance to people's perception of the environment, and some remarked that the soundscape was changing for the better. But are these anecdotal reports based in truth? Has traffic noise from cars and airplanes really gone down, so that more birdsong can be heard? Have socially distanced people quietened down? This article presents a case study of the human perception of environmental sounds in an urban neighborhood in the Basque Country between 15 March and 25 May 2020. The social restrictions imposed through national legislation divided the 69-day period into three phases. We collected observations, field audio recordings, photography, and diary notes on 50 days. Experts in soundscape and architecture were presented with the recordings, in randomized order, and made two separate perceptual analyses. One group ($N = 11$) rated the recordings for pleasantness and eventfulness using an adapted version of the Swedish Soundscape Quality Protocol, and a partly overlapping group ($N = 12$) annotated perceived sound events with free-form semantic labels. The labels were systematically classified into a four-level Taxonomy of Sound Sources, allowing an estimation of the relative amounts of Natural, Human, and Technological sounds. Loudness and three descriptors developed for bioacoustics were extracted computationally. Analysis showed that Eventfulness, Acoustic Complexity, and Acoustic Richness increased significantly over the time period, while the amount of Technological sounds decreased. These observations were interpreted as reflecting changes in people's outdoor activities and behavior over the whole 69-day period, evidenced in an increased presence of Human sounds of voices and walking, and a significant shift from motorized vehicles toward personal mobility devices, again evidenced by perceived sounds. Quantitative results provided a backdrop against which qualitative analyses of diary notes and observations were interpreted in relation to the restrictions and the architectural specifics of the site. An integrated analysis of all sources pointed at the temporary suspension of human outdoor activity as the main reason for such a change. In the third phase, the progressive return of street life and the usage of personal mobility vehicles seemed to be responsible for a clear increase in Eventfulness and Loudness even in the context of an overall decrease of Technological sounds. Indoor human activity shared through open windows and an increased presence of birdsong emerge as a novel characteristic element of the local urban soundscape. We discuss how such changes in the acoustic environment of the site, in acoustic measurements and as perceived by humans, point toward the soundscape being a crucial component of a comprehensive urban design strategy that aims to improve health and quality of life for increasingly large and dense populations in the future.

Keywords: soundscape, urbanism, perception and cognition, COVID-19, pandemic, social response, case study

INTRODUCTION

"Media that emphasize space are apt to be less durable and light in character. . . such as sounds, for the true character of sound in shaping societies is in its spatial spread. . . and the real paradox is that although sounds are pronounced in time, they are also erased by time" (Schafer, 1977 p. 162). How does the soundscape change over time? In *The Tuning of the World*, Schafer (1977) describes measuring a collection of fire engine sirens covering seven decades. He found that their signal had gotten louder by "nearly half a decibel per year on average" (idem p. 186). This observation supported his general thesis that urban noise levels have increased in industrialized societies, to the detriment of animals and human inhabitants alike. In what may seem as a reply, Arana (2010) analyzed a large number (876,480) of noise measurements taken in the Spanish cities of Pamplona and Madrid between 1999 and 2003. Contrary to Schafer's results, he found a statistically significant decrease in the overall sound level. The findings were translated to inspire politicians and designers; for example, the "remarkable reduction" of noise that had been observed in one district was attributed to the implementation of pedestrian areas.

Such investigations are part of a larger movement. The approach to sound and listening that Schafer pioneered in the mid-1960s has broadened out, in particular through the World Soundscape Project and the World Forum for Acoustic Ecology (see Truax, 2019 for a history), and has become a multifaceted field that is deeply connected with urban planning, policy, health, architecture, activism, and art. Interdisciplinarity has suffused research in the past decade, such as the Soundscape Support to Health program (Berglund and Nilsson, 2007) and the Positive Soundscapes Project (Davies et al., 2013). Viewed from this perspective, soundscape studies have a natural affinity with environmental psychology, even if goals and methodologies are sometimes different.

From the perspective of urban and city planning, as the concern for a more active awareness on the perception of environmental sound grows in the 1960s, perhaps influenced by soundscape studies, urban designers propose a more subjective and qualitative approach to the city. In the United States, Jacobs (1961) advocates in *The Death and Life of Great American Cities* (1961) that while looking at real cities "you might as well also listen, linger and think about what you see." Not far in time, Gehl in Europe claims for a closer attention to the *Life Between Buildings* (1971). The perceivable, intangible aspects of the city environment are linked with physical, tangible components of the architecture as well as the urban design of our human ecosystem. We consider the soundscape as one of the intangible layers of the city. Ultimately, the way we arrange the invisible linkages will determine crucially the form of the city. It will revert to us as a society and will shape behaviors and habits. Thereby, the urban ecosystem we design today is intended for future generations. As shown by Arana (cit.), soundscape research can contribute to improving people's quality of life through urban planning initiatives.

There are few examples of experimental soundscape studies of the kind found in, e.g., medicine or psychology (but see Aletta et al., 2016b, for a covert intervention study). It would be impractical or unethical, or both, to try to implement a double-blind study on the physical scale of cities or nations, or on the temporal scales of decades that Schafer imagined. That being said, we find ourselves today in an extreme situation, with the coronavirus disease 2019 (COVID-19) pandemic sweeping through human societies in every country and every city. It offers an unusual opportunity to put the acoustic environment to a test, almost as if the pandemic lockdown restrictions in various places were conditions in a social experiment at huge geographical and temporal scales. From the urban design point of view, the situation triggered by the COVID-19 outbreak has opened up possibilities to observe how the city environment changes under extreme circumstances. "For a few weeks, the world has rehearsed a post-carbon world, a world not dependent on the car, a world that only consumes what is necessary, a world that only produces the essential, a contained and self-limited world, a world that understands what it is socially relevant and productive" (Fernandez, 2020). Such a context can be used as a testbed not only of environmental changes but also of a desired potential future city with other types of mobility (maybe the city without petrol cars) and different behaviors and proxemics (social distancing). The soundscape is a significant first intangible tester of these changes in a new imposed situation as it happens to be the COVID-19 pandemic. Until now, the tangible–intangible duality has been used for two purposes, not far from the city ecosystem. One is in the field of marketing, in the form of tangible and intangible assets. Another is the field of cultural heritage, which involves physical constructions that intertwine with intangible values and social traditions. In the latter, the "intangible" has already earned a certain rank as a formal category. Our goal should be to broaden the term and to define as "intangible" any feature of the city fabric that is not directly physical. This includes the soundscape.

The present study responds to calls for contextual specificity in soundscape research (Hong and Jin, 2015). Through a case study of the Getxo site, we aim to identify how people's activities and their perception of the acoustic environment can determine whether some aspects of the soundscape have indeed changed significantly during the time of the pandemic lockdown.

We framed the study within the overarching idea that, due to the pressures of the COVID-19 lockdown restrictions, humans and animals would respond by changing their activities and behavior and that the soundscape would indicate the character and magnitude of those changes. This directed our attention toward five assumptions formulated at the beginning of the data collection. Firstly, that loudness would decrease; secondly, sounds linked to machinery and human interaction would decrease; thirdly, human outdoor activity would decrease; and fourthly, birdsong would increase. A fifth assumption emerged from observations, namely, that personal mobility devices (scooters and bikes) would be increasing.

The next section recapitulates the development of the pandemic and lockdown restrictions in response and describes the site of our study. This is followed by a *Methods* section in four parts; outlining the procedures for collecting audio recordings; making diary observations; conducting two analyses by an expert group, which allowed the construction of a Taxonomy of Sound Sources; and extracting computational soundscape descriptors. We report integrated results and make comparisons in relation to the lockdown phases. In the *Discussion*, we return to the five assumptions and attempt to provide answers.

Pandemic Lockdown Response Phases

In early March 2020, cases of COVID-19 started to be detected in Spain. Within 1 week, there were 589 confirmed cases and 10 deaths, and within 2 weeks, Spain recorded 7,753 cases and 288 deaths (Kassam and Burgen, 2020). The accelerating severity of the situation pushed the national government to proclaim a state of emergency. Learning from nearby countries that were "ahead of the curve," notably Italy, strict measures were introduced on 14 March. The project presented in this article started immediately thereafter, on 15 March at noon, on the first day of official application of confinement measures in Spain. In the following weeks, the country moved through different phases of restriction to activity and mobility (see **Table 1**). At around noon every day, the first author made a 5-min audio recording of the sonic environment and made observations from the window of her residency.

Undoubtedly, the coronavirus causes death and a great deal of psychological suffering. The restrictive responses inflict traumatic changes to human lives, with businesses closing, plans being canceled, and the stress of being confined to staying at home. These changes tend to affect those most vulnerable in society the most. By the end of April 2020, most countries in the world had declared partial or total lockdown regimes, causing half of the world population to live in confinement (Sandford, 2020). These measures impose a drastic reduction in human activity, "causing multiple cities across the globe to simultaneously go into hibernation" (Weston, 2020).

There are subtle changes underway that may have less dramatic yet deep and long-lasting influence. The soundscape is both an indicator of environmental quality and a component of cultural identity. For example, one newspaper stated that "life has changed, too: The city no longer sounds the same. And that realization is as jarring as the sight of empty streets" (Bui and Badger, 2020). Others have commented how animals react to the changing environment, such as "birdsong, for instance, seems louder than ever before. Some birds are actually likely to be lower in pitch than before, since they have fewer cars, planes, jackhammers, and leaf-blowers to compete with" (Ro, 2020). As for marine life, "evidence of a drop in underwater noise pollution has led experts to predict [that] the crisis may [. . .] be good news for whales and other sea mammals" (McVeigh, 2020). Reduced human activity affects not only the living conditions for animals but also might even impact the crust of planet Earth itself, as "seismologists are reporting less seismic noise, or vibrations in the Earth's crust" (Ro, 2020).

As mentioned, Spain entered a several-weeks-long sequence of different lockdown phases on 15 March. Citizens are generally confined in their residencies; schools, shops, and other services are suspended; leisure and sports activities are forbidden. Restrictions to local, national, and international mobility vary. In this article, we refer to three phases defined by the severity of the restrictions. **Table 1** briefly illustrates the restrictions to mobility and activity applied in each phase. Phase 1, which lasted approximately 2 weeks, initially allowed citizens' mobility to reach their workplace while suspending any other activity. Phase 2 saw an increase in the severity of the measures with the total suspension of any non-necessary activity. We identify the beginning of Phase 3 with the ease of the most restrictive measures – notably, the permission for kids to spend up to 1 h a day outdoors – which preceded the launch of the so-called "Plan de Desescalada" (Consejos de Ministros de España, 2020) on 28 April.

A detailed analytical overview of the phases with restrictions applied in each phase as adapted by the government of the Basque Country (Mallo, 2020a,b) is available in the **Supplementary Material**.

Site of the Case Study

The case study took place at Plaza de las Escuelas (coordinates 43.325777, −3.014046[1]). Our observation point (see **Figure 1**) is located in Las Arenas, a neighborhood in Getxo, which is a municipality on the estuary of the River of Bilbao. Getxo, a traditionally residential agglomeration with 77,000 inhabitants, lies 12 km from Bilbao, the largest town of the Basque Country. The area is gentrified with few high-rise buildings. Following a trend that is common all over Europe, it is home to an aging population where almost a quarter of the inhabitants are 65 years old or above. The economy of Getxo is essentially based on the third sector (services), which accounts for 92.4% of the turnover of the municipality, with a very

[1] https://www.google.com/maps/@43.3257696,-3.014469,169m/data=!3m1!1e3

TABLE 1 | Description of each phase of lockdown in Spain.

Phase	Duration (date)	Duration (days)	Outline
Phase 1	15-03 to 29-03	Day 1 to Day 16	Restrictions to mobility and to activity.
Phase 2	30-03 to 22-04	Day 17 to Day 39	Further restrictions to mobility, all non-necessary activity suspended.
Phase 3	23-04 to 21-06	Day 40 to Day 100	Progressive release of mobility and activity restrictions, starting with children allowed outdoors 1 h a day.

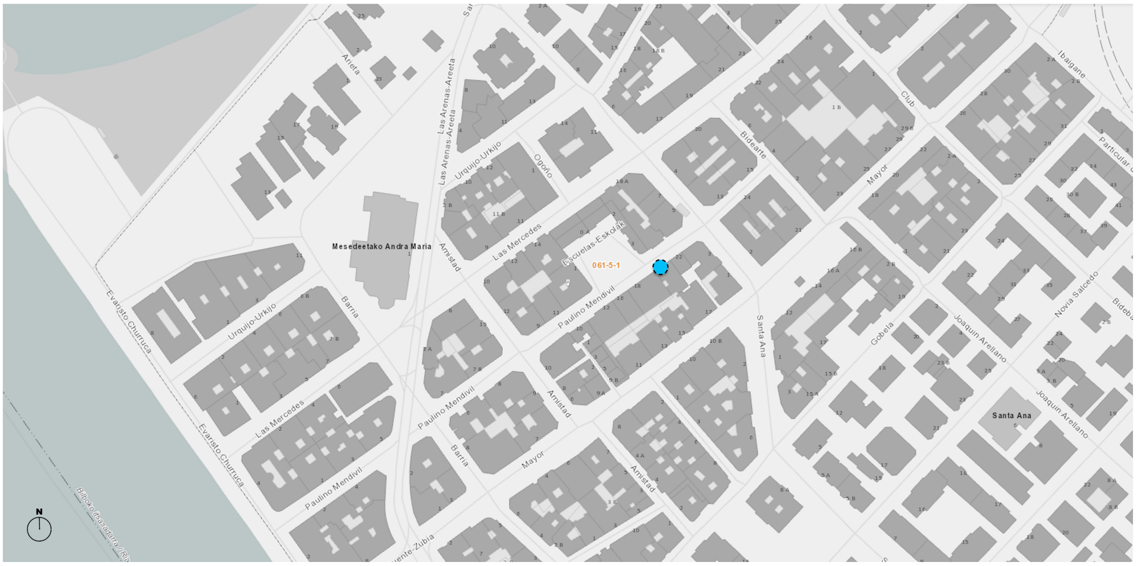


FIGURE 1 | Map of the area surrounding the site.

weak presence of production from the primary and secondary sectors (EUSTAT, 2020).

In order to be able to analyze the soundscape variations from an urban spot that is representative of the city as a whole, we need an observation point that can record different layers of activity, granting that none of them overlaps and covers totally the others. From a mobility point of view, coexisting light traffic [pedestrian, bicycles, and personal mobility vehicle (PMV)] and traditional heavy traffic (cars and freight) must be present. As for human activity, public space should offer both a static gathering space (plaza or alike) and a transit. Regarding economic activities, the spot ought to host shops and bars, mirroring the usual premises of an average residential street of Getxo, including delivery services. Recordings should, in an optimal scenario, be taken from a first floor to avoid distortions and to better engage with street activity.

Paulino Mendívil is a pedestrian road, allowing only for limited vehicle access (08:00 to 12:00 for deliveries) in all its 180-m length. It intersects with Andres Larrazabal street, forming a pedestrian cross that is encircled by the wheel traffic of the surrounding roads. Our observation point is halfway (95 m) to both wheel traffic roads and on the point of convergence (gradient) of the isophonic lines defined by the "Noise White Paper" of the municipality of Getxo (see *Acoustic Environment* section). Hence, we can observe the consecutive layers of sound coming from heavy traffic but also record and analyze dynamic and static public space occupation on the plaza together with commercial activity on the ground floors. The observation point also gives the opportunity to record human activity related to work and leisure at a wide range of time without dealing with a too-dominating traffic noise. Nonetheless, being our observation point halfway to both extremes of the road (where car traffic is allowed), we still can identify motor vehicles flow if existing. The residence of the first author is also on a first floor, which enabled a perfect reference observation point for the study.

With the exception of the children's playground (see **Figure 2**), covered with rubber flooring, the rest of the surfaces are stone slabs. It is important to take these materials into consideration. Materials affect the acoustic properties and hence perception of sound (see, e.g., Lindborg, 2015). A double line of trees at both sides of the road offers a sound and visual cushion (**Figure 3**). This changing green environment also affects the cushioning and filtering of sound that the tree leaves account for. The present study takes place during the growing of leaves and blooming of trees, starting from no leaves at all in mid-March to full coverage at the end of May.

As **Figure 4** illustrates, the site is only 250 m from the sea, which provides wind and the sonic presence of seagulls, waves, and boats. There is also a church tower with bells nearby, and the main activity road of the neighborhood (Calle Mayor, see **Figure 2**) is one block apart. The surrounding larger area is mainly residential.

The weather in Getxo is mild, with temperatures oscillating between a minimum of 6°C and a maximum of 22°C on a yearly average. The quantity of yearly rain precipitation is not low, going

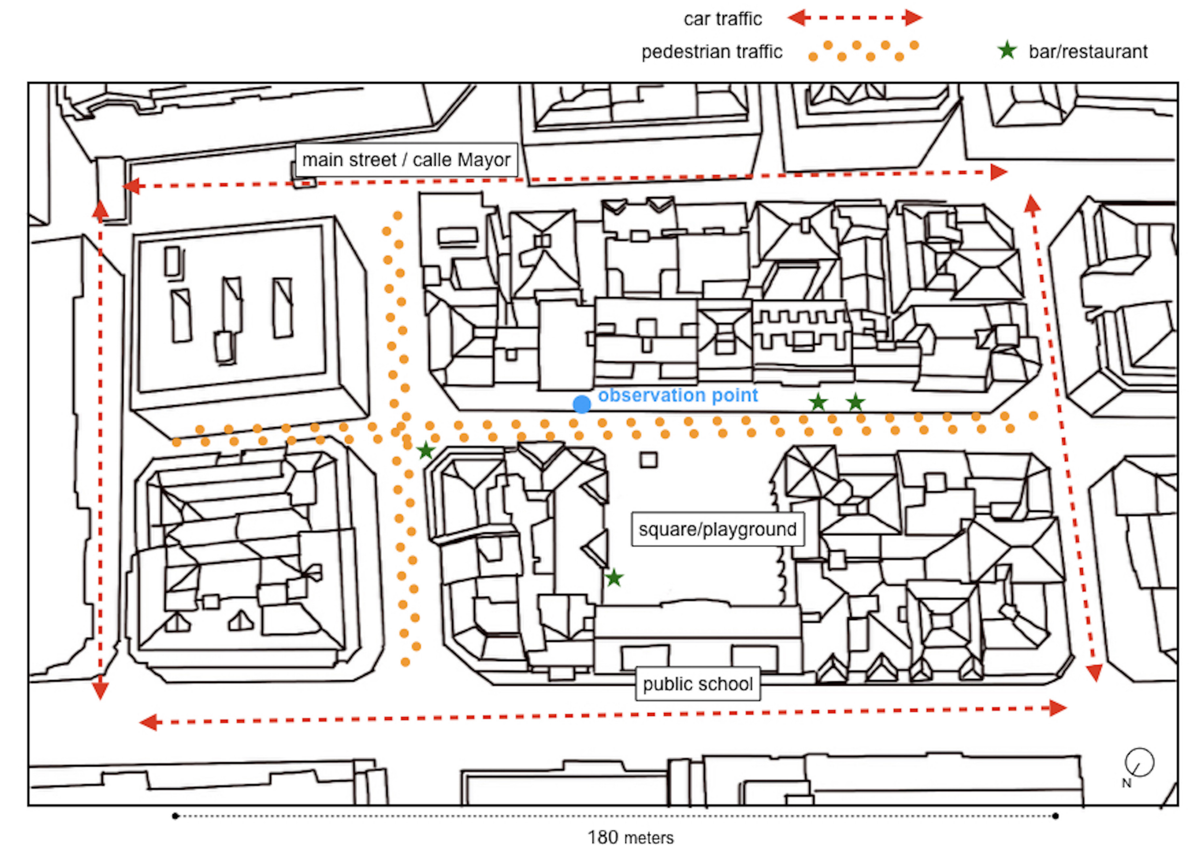


FIGURE 2 | Urban blocks and traffic at the site. The site is mainly residential, with a public school overlooking the plaza. Almost half its area is dedicated to a playground.

from a minimum of 50 to 170 mm. This means that outdoor activities are limited by rain and that architecture responds to this aspect, as we can see in the main plaza (next to the observation point, **Figure 3**) with an arcade (echoing sounds) surrounding the space and allowing urban life during rainfalls. Social life is intense, but squares and plazas are not constantly occupied, and usually people do not interact from their windows or balconies. Office work starts at 08:00–09:00, retail opens at 10:00, and lunchtime is late by European standards: at 14:00 on working days and 15:00 on weekends. Dinnertime also is late evening, at 21:00–22:00. Children attend school all day until around 17:00, when they join extracurricular activities or go to the playground if the weather allows. The local habit of joining family, friends, and colleagues for pre-lunch and pre-dinner drinks is particularly well established. Given the presence, at the site, of several bars and restaurants as well as an outdoor playground, the area of the case study tends to be crowded with children and adults during the hours that precede lunchtime and dinnertime.

Acoustic Environment

The municipality of Getxo has shown great concern with the acoustic environment. The "Noise Map" (AENA, 2013) and the "Acoustic Zonification of Getxo" (Municipal de Getxo, 2016) are two white papers emanating from the "Basque Government Law of Noise" (Jefatura del Estado, 2003). The two white papers provide a benchmark on the acoustic environment in Getxo, focusing mainly on sound levels and noise. In particular, the second document identifies the location of the site of the present study as residential. For each type of zone, there is an acoustic quality objective (AQO). Three measures are defined for different parts of the day. Ld is the A-weighted level-equivalent sound pressure level (SPL) (in dB *re* 20 µPa) during daytime, 07:00 to 19:00. Similarly, Le is for evening time, 19:00 to 23:00, and Ln is for nighttime, between 23:00 and 07:00. For the neighborhood of Las Arenas, AQO is set at Ld = 65 and Le = Ln = 55. The goal of the municipality is to lower these limits by another 5 dB in future residential developments.

The Acoustic Zonification includes a "'Noise Map of Getxo" that indicates sound levels (Total Ambient Noise) and zoning in different parts of the city. A part of the map is shown in **Figure 5**. The observation point is inside an area where Ld is indicated to be in the range of 45–50. It is equidistant from areas characterized by much higher noise levels (Ld = 60–65) and areas with higher (55–60) or slightly higher (50–55) levels of total ambient noise. However, in measurements at 50 midday recordings over the 69-day period, we found the daytime level to be 56.3 dB (A-weighted level-equivalent SPL), while the nighttime level (mean across four separate recordings toward the end of the period)

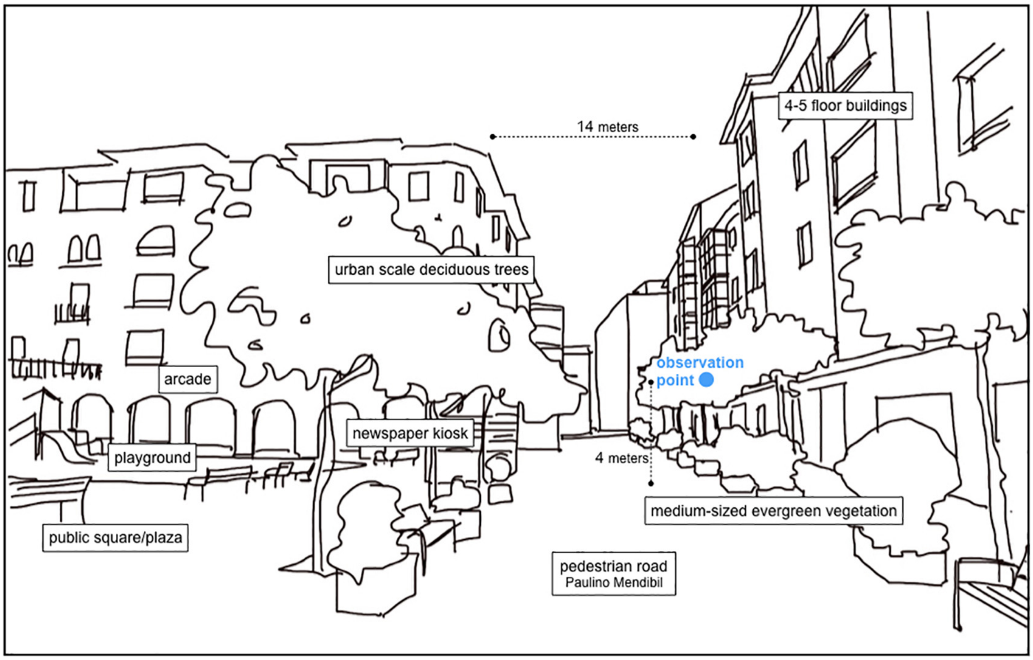


FIGURE 3 | Urban elements and proportions at the site. The buildings in the street are mid-20th century, showing brick and stone constructions with a balanced composition of walls and voids. The average height of the building is five to six floors, including the ground floor.

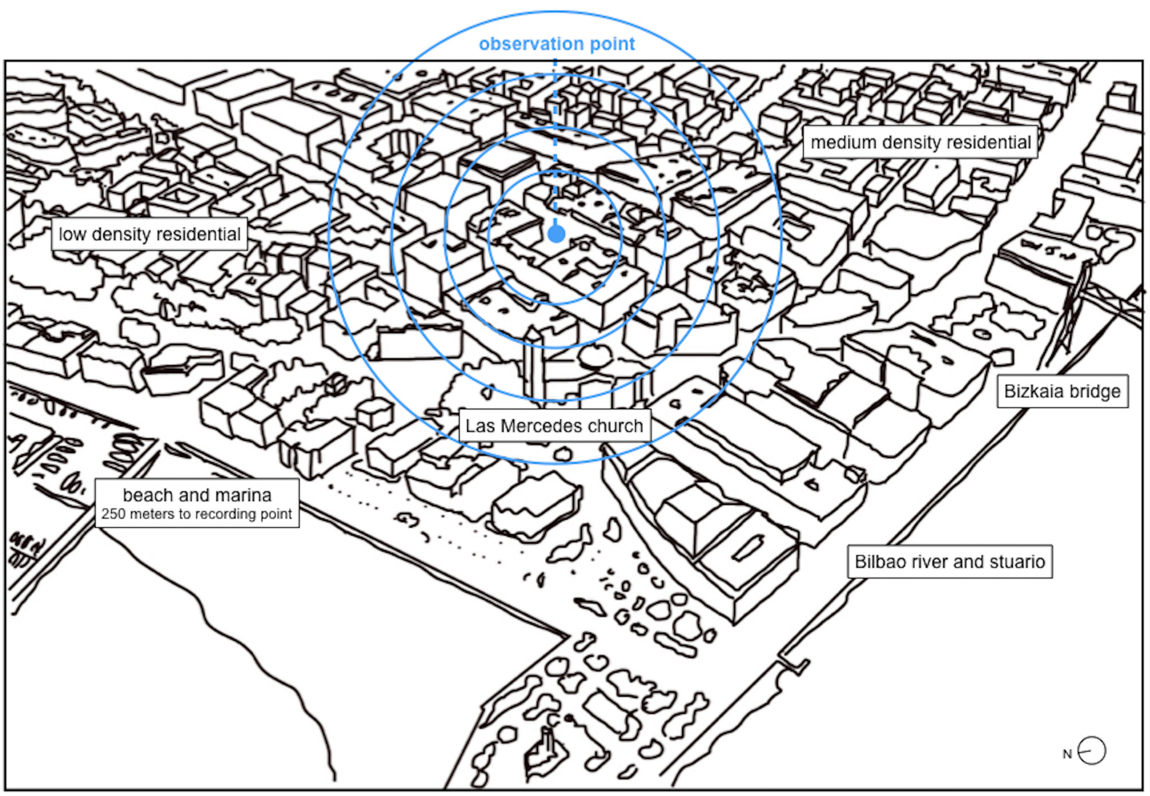


FIGURE 4 | Spatial location of the observation point.

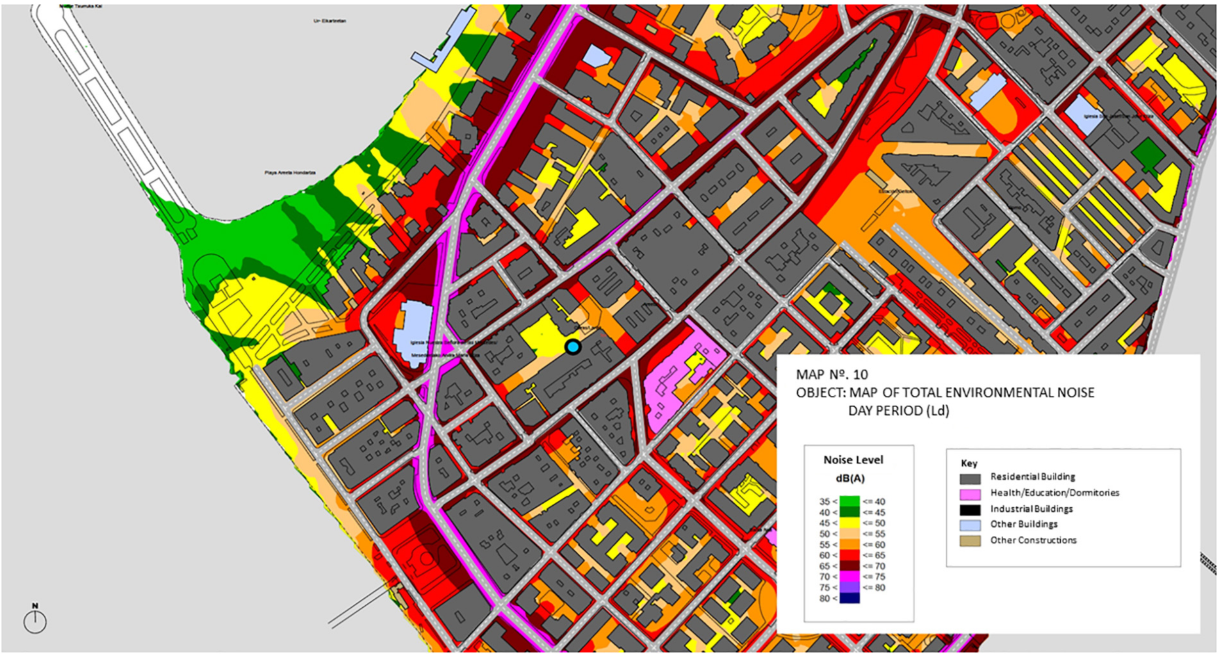


FIGURE 5 | Excerpt from Map 3 of the Acoustic Zonification document, showing the ambient levels in the neighborhood around the site. Colors of blocks indicate zoning (residential, health/mixed, industrial, commercial, and other). Colors of urban arteries indicate noise levels. Image used with permission.

was 40.0 dB. Further research might show if the slightly higher daytime levels reflect a general effect of the lockdown on changed behaviors by users.

MATERIALS AND METHODS

Audio Recordings

Audio recordings were made at the observation point (see *Site of the Case Study* section). We used a Zoom H4 recorder with integrated stereo microphones positioned at a 90° angle, with a sample rate of 48 kHz and bit depth of 24 bits. The recorder was placed in the same spot every day, at a windowed balcony, with the window open. The pre-amplifier level was set at mark 68 for the first 10 days and then adjusted to mark 65. No other position or amplifier adjustments were made. Calibration recordings and SPL measurements were made to account for this slight difference in the computational extraction of loudness (Section "Computational Descriptors" below).

Forty-two of the 50 recordings were taken between 11:45 and 14:00, six between 16:00 and 18:00, and two between 19:15 and 22:00. While there might be reasons to exclude the late-hour recordings, initial analysis revealed that the statistical results did not change in any significant way by their exclusion. Therefore, the analysis proceeded with the full set. The audio files were scrutinized and trimmed, so that a clean section of adequate duration could be taken from the beginning of each file. The excerpts needed to be sufficiently short to avoid fatiguing the volunteers in the annotation exercise (see below) yet long enough to provide representative data. We decided on a target duration of 120 s, though in three cases shorter files (90, 45, and 38 s) were deemed acceptable for inclusion. Compressed versions of the 50 soundscape recordings are included as **Supplementary Material** to the article.

Diary Notes

Diary notes, in the form of short catchphrases, were written at the time of soundscape audio recordings by the author of the recordings as a spontaneous collection of traces inspired by the experience itself. Such observational field notes (Flick, 2014) express the "researcher's own thoughts, feelings, impressions and insights" (Maharaj, 2016) and highlight elements that from time to time appeared to be the most striking. Occasionally, they serve as a poetic deepening of the acoustic and visual experience. Furthermore, they tend to express overall changes in the neighborhood as witnessed by the researcher that neither visual nor sonic material in their own fully capture. Gehl and Svarre (2013) identify several tools to extract a deeper knowledge of the use of public space in urban research. Among others, they describe the action of keeping a diary that can "register details and nuances about the interaction between public life and space, noting observations that can later be categorized and/or quantified" (idem, page 24).

The diary notes and recordings have been continuously shared on the blog of the first author (Lenzi, 2020) and via social media channels. The complete list of diary notes is included as **Supplementary Material**.

Soundscape

We are interested in the relationship between perceived sound sources and the perception of the soundscape as a whole (ISO, 2014). In order to capture the human perception of the acoustic environment at the site, we gathered a group of 14 experts in soundscape, music, and architecture and tasked them to analyze the set of soundscape recordings. Two separate procedures were carried out: evaluations of soundscape quality and annotations of sound sources. The three authors participated as well, two of whom having knowledge about the site. Among the others, six were professionals in architecture or music and five were graduate students in these fields. The median age in the group was 40 years, in a range between 24 and 53, with equal numbers of men and women. Having received full information all declared consent in writing before commencing, the two tasks were completed several weeks apart. The instructions are available in **Supplementary Material** to this article. The collection of diary notes and soundscape recordings did not require an ethics approval from the institution of the first author. The procedures for evaluation and annotation were approved by the Research Ethics Committee of City University of Hong Kong (ref. 13-2020-08-E).

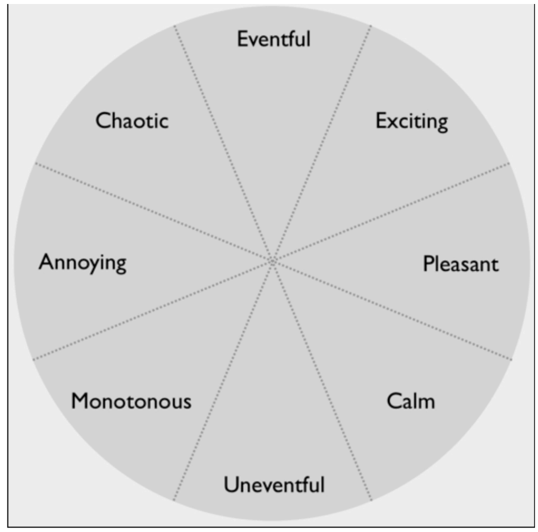


FIGURE 6 | Graphical user interface for the ratings of soundscape quality. The eight adjectives in the circumplex yield two dimensions, Pleasantness and Eventfulness, which are considered to be orthogonal.

Evaluations of Soundscape Quality

Eleven experts rated the 50 soundscape recordings using an adapted version of the Swedish Soundscape Quality Protocol (SSQP; Axelsson et al., 2010). This task calls for a mode of semantic listening (Chion and Gorbman, 2009; Lindborg, 2019) and took just under 2 h to complete. The SSQP includes eight adjectives: pleasant, exciting, eventful, chaotic, annoying, monotonous, uneventful, and calm. These words were originally selected for representing equidistant and equally strong semantic concepts, spanning a circumplex with the horizontal axis labeled Pleasantness and the vertical axis Eventfulness. In our adapted version, the circumplex and adjectives are presented as shown in **Figure 6**. While listening to the recordings (in individually randomized order), the rater continuously pointed with the computer mouse to the adjective that "best described what you feel the soundscape is like" (Axelsson et al., 2010). From the response time series, the mean angle and distance from the center were calculated for each of the 50 soundscapes across raters. This yielded values for Pleasantness and Eventfulness for each soundscape recording. The agreement was good among the expert raters (N = 11), as indicated by Cronbach's alpha = 0.87 for Pleasantness and 0.83 for Eventfulness.

Annotations of Perceived Sounds

Twelve experts scrutinized the same set of soundscape recordings to identify individual sound events and describe them according to their perceived source. Labels were recorded as free-form text, almost all consisting of three words or less, along with begin and end times. This task called for a mode of causal listening (Chion and Gorbman, 2009; Lindborg, 2019) and was considerably more time-consuming than the previous task. Three of the experts completed the whole set of 50 soundscapes in 4–5 h. Others completed on average 26; no one less than 20. Each made between 7 and 22 labels per recording (median = 14), producing a total of 5,581 annotations of perceived sounds.

The labels were pre-processed by removing non-letter symbols (such as question marks, citations, parentheses, and trailing spaces) and transcribing to lowercase. There were a total of 10,441 individual words, out of which less than a thousand were unique. The 23 most common were as follows: birds (4.3%); voice (3.3%); door (3.0%); bird (2.8%); car (2.6%); voices (2.5%); dog (2.2%); talking (2.0%); human (1.7%); distant, child (1.6%); traffic (1.5%); chirping (1.4%); man, noise, male (1.3%); woman (1.2%); barking (1.1%); footsteps, closing (1.0%); and passing, female, children (0.9%).

A taxonomic classification with interconnected levels can serve as a bridge between a detailed description and a holistic description. To build a taxonomic classification of perceived sounds, we chose an empirically grounded approach (Scott-Ram, 1990; Atkinson et al., 2000; see also Lindborg, 2016). With the frequency counts in mind, we sorted each of the original 5,581 annotations within exactly one of the following 22 categories constituting Level 1: bird, animal, geophony, conversation, communication, body, individual, group, crowd, music, onomatopoeia, noise, action, object, material, signal, wheels, vehicle, machine, acoustic, spatial, and rest. Keywords for inclusion or exclusion speeded up the process so that ∼55% could be automatically matched, while the remainder required individual attention. Next, we developed Level 2 of the taxonomy and automatically sorted (by keywords) each of the Level 1 categories into exactly one of the Level 2 categories: nature, voice, people, sonic, physical, traffic, and modifiers. Finally, we defined three categories in Level 3 Natural, Human, and Technological, to correspond to the typology for sounds in soundscape first advanced by Schafer (1977) and developed by, e.g., Krause (2008) and Axelsson et al. (2010). This process yielded the taxonomic classification illustrated in **Figure 7**, with examples given in **Table 2**. The agreement was very good among the expert raters

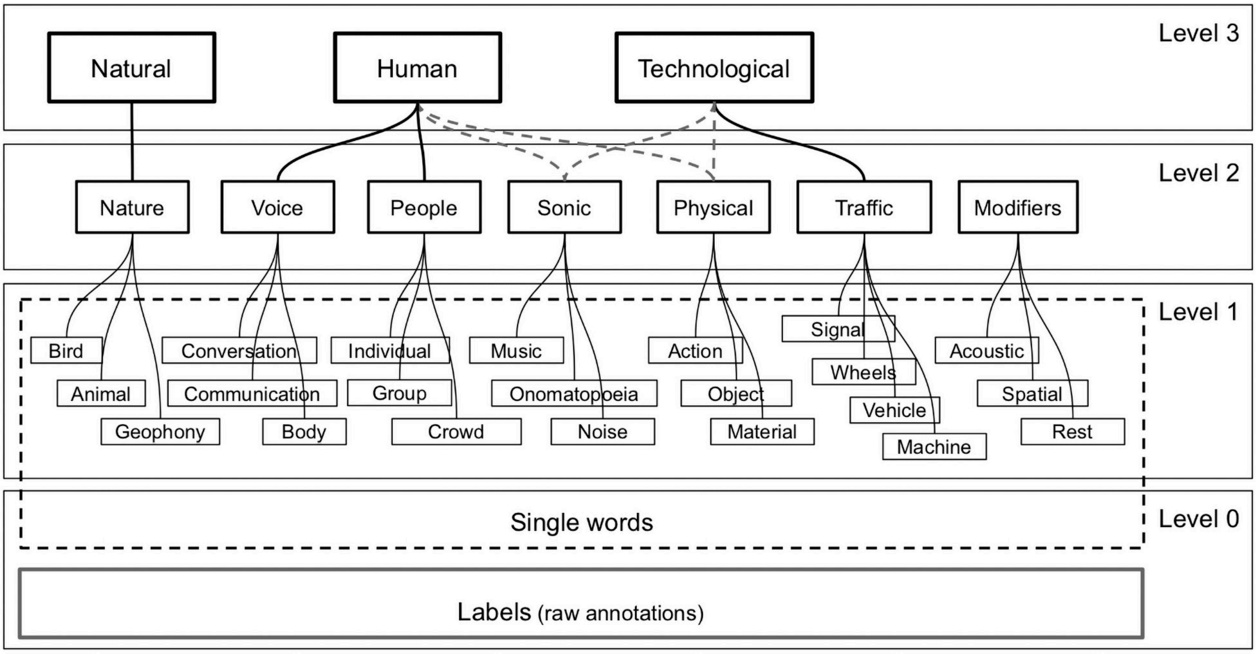


FIGURE 7 | Schematic representation of the Taxonomy of Sound Sources.

(N = 12), in terms of the proportions of Natural, Human, and Technological sounds in their annotations of soundscapes, as indicated by Cronbach's alpha = 0.92, 0.95, and 0.93 for the three categories, respectively.

Note that unique associations between categories in Level 2 and Level 3 could not be made. For example, a perceived sound with a label including the word "music" might refer to a recording played on a radio and thus sort under Technological (Axelsson et al., 2010). Or it could be someone playing an instrument or singing and thus evidence of someone's action with an object or their own body and thus sort under Human. Interpretative challenges such as these point to the difficulty of marrying a cladistic taxonomic classification (bottom-up) with a previously given typology (top-down). A datafile including all the annotated labels and levels of the Taxonomy of Sound Sources is available in the **Supplementary Material**.

From the taxonomy, we report results on four descriptors.

Natural, which relies on annotation of perceived sound sources that are related to birds of different kinds, as well as insects and geophony, water, waves, and so forth. Despite there being good reasons why domestic animals should not be categorized as part of the biophony (Schafer, 1977), we decided to include dog barks in this category to keep the taxonomy parsimonious.

Human, which covers a large range of sounds perceived to be produced by the human body, i.e., footsteps, speaking, and other vocalizations such as coughing and laughter. Please note that we decided to exclude music instruments and indeed any kind of sonic objects that might be manipulated by humans, since the source of such sounds is outside of the human body.

Technological, which includes sounds produced by machines, tools, cars, traffic, and so forth, a.k.a., technophony. While being propelled by a motor or an external energy source is a main characteristic of this category, we also included sounds from non-motorized wheeled vehicles such as bicycles, skateboards, and delivery carts.

Perceptual normalized difference soundscape index (pNDSI). We introduce a perceptual counterpart to NDSI (Kasten et al., 2012; see below), defined as

$$\text{pNDSI} = (\text{Natural} - \text{Human})/(\text{Natural} + \text{Human}) \quad (1)$$

where each variable is a time series generated by the taxonomy. With Natural and Human in the range [0–1], pNDSI is in the range [−1 to 1]. A value close to −1 indicates that the soundscape is dominated by sounds associated with humans, and a value close to 1 indicates prevalence of natural sounds.

Computational Descriptors

We used six computational indices from soundscape and bioacoustics research (Sueur, 2018; Kang et al., 2019) obtained in R (R Core Team, 2020).

NDSI (Kasten et al., 2012; Sueur et al., 2019) estimates the level of anthropogenic disturbance on a natural environment with the ratio:

$$\text{NDSI} = (b - a)/(b + a) \quad (2)$$

where b is the sound intensity in the 2–8 kHz range (where biogenic sounds are prevalent) and a is the intensity in the 1–2 kHz range (where anthropogenic or mechanical sounds are most prevalent). NDSI is scaled between −1 and 1, with 1

TABLE 2 | Part of the Taxonomy.

Annotations, raw (random sample)	Keywords/inclusion	Keywords/exclusion	Level 1	Level 2	Level 3
Birds and voices almost muttered, bird chirps, dog roaming, bird chirp, chirp birds, bird chirps, birds plenty, birds intense, maybe it is raining, cat meow, dog yelp, dog's footsteps, birds faint, cocorita, bird chirping, slightly differently, dog walking,	animal, bark, bees, bids, bird, birs, brids, burds, cat, chirp, cocorita, crow, dog, flap, fly, gull, gust, gut, insect, miaow, mosquito, nightingale, pigeon, rain, sea, seagul, seagull, trill, tweet, waves, wind, wings, yelp	Crowd, train, winding, window	Bird (856), animal (269), geophony (36)	Nature (1189)	Natural
Chatting + children voices, people talking and birds, woman talking with man, voice, female, child talking, human male voices chatting, adult footsteps, steps, mam speak with her children, Girl shout, woman or child humming, kid's voice, playful, woman voice, children voices, clear conversation, Man talks, kids voices (playing, shouting), female voices,	adieu, burp, bye, chat, chuckl, clap, complain, convers, coo, cough, count, creaing, cry, dialog, exclaim, exhalation, foot, footstep, giggl, heel, humm, laugh, nose, running, scream, screem, shout, shriek, singing, sneez, sole, speak, step, stpes, talk, tantrum, throat, ululat, vioce, vocal, voice, walk, whin, whisper, whist	Passing, passing, closing, rising, creasing, machine, music, thump, traffic, scooter, wheel, cart, passing	Conversation (358), communication (826), body (265)	Voice (1490)	Human
children, indistinct voices, baby babbling, little scream, rather loud female voice, many voices, human voice (close), whining/seagulls, distant, young child, human voice, kids, shouting, voice, male, close, kid's voice (faint), amplified female, children chasing, human voices, close, human whistle, man sings	adult, adults, anthropic, babbl, babies, baby, backg, boy, chid, child, children, crowd, din, faemale, family, father, femail, female, folk, girl, group, humans, kid, kids, male, man, market, mather, men, mom, mother, mumbl, murmur, neighbors, neighbors, owner, parents, people, person, police, somebody, someone, women	Human, instrument, movement, winding, descending, sliding, bounding, receding	Individual (922), group (147), crowd (91)	People (1356)	
Click, dops-like, human activity sounds, starts music, creaking, blinds closing sound, human activity/dishes, melody continues, clicking, sound by TV, scan sound, background buzz, clacking, radio, indistinct human and non-human noise, squeaking, melody, nice, like playing background music, indistinct non-human, clink, distant	accordeon, airflow, bang, bash, boom, boum, bump, burst, buzz, carillon, choir, chor, chorale, clack, clang, clash, classic, click, cling, clink, cordion, crash, creak, crink, cump, doum, drip, flute, guitar, harmonica, howl, hum, instrument, jing, jingl, knoc, major, melod, music, nois, organ, patter, pjoff, puff, radio, rhythm, rumbl, sbam, scan sound, schreech, scratch, screech, shot, shrill, slam, snap, song, sound, sqee, squeak, squee, squirr, squoink, swish, teardrop, thud, thump, tick, tone, trumpet, tv, undefined, undetermined, unidentified, vent, whir, whosh, wind, woosh		Music (116), onomatopoeia (517), noise (475)	Sonic (1108)	NA

(Continued)

TABLE 2 | Continued

Annotations, raw (random sample)	Keywords/inclusion	Keywords/exclusion	Level 1	Level 2	Level 3
Door slam, bash car door, doorlock again, cutlery plates, gate, Something swipes, rattling, high pitch, unlocking, car door, hit, indoor, keys jingling, hits (glass falling), objects bashing (faint), scratching of something on the ground and, pluck, cups, plates, 2 rattles, hammering	activ, ball, bike, blind, bottle, bounc, break, buck, can, chain, chair, crockery, cup, cut, cutlery, dish, door, drag, driv, drop, fall, flutt, gate, glass, hammer, hit, house, iron, item, key, kick, kitchen, knick, lagguage, lock, material, metal, moving, newspaper, object, paper, pladtic, plastic, plate, play, pluck, pull, rattl, roll, rubb, saw, scissor, scrap, shak, shaker, shuffl, shut, shutter, sifting, solder, something, spoon, start, steel, stomp, stone, strap, sunblid, swip, tennis, throw, tool, toy, trunk, water, wood, work, wrapper	Scan	Action (557), object (564), material (106)	Physical (1243)	NA
Stroller wheels, van parked, engine still running, bus slowing down, beep, bike in water puddle? Tram arrival, engine truck close, ambulance siren, indistinct voices/busy people chatting, siren faint, bus whistling breaks(?), trolley little wheels - metallic, plane, car brake, battery car, bip, no-human, gear, playing with skate	alarm, ambulance, atm, beep, bell, bicycle, bike, bip, brakes, braking, bus, calls, car, cart, chart, chopper, church, clacson, claxon, construction, delivery, drone, electr, engine, gear, hawking, honk, horn, jet, machine, mechanic, message, motocycle, motor, motorbike, motorcycle, mower, parked, phone, plane, raffic, revs, revv, ringtone, road, scooter, signal, siren, skate, skatebaord, skating, stroller, traffic, train, tram, trolley, trolly, truck, van, vehicle, warning, weel, wheel	Atmosphere, motorbike, cart, door, key, carrousel, carillon	Signal (259), wheels (162), vehicle (562), machine (33)	Traffic (1062)	Technological
Mic handling noise, bump on mike, microphone manipulation, around eight beats, regular on a pitched drum (blank), noise on microphone, undetermined urban noise, croak/fart	accelerat, acoustic, approach, arrival, audible, away, backg, behind, between, bit, blank, circular, city, clear, clos, creasing, departure, distanc, distant, doppler, echo, exotic, faint, far, foregr, freq, from, front, hard, heavy, high, hollow, indistinct, indoor, intense, large, light, little, loud, louder, lound, low, medium, mic, mike, mobile, movement, moving, muffl, near, open, passing, past, pitch, quiet, reced, reverse, soft, soundscape, sparse, street, strong, surface, sustain, toward, undistinct, undistinguished, urban, very, volume, weak	Rhythmic	Acoustic (445), spatial (636), rest (45)	Modifiers (1129)	NA

In this example, rows correspond to the seven categories of Level 2. The leftmost column shows a random sample of the labels exactly as written. They are filtered through Level 0 (not shown) and auto-matched via the inclusion and exclusion keywords to aggregate into Level 1. The procedure is repeated to yield Levels 2 and 3, shown in the last two columns. Numbers in parentheses are counts.

indicating pure biophony (cf. Sueur, 2018 p. 491–2; Remote Environmental Assessment Laboratory [REAL], 2020).

Loudness (N). The Loudness model (Zwicker and Fastl, 2013) was originally limited to sounds of short duration initiated, has been extended to model the perception of sounds with time-varying and complex spectra, and has been widely employed in soundscape studies (Axelsson et al., 2010; Davies et al., 2013; Lindborg, 2015; Aletta et al., 2016a; Lindborg, 2016; Anikin, 2017; Kang et al., 2019).

SPL. A-weighted SPL in dB *re* 20 μPa is reported to facilitate comparisons with other research.

Loudness variability (N_{10-90}) is the difference between the loudness exceeded 10% of the time and that exceeded 90% of the time. While the former captures short and loud sounds, the latter captures the background. The range indicates the amount of foreground sources emerging from the background (Axelsson et al., 2010; Lindborg, 2015, 2016).

Acoustic richness (AR) is calculated from amplitude (M) and acoustic entropy (Ht) and ranked over several files (Sueur et al., 2019). M is scaled between the median Hilbert amplitude and the maximum. Ht increases with signal entropy, or heterogeneity, so that a higher value indicates a richer acoustic environment.

Acoustic complexity (ACI) is an index based on the "observation that many biotic sounds, such as bird songs, are characterized by an intrinsic variability of intensities, while some types of human generated noise (such as car passing or airplane transit) present very constant intensity values" (Pieretti et al., 2011 p. 869; Sueur et al., 2019).

RESULTS

Table 3 gives mean values of the soundscape descriptors determined from evaluations and annotations by the expert group (N = 15) and extracted computationally. Cross-correlations for Pleasantness and Eventfulness against other variables were calculated using Spearman's rank-order correlation (since variable distributions were non-normal) and interpreted after Dunn–Šidák correction for each pairwise comparison significance level (α = 0.0022, for 23 comparisons at α_{FWE} = 0.05). For Pleasantness, there was a significant positive association with the amount of perceived Natural Sounds (ρ = 0.44, p = 0.0013) and negative associations with technological sounds (ρ = −0.44, p = 0.0015), SPL (ρ = −0.53, p = 0.0001), and Loudness Variability (ρ = −0.51, p = 0.0002; larger range correlated with less pleasant soundscape). Eventfulness was associated with SPL (ρ = 0.44, p = 0.0016), Loudness (ρ = 0.50, p = 0.0003), and Loudness Variability (ρ = 0.48, p = 0.0005). No other correlations were significant at the predetermined level. These findings are in line with previous soundscape research (e.g., Cain et al., 2008; Axelsson et al., 2010; Davies et al., 2013).

We proceeded with an analysis of the change over time for the 13 descriptors. A multivariate analysis of variance (MANOVA) revealed that there were differences in the data [Pillai's trace = 0.52, F(13) = 3.09, p = 0.0037**]. Spearman's ρ was used to evaluate univariate correlations against a dummy variable for time, i.e., the 69 days. Results are given in **Table 4**, and **Figure 8** plots the development of the variables against the three main phases of the lockdown.

On-site diary notes were compared with perceived sound sources in Level 0 (original annotations) and Level 3 (categories of Natural, Human, Technological sounds) of the Taxonomy. This highlighted correspondences between the elements of the soundscape as they emerge from annotations and the subjective impressions as they were noted by the first author during the recording process throughout the different phases. Additionally, notes were compared with the evolution of the perceived quality of the soundscape (see *Evaluations of Soundscape Quality* section) over time and during specific days. See **Table 5**, which shows details for soundscapes captured on 8 of the 69 days.

We found several correspondences between the author's and evaluators' perception of the soundscape in terms of Pleasantness and Eventfulness. Such correspondences are also sustained by the Taxonomy obtained from the annotations. On Day 4 of the confinement, the first author notes "I never noticed how much human voice resonated in the little plaza in front of our window. [. . .]" Interestingly, in a Phase 1 marked by restrictions to human mobility and activity, the predominant perception of the soundscape of Day 4 is "extremely annoying" (see **Figure 9**). Annotations for the same day register a clear predominance of Human (annotated as a source of sound 28 times) over Natural (15) and Technological (12) sounds with the indication, at Level 0 of the taxonomy, of words such as "people talking," "voices," "whistle," "human conversation," "human voices," and "dialog between man and woman." On Day 14, an "Extremely Calm" day for soundscape Evaluation is described as "near silence" in the author's notes, while annotations show a prevalence of technological sounds further described as "traffic," "distant traffic," and "car passing distant" (or "medium distance"). Day 16 stands out in the diary notes, in the context of the restrictions imposed by Phase 1 ("making noise is feeling alive") as well as in the evaluators' assessment ("clearly annoying"). Annotations register a clear predominance of technological (32) over Natural (13) and Human (7) sounds. As mentioned, Phase 2 was characterized by a strengthening of restrictive measures both to mobility and human activity. Day 19, the third day of Phase 2, is described as "Extremely Uneventful" by evaluators, while annotations highlight both human and technological sounds, the latter mainly referring to indoor activity ("glass bottle," "objects on surface," "hit plate," etc.). Notably, the diary note for that day read "In the silence, someone's getting ready for lunch." Day 28 registers a clear predominance of Human (39) over Natural sounds. On the contrary, diary notes define it as a fully natural experience ("Birds, birds, birds."). Evaluators describe it as "Extremely Pleasant," leaving us in the doubt as to the perceptual dimension that is responsible for the pleasantness. The first day of Phase 3, when restrictions start to be lifted with kids being allowed outdoors for 1 h a day, marks a turning point in the lockdown diary notes ("Is it excitement I am hearing in the air?"). Evaluators define the same day as "Extremely Chaotic," while Human and Natural sounds appear almost equally in the annotations (27 to 26). As restrictions are progressively lifted, Human sounds emerge as the prevailing source in the soundscape evaluations, as well in the diary notes. On Day 48, the author

TABLE 3 | Mean values for 13 descriptors of 50 soundscapes.

Days	Lockdown	Pleasantness	Eventfulness	Natural	Human	Technological	NDSI	pNDSI	Wheels vs. Vehicle	Loudness (sone)	SPL (dBA)	Loudness variability	Acoustic complexity	Acoustic richness
1	Phase 1	−0.62	−0.07	0	0	1	−0.13	0	−1.00	2.22	62.4	4.61	12,993	0.02
2	Phase 1	−0.02	−0.58	1.13	0	0.25	−0.35	1	0	0.79	55.6	0.36	12,762	0.49
3	Phase 1	−0.48	0.33	0.2	0.3	0.3	0.67	−0.20	−0.10	1.09	58.3	2.4	13,651	0.08
4	Phase 1	−0.62	0.08	0	1	0.29	−0.12	−1.00	−0.29	1.03	57.6	0.76	13,315	0.61
5	Phase 1	−0.44	0.28	0	1.17	0.33	−0.60	−1.00	−0.17	0.85	55	0.32	13,028	0.08
6	Phase 1	−0.51	−0.31	0	0.57	0.57	−0.66	−1.00	−0.57	0.65	57.8	0.34	13,028	0.26
7	Phase 1	−0.53	−0.21	0	1	0.17	−0.71	−1.00	−0.17	0.71	55.4	0.75	12,963	0.17
8	Phase 1	0.46	−0.27	1	1.14	0	−0.49	−0.07	0	0.52	51.3	0.26	13,409	0.11
9	Phase 1	−0.37	0.45	1	0.33	0.56	−0.63	0.5	−0.44	1.88	58.9	1.63	13,470	0.28
10	Phase 1	−0.60	−0.14	0.63	0.88	0.63	0.51	−0.17	0	0.59	53.2	0.9	13,314	0.11
11	Phase 1	−0.46	0.4	0.89	0.22	0.56	−0.19	0.6	−0.44	1.81	56.1	0.93	14,993	0.11
12	Phase 1	−0.56	−0.01	0.13	1.25	0.13	−0.44	−0.82	0	0.74	56.4	0.38	13,442	0.47
13	Phase 1	−0.53	−0.16	0.25	0.88	1	−0.43	−0.56	0	0.6	53.1	0.74	13,292	0.02
14	Phase 1	0.31	−0.55	1.2	0	0.8	−0.41	1	−0.80	0.36	50.7	0.33	12,980	0.05
15	Phase 1	0.07	−0.53	0	0.56	0.22	−0.27	−1.00	−0.22	0.9	50.9	0.51	14,294	0.08
16	Phase 1	−0.58	0.08	0.57	0	0.86	−0.74	1	0.43	1.79	61.3	1.73	13,126	0.49
17	Phase 2	0.02	−0.61	0.63	0	0.5	−0.56	1	−0.25	0.56	53.6	0.22	12,744	0.35
18	Phase 2	−0.50	0.25	0.11	0.33	0.78	−0.57	−0.50	0	1.65	60.4	2.04	13,150	0.24
19	Phase 2	−0.23	−0.60	0	0	0.64	−0.44	0	−0.64	0.28	50.9	0.62	12,839	0.01
20	Phase 2	0.37	−0.46	0.3	0.9	0.2	−0.27	−0.50	−0.20	0.26	49.7	0.14	13,408	0.05
21	Phase 2	−0.14	−0.55	0.57	0.57	0.43	−0.52	0	−0.29	0.79	52.8	0.27	13,469	0.07
22	Phase 2	0.2	−0.54	1	0.5	0.17	−0.24	0.33	0	1.21	53.6	0.45	14,548	0.17
23	Phase 2	−0.58	0.25	0.8	0.7	0.2	−0.24	0.07	−0.20	1.66	57.4	0.98	13,754	0.37
24	Phase 2	0.47	−0.35	0.3	0.6	0.2	−0.45	−0.33	−0.20	1.5	56.2	0.64	13,660	0.02
25	Phase 2	−0.09	0.57	0	0	0	−0.54	0	0	0.71	52.1	0.37	13,277	0.04
26	Phase 2	0.31	−0.53	0.29	0.71	0	−0.30	−0.43	0	0.16	47.4	0.07	12,845	0.02
27	Phase 2	−0.36	−0.46	0.71	1	0	−0.60	−0.17	0	0.5	51.7	0.39	13,477	0.1

(Continued)

TABLE 3 | Continued

Days	Lockdown	Pleasantness	Eventfulness	Natural	Human	Technological	NDSI	pNDSI	Wheels vs. Vehicle	Loudness (sone)	SPL (dBA)	Loudness variability	Acoustic complexity	Acoustic richness
28	Phase 2	0.64	−0.23	1	0.71	0	0.67	0.17	0	0.27	49.2	0.22	13,736	0.03
29	Phase 2	−0.57	−0.08	0.29	1.29	0.57	0.24	−0.64	−0.43	0.57	53	0.36	13,833	0.15
30	Phase 2	−0.22	−0.59	0.33	1.33	0.67	−0.27	−0.60	−0.17	1.22	54.1	0.44	13,700	0.27
31	Phase 2	−0.40	−0.41	0.57	0.57	0.43	−0.50	0	−0.43	1.44	56.8	0.6	12,950	0.08
34	Phase 2	−0.10	−0.59	0.22	0.22	0.44	−0.46	0	−0.44	0.72	55.5	0.34	13,507	0.44
35	Phase 2	0.55	0.31	0.83	0.33	0	−0.11	0.43	0	0.61	52.9	0.21	13,475	0.23
37	Phase 2	−0.49	−0.21	0	0.44	0.56	−0.39	−1.00	−0.33	0.76	56.9	0.31	13,307	0.65
38	Phase 2	0.41	0.46	1	0	0.43	0.53	1	−0.29	0.76	54.9	0.41	13,837	0.38
39	Phase 2	0.47	−0.33	0.86	0.57	0.14	−0.02	0.2	−0.14	0.66	52.3	0.42	13,849	0.15
41	Phase 3	−0.52	0.33	0.5	0.13	0.88	−0.54	0.6	0.13	2.15	62.2	2.31	13,415	0.13
46	Phase 3	−0.12	0.61	1	1	0	0.01	0	0	1.63	57.3	0.78	13,968	0.44
47	Phase 3	−0.53	0.4	0	1.38	0.38	−0.26	−1.00	0.25	7.55	66.7	6.89	14,424	0.22
48	Phase 3	−0.30	0.55	0.17	1.17	0.17	−0.51	−0.75	−0.17	3.09	60.2	2.94	14,860	0.38
49	Phase 3	−0.64	0.37	0	0.33	0.67	−0.59	−1.00	0.67	2.3	60.6	2.3	13,871	0.45
51	Phase 3	−0.39	−0.49	0	1	0.11	−0.29	−1.00	0	1.61	57.2	0.76	13,987	0.61
52	Phase 3	−0.62	0.03	0	0.67	0.11	−0.24	−1.00	0	0.72	49.8	0.43	14,499	0.04
54	Phase 3	−0.36	0.44	0	0.7	0.1	−0.61	−1.00	0	1.35	55.3	0.86	13,701	0.14
56	Phase 3	−0.31	0.48	0	1	0.29	−0.55	−1.00	−0.14	1.01	55.6	0.57	13,266	0.35
60	Phase 3	−0.30	0.45	0.33	1	0.33	−0.47	−0.50	0.17	1.43	57.1	0.94	13,549	0.29
62	Phase 3	−0.45	0.41	0.25	1	0.13	−0.43	−0.60	0	2.1	56	0.65	13,968	0.45
64	Phase 3	0.02	0.54	0.13	1	0	−0.22	−0.78	0	0.69	53.2	0.39	14,037	0.26
65	Phase 3	−0.56	0.15	0.33	0.56	0.33	−0.25	−0.25	−0.22	0.61	52.5	0.2	13,769	0.21
69	Phase 3	−0.33	0.54	0	1.14	0.29	−0.49	−1.00	−0.29	2.3	59.6	1.25	13,964	0.58

Pleasantness and Eventfulness (evaluations by expert group, N = 11) are in the range [−1 to + 1]. Natural, Human, Technological, pNDSI, Wheels vs. Vehicles (annotations by expert group, N = 12), SPL in A-weighted dB re 20 μPa, Loudness and Loudness Variability in sone. For Acoustic Complexity and Richness, see Section 2.5.

TABLE 4 | Correlations between 13 perceptual and computational soundscape descriptors and a dummy variable for time.

Descriptor	S	ρ	p
Pleasantness	18,580	0.108	0.45
Eventfulness	12,942	0.379	0.007**
Natural	23,587	−0.133	0.36
Human	15,342	0.263	0.065.
Technological	26,751	−0.285	0.045*
NDSI	19,872	0.0458	0.75
pNDSI	25,947	−0.246	0.085.
Wheels vs. Vehicle	14,637	0.297	0.036*
Loudness (sone)	16,442	0.21	0.14
SPL (dBA)	20,736	0.00427	0.98
N10m90	19,764	0.0509	0.72
Acoustic Complexity	9,382	0.549	0.00005***
Acoustic Richness	15,053	0.277	0.051.

*S is the value for Spearman's test; ρ is the rank-order correlation; p is the probability, if the null hypothesis of independence were true, of a result as extreme as the one obtained. By convention, a p-value lower than 0.05 indicates a significant association between variables. Asterisk codes for degree of significance: ***p < 0.001; **p < 0.01; *p < 0.05.*

notes that "{. . .} little by little, people are taking back the streets." Human sources are clearly predominant (87) over Natural (23) and Technological (4) sources. Evaluators define the soundscape as "Strongly Eventful." On Day 60, toward the end of our recordings and with most of the restrictions lifted, with "[. . .] small retailer, hairdressers, hospitality open and people happily and maybe unwisely taking the road [. . .]" and the soundscape is "clearly chaotic" with a prevalence of Human and Technological sounds over Natural source.

The results obtained through quantitative and qualitative analyses were interpreted in relation to the five assumptions we had previously made, to recall, (1) loudness will decrease; (2) machinery and human interaction will decrease; (3) human outdoor activity will decrease; (4) birdsong will increase; and (5) transport will shift from fuel vehicles to light mobility.

1. Over the 69 days covered in the study, overall Loudness did not significantly change ($\rho = 0.21$, $p = 0.14$ n.s.; likewise for SPL), contrary to expectations. Across the whole period, the average daytime sound level was 56.3 dBA, which is in line with the pre-lockdown municipality measurements of urban noise levels at the site (see *Acoustic Environment* section). The stable Loudness, contrary to expectations, might be explained by changes in people's behavior, evidenced in the significant decrease in Technological sounds ($\rho = -0.285$, $p = 0.045$*) being to a sufficient extent compensated by the increase in Human sounds ($\rho = 0.263$, $p = 0.065$). See **Table 3** and **Figure 8** (fourth row).
2. The decrease in Technological sounds during the time period, which was expected, indicates that the mandated restrictions on traffic circulation and human activities (with, for example, the temporary suspension of all construction works) had a noticeable influence on the soundscape at the site. Despite the site being a pedestrian area, the nearby main road is characterized by both private and public traffic circulation that can be heard from Calle Paulino Mendibil. Additionally, and perhaps due to the absence of direct sources in traffic noise at the site, the occasional construction works have a notable impact on the local soundscape. See **Figure 8** (second row, right).
3. Human activity increased at the site during the studied period, as evidenced both in the diary notes and by the slight increase in sounds from humans ($\rho = 0.25$, $p = 0.076$). Human is a category in Level 3 that aggregates annotations from two Level 2 categories: Voice and People. The former showed no change over time ($\rho = 0.16$, $p = 0.25$ n.s.), while the latter increased ($\rho = 0.35$, $p = 0.013$*). A closer look to the evolution in the perception of Human sounds during the different phases can help explain the results. **Figure 8** (second row, middle) suggests that people's activity level increased steadily throughout the time period, as evidenced by Human sounds, and was the highest in Phase 3. The diary notes substantiated this interpretation. The word "children" progressively appears from Phase 3 onward, when children started to be allowed outdoors after having been confined indoors in the first two phases.
4. The results in regard the expected increase in birdsong were not conclusive. We have already noted during the development of the Taxonomy that the most commonly annotated word overall was "bird," which (together with "birds") appeared in ∼7% of the original labels. The amount of perceived bird sounds did not change significantly over time ($\rho = -0.09$, $p = 0.54$ n.s.), as evidenced by annotations in the Level 1 category Bird. Neither did the Level 3 category Natural show a significant trend overall. However, inspecting **Figure 10**, there was an increase during Phases 1 and 2, followed by a much lower level in Phase 3. This might be explained by factors regarding avian activity: seasonal shifts, mating periods, and noon being something of a siesta time for birds See **Figure 8** (second row, left). As for the computational descriptors, AR increased slightly over time ($\rho = 0.28$, $p = 0.051$), while Loudness Variability did not change appreciably (see **Figure 8**, fifth row), and the two NDSIs (NDSI and the proposed perceptually based pNDSI) showed similar patterns and no significant change overall (**Figure 8**, third row, left and middle).
5. Regarding a possible shift from fuel vehicles to light mobility (non-fuel vehicles), we analyzed two categories in the taxonomy, "Vehicle" and "Wheels," both in Level 1. Recall from **Table 2** that the former aggregates annotations about sounds from motorbikes, cars, and traffic, while the latter tracks sounds associated with bicycles, skateboards, and carts. The difference score (medians across 50 recordings) was significantly different from zero (Wilcoxon signed rank test $V = 490$, $p = 0.0002$***), and there was a significant trend over time ($\rho = 0.31$, $p = 0.03$*), giving evidence for the assumption of an increase of activities involving non-motorized "wheels" such as bicycles vs. motorized mobility. See **Table 3** and **Figure 8** (third row, right).

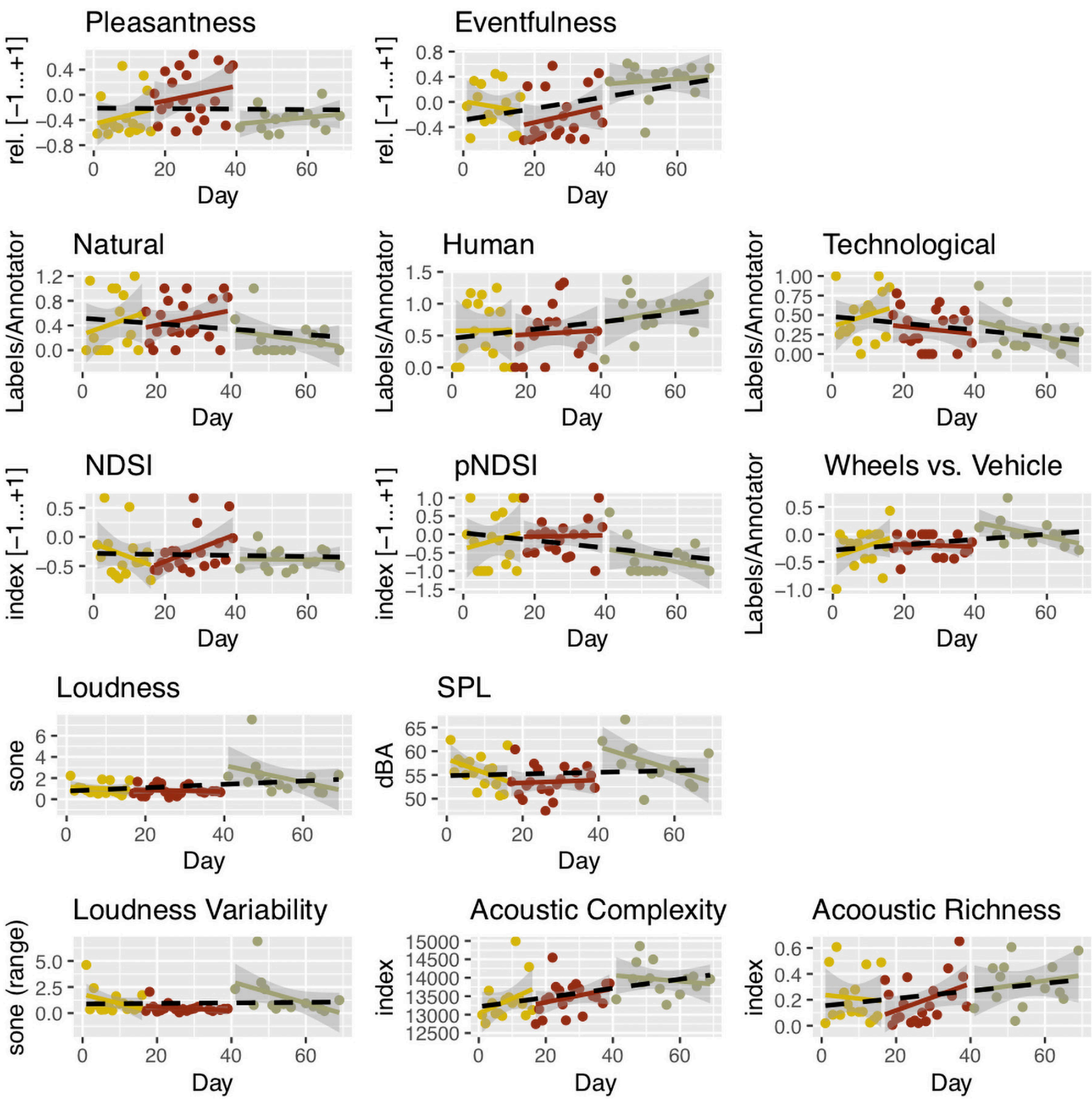


FIGURE 8 | Six computational and seven perceptual descriptors over the 69-day period under study. Loess regressions are given for each of the three lockdown Phases, with smoothing $f = 0.67$ and 95% confidence interval. To facilitate interpretation, a black dotted line indicates the overall linear regression, but note that non-parametric statistics were used in evaluations of descriptor change over time. Image produced with ggplot2 (Liu and Kang, 2016; Wickham et al., 2016).

To sum up, we found that during the 69-day period during the lockdown at the site, overall loudness remained stable. There was a reduction in perceived sounds of machinery, especially traffic, and a shift from fuel vehicles to light mobility. Contrary to our expectations, human outdoor activity increased. There was no appreciable change to the amount of birdsong.

Integrated Analysis

Finally, we present an integrated analysis of qualitative and quantitative results, diary notes, and phases of lockdown. It takes the form of the diagram shown in **Figure 10**, aiming to capture the essentials of the multifaceted aspects of our collaborative case study. Diary notes and Level 3 taxonomic categories (Human, Natural, and Technological; described in the *Methods* section) were organized by phases and further compared with the perceptual analysis results and with Loudness. We analyzed the text of each Diary note in order to assign it to one of the three categories, where possible. Associations were made based on human sounds, i.e., perceived to be produced by people. This includes voices, footsteps, and laughing (see *Annotations of Perceived Sounds* section for details). Natural sounds include birds, seagulls, rain, and dogs. Technological sounds include traffic, various objects, and noises. **Figure 10** places the diary notes on a timeline, by phases. Illustrations visually represent the keywords associated with the corresponding categories in the diary notes, telling the story of each category's change over

TABLE 5 | Correspondences between diary notes and annotations on selected days.

Day	Diary Note	Level 0	Level 3	Evaluation
4	I never noticed how much human voice resonated in the little plaza in front of our window. Interesting how the still image always look the same, day after day, while soundscape is so varied.	Birds (6), close loud hitting (3), whistling, bird chirps, people talking, voices, conversation between man and woman	Human (28), natural (15), technological (12)	Extremely annoying
14	Near silence.	Traffic passing (6), car driving by (3), bird (2), voices faint	Technological (28), natural (16), human (9)	Extremely calm
16	Making noise is feeling alive.	Birds (5), trolley, cart (4), stroller wheels (3), beep (2), dog bark, hits and bumps, music	Technological (32), natural (13), human (7)	Clearly annoying
19	In the silence, someone's getting ready for lunch.	Child (3), glass bottles, object on a surface, car (2), child voice, clacking, hit plate	Human (23), technological (20),	Extremely uneventful
28	Birds, birds, birds.	Child (5), child voice (4),bird (3), birds, birds chirping	Human (39), natural (19), technological (7)	Extremely pleasant
39	Starting Sunday children will be allowed out. Is it excitement I am hearing in the air?	Birds (6), children talking, dog, human voices distant	Human (27), natural (26), technological (14)	Strongly chaotic
48	It's Labor Day, little by little, people are taking back the streets.	Birds (8), bird (6), female, bird chirping (5), footsteps, child (4), talking, children shouting (3), children voices	Human (87), natural (23), technological (4)	Strongly eventful
60	After 2 months of (almost) daily recording the soundscape of the little plaza in front of my window, as the Basque Country rolls out what in Spain is called "Phase 1," with small retailers, hairdressers, hospitality open and people happily and maybe unwisely taking the road, I decided to stop publishing - even though I'll keep recording as we move toward "the new normality". Good luck everybody, who knows what a brave new world is awaiting us out there!	Bird (8), child (6), bird chirping (5), children, children shouting, footsteps (4), kids scooter (3), baby (2)	Human (58), technological (21), natural (20)	Clearly chaotic

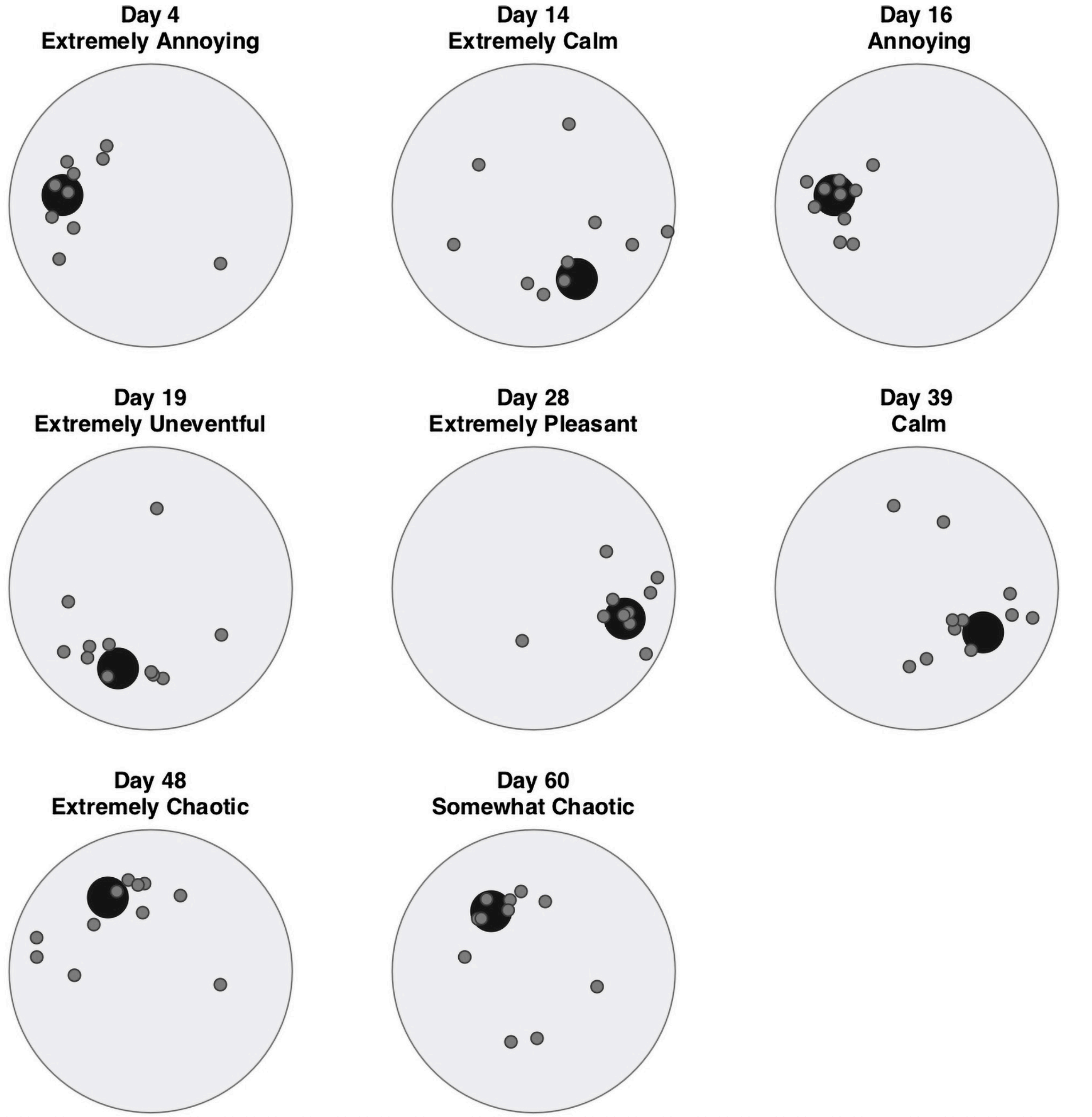


FIGURE 9 | Pleasantness–Eventfulness circumplex for select days that are discussed in the article. See **Supplementary Material** for plots of all the days in the study. Small gray circles are the mean evaluations for each expert evaluator (*N* = 11) calculated as the circular mean of the continuous response and its mean distance from the center. The large black circle is the overall mean. The words in each subtitle indicate in which adjective sector the overall mean is located and its strength.

time. In particular, the category "human" is characterized in the diary notes with keywords associated with the outdoor presence of people at the beginning of lockdown (Phase 1), while during the more restrictive Phase 2, human sounds are described as coming from indoor through open windows. In Phase 1 and Phase 2 notes, such keywords appear as little as five times and one time, respectively. On the other hand, during the most restrictive Phase 2, Natural sounds (birds, but also meteorological elements such as wind, rain, and thunderstorms) appear 11 times, emerging as the most prominent taxonomic category. As shown in **Figure 10**, in Phase 2, we can also observe a temporary decrease in Loudness, which, as illustrated in *Results* section, might be interpreted as related to the concurrent decrease in human activity. Perceptual indicators for Eventfulness also sharply decrease during Phase 2, while indicators of Pleasantness do significantly increase. Likewise,

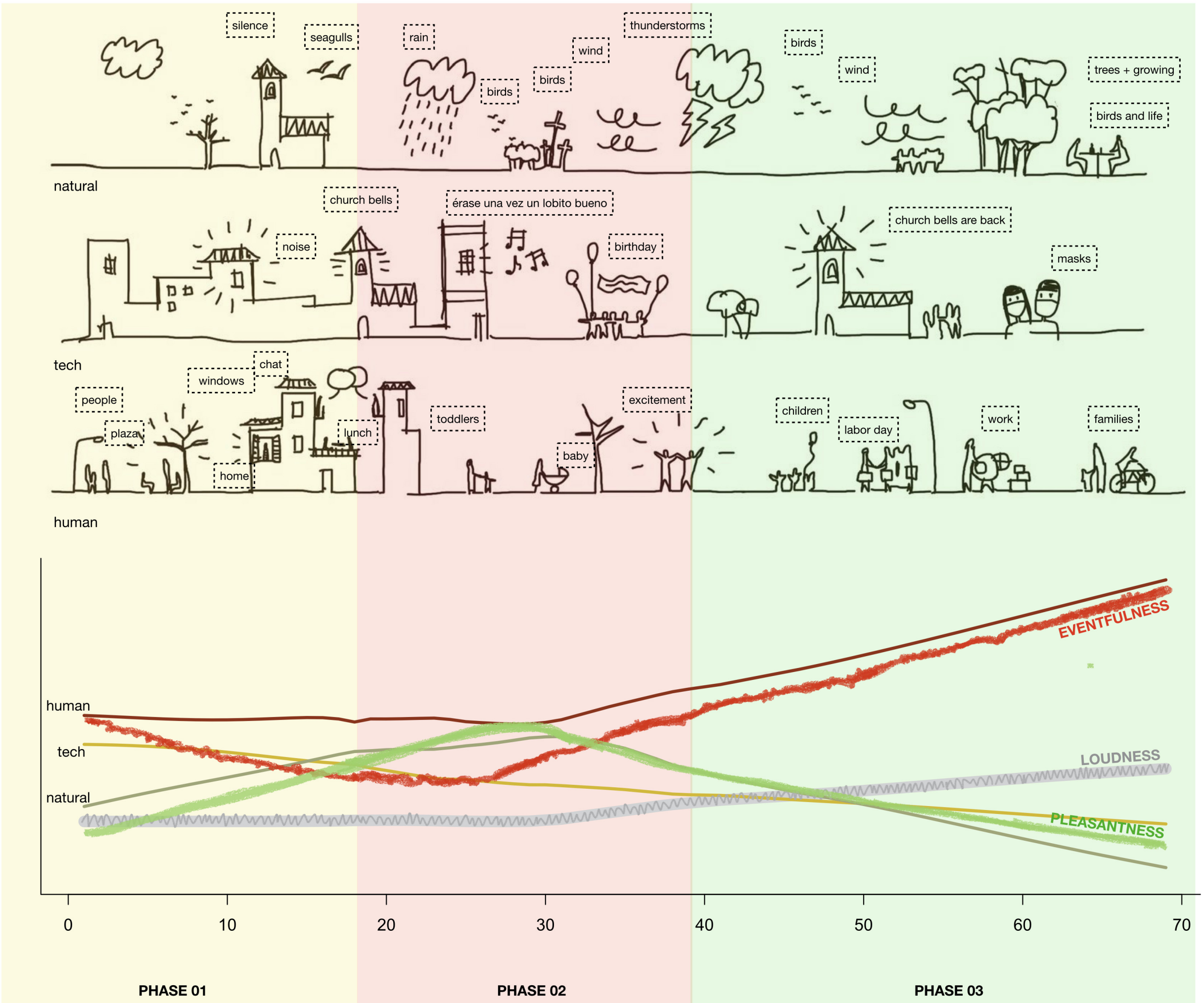


FIGURE 10 | Timeline integrating select descriptors, condensed diary notes, and lockdown phases.

Natural sound sources, mainly related to birds, seem to be more dense in the phases leading up to Phase 3. With the progression of the "Plan de Desescalada," Human sources are described as coming more and more from the street as their frequency increases, with human-related keywords appearing 14 times in the diary notes of Phase 3. A slight, temporary increase in technological sounds during the initial phase of confinement seems to be reflected in the diary notes, with related keywords appearing as much as four times compared with only two mentions of natural sounds. Finally, changes in the diary notes over the whole lockdown period seem to be reflected in the results of qualitative analysis. The Loess curve associated with Human sounds shows a slight decrease throughout Phase 1 and the following Phase 2, while it increases steadily in Phase 3. The increase in Human sounds is mirrored by a sharp increase in perceptual indicators of Eventfulness in Phase 3, while Pleasantness clearly decreases. Overall Loudness increases during the entire period.

DISCUSSION

The increase in Human-generated sounds appears to be the reason for a perceived increase in Eventfulness within the soundscape of the area and a higher degree of Loudness in Phase 3. In the initial Phase 1, leaving one's house was allowed only for work-related tasks and essential shopping. This represented a great change in the habits of local residents. As described in *Site of the Case Study* section, locals gather around the plaza in the hours before lunch (between 12:00 and 14:00, the time of the recording). Adults occupy local cafes for drinks, while kids play at the playground (during non-school days). During Phase 2, these habits were forcibly suspended, a fact that might explain the generally higher level of Pleasantness, which increased in this phase, and the generally lower level of Eventfulness, which also increased. Diary notes of Phase 1 capture the presence of human voices as an indicator of "too many people" being "still around," "enjoying a chat," and "resonating in the little plaza." Human voices, perhaps amplified by the architectural features of the location (the arcade, see *Site of the Case Study* section), seem to be the most relevant image in the soundscape. In Phase 2, with additional restrictions to activity, humans seem to finally leave the scene in favor of birds and other animals (mainly dogs), which become prominent in the Diary Notes. Diary notes ("Day 7. Of voices populating windows" and "Day 8. Stay inside but keep your window wide open, especially if it's Sunday") remind us that neighbors tended to leave their windows open more often. Voices and human sounds were recorded as emanating from the inside of other apartments, revealing to the author the intimacy of family life ("Day 19. In the silence, someone's getting ready for lunch"). It is worth noticing that during Phases 1 and 2 (mid-March to mid-April), temperatures in the area were warmer than the average (Agencia Estatal de Meteorología [AEMET], 2020). This might have contributed to residents moving part of their daily activities out onto their balconies and keeping windows open more often, which caused activities to be heard. As illustrated in *Site of the Case Study* section, the habit of sharing indoor life with the world outside is not common in this area of Spain. Diary notes confirm how exceptional such a behavior was: "Day 22. A good thing of these days is that I am finally seeing my neighbors, from window to window. Nobody used to lean out of the window in this *barrio* [neighborhood] of mine, before" and "Day 65. Last night we had the last collective clapping for healthcare workers. What will happen to my neighbors now?" An increased sense of community during and after lockdown has been reported by several sources. One study found that "a substantial proportion of people felt that they had become more involved in neighborhood life following the lockdown" (Jones et al., 2020) and that support among neighbors include "raising morale through humor, creativity and acts of kindness and solidarity" (idem). Another observer wrote that "It's not just the help and practicalities, the socializing too is vital. Conversations across balconies, news being discussed and sometimes neighbors humming along to music being played next door" (Banerjee, 2020). This apparently new attitude might be connected to a psychological reaction to isolation during lockdown (Henley, 2020) and the need to share with neighbors during such a unique time. The limitations of the present study do not allow for a conclusive interpretation of results. It is difficult to say whether the local residents of the road kept their windows open due to the exceptional meteorological conditions or as the result of psychological reaction to isolation. Note that both Diary Notes and Annotations by the expert group indicate the influence of the indoor soundscape on the outdoor soundscape, which might be due to a reduction of other common sources of outdoor sounds.

In fact, machinery and human interaction sounds decreased during the lockdown. This result needs to be considered within the context of the specific location under study, a pedestrian road where traffic noise even during normal times is within regulatory levels. Restrictions such as Phase 2 total ban of any non-necessary activity and a stop on constructions and home renovations should be taken into account. It is only in the last phase of the "Plan de Desescalada," in mid-June, that this ban was lifted. It is well known by now (Garcia, 2020) that such restrictions on human activity had a strong impact on the local mobility of people and vehicles. In the whole Basque Country, public transport was reduced by 50% during Phase 1 and only recovered full capacity with the entrance in the so-called "new normality" in late June. Conversely, traffic of light vehicles in the Basque Country decreased by 95% (Dirección General de Tráfico, 2020) during Phases 1 and 2, along with traffic of heavy vehicles, which is a well-known source of noise pollution (Jacyna et al., 2017; Kulauzović et al., 2020), decreased by more than 50% (Ortega Dolz, 2020). With the ban on mobility and closure of international borders, the local airport was exceptionally quiet, and the reduction in international air traffic was 95% (Alonso, 2020). By contrast, maritime activity at the port of Bilbao fell by only 5% (Alvarez, 2020). The disparity in reduction between these two types of trade and travel (air or sea) reflects not only their influence on the economy but also their very different levels of impact on the acoustic environment. In fact, if the maritime traffic does not affect the acoustic environment of Las Arenas, traffic from the airport can be heard at times over the area, mostly when, due to specific meteorological conditions, aircrafts land from the sea, thus flying over Calle Paulino Mendibil. Interestingly, the acoustic presence of aircrafts is also reflected in the diary notes, on Day 6: "Day 6 of lockdown from my window in Getxo, Basque Country. It sounds someone [*sic*] is still flying out from here. Or in, who knows."

Looking at our results, the perceptual analysis seems consistent with an urban soundscape where acoustic events are more rarefied during Phases 1 and 2 of the lockdown, while Eventfulness grows in Phase 3, when mobility for leisure is allowed and public as well as private transport is progressively restored. As for human interaction, two factors are worth noticing: the influence of the reopening of cafes and restaurants, with outdoor spaces on the public pedestrian road being allowed extra hours in order to make up for the economic loss of the lockdown weeks; and the reopening of the children playground in the square located opposite the observation point, in a moment (Phase 3) where schools were closed, due to the end of the academic year, which coincided with the end of the "Desescalada." Toward the end of Phase 3, the increase in Eventfulness is testimony to the progressive return of local habits and behaviors, described in *Site of the Case Study* section.

As for the category of Natural sounds, undoubtedly, the sonic imagery of birds singing and birdsong has grown in importance during lockdown. Media have widely reported an increase in attention toward the singing of birds by both the scientific and artistic communities. From the launch of the first international global soundscape of spring dawn chorus created by artists and scientists (Morss, 2020), to birdsong becoming "more beautiful" (Cockburn, 2020) or "sexier" (Chrobak, 2020) thanks to the absence of human activity, birds seem to have grown to represent the essence of the urban soundscape in lockdown to the point that they could condition political choices (The Economist, 2020). The results of the present studies seem to support these claims at least for Phase 2, when the most restrictive measures were applied to human mobility. In this phase, an increase in the presence of natural sound sources (mainly birds) is observable in the annotations as well as in diary notes, where the word "birds" is often noted down in isolation, as if it could contain by itself the whole imagery of a pleasant soundscape. In our study, Pleasantness increased during Phase 2, while Eventfulness decreased, allowing for relating the sonic image of birdsong with that of pleasantness and calm.

A separate reflection should be dedicated to the consequences of lockdown regulations on light mobility. Findings seem to indicate that a growth in perceived sounds of Wheels, a.k.a. manual (non-electric) PMVs, contribute to a perceived increase in Eventfulness, an increase in Loudness, and decrease of Pleasantness of the soundscape in Phase 3. Rather unexpectedly, scooters and other light vehicles seemed to produce a considerable perceptual annoyance, at least when used on pedestrian non-PMV-specific surfaces and nearby residential dwellings, such as the observation point of this study. During the same time period as the present study, according to a study by the Spanish insurance company Acierto and widely circulated by the media (Gutierrez, 2020), the use of bicycles grew by as much as seven times. Purchases have increased by 30% since Phase 1 (Blanchar, 2020). Usage of manual and electric scooters as well as other light mobility has also grown. Additionally, from Phase 3 onward, children were allowed out and expressly permitted, if not encouraged (Lucas, 2020), to use PMVs such as skateboards, roller skaters, and manual scooters. At the time, extended media coverage was dedicated to the claim that "the pandemic and this crisis is causing a rethinking of many issues in life. It will give much more voice to people who do not use but who suffer from the presence of the car. This element, essential in our culture, will no longer be the privileged element of the city. The post-COVID city will be the post-car city" (Spanish urban planner Jose Ezquiaga interviewed by Mendoza Pérez, 2020). In our analysis, we highlighted how evidence could be found that in Phase 3 there was an increase of perceived sounds of Wheels (light vehicles) vs. perceived sounds from Vehicles (traditional fuel vehicles, cf. **Figure 8**, third row, right). It is difficult to say whether these changes will be permanent, or whether a return toward private (mainly fuel-based) transport is to be expected, given the risks associated with traveling on public transport while the pandemic is still unresolved. In the months following lockdown, the motorcycle sector in Spain "witnessed a growth above double-digit, still among great economic uncertainty" (Asociación Nacional de Empresas del Sector de Dos Ruedas [ANESDOR], 2020). We can posit though that the environmental presence of light mobility is important and that specific lane/pavements might be considered to be included as urban design criteria. It is indeed expected that in future cities, the traditional prominence of fuel cars will be superseded by a range of other kinds of mobility devices. "The world of vehicles is exploding in thousands of shapes and sizes. We are witnessing our cities being more and more conquered by vehicles of different kinds, dimensions and number of occupants" (Sádaba, 2019). Fumihiko Maki talks about the intangible "linkage" as the glue, "the act by which we unite the different layers of activity and resulting form of the city" (Maki, 1964). Changes in mobility, means of communication, and social interaction will probably soon redefine the eventual shape of cities, a shift that the COVID-19 pandemic seems to have accelerated. On the one hand, the increase in light mobility will have an impact on the width and layout of traffic lanes and on the design and occupation and management of sidewalk curbs (Goffman, 2018). On the other hand, smart working and a consequent decrease in face-to-face meetings might reduce the need for human displacement within cities. Koolhaas' "event structure" of cities (Kipnis, 1996) will become diverse and multifaceted. The limitations of the present case study do not allow us to generalize results in order to imagine potential scenarios of changes in the urban soundscape in the case of a decrease in fuel cars. Nonetheless, we believe that by measuring and analyzing *intangible* aspects of the city, such as the acoustic environment, we can gather precious insights to optimize the design of appropriate indicators for future quality of life, as well as health, and a better management of finite resources. In our study, we combined what Gehl and Svarre call "Keeping a Diary," "Photographing," and "Tracking" (our soundscape recordings) with quantitative and qualitative analyses, to understand the social behavioral changes triggered by the lockdown. Soundscape research provides crucial knowledge, allowing a better understanding of city life. This study contributes to opening up further research on the tangible–intangible duality in order to offer improved urban indicators for city and mobility design.

CONCLUSION

When the COVID-19 crisis subsides, we might be able to compare and synthesize results from these varied endeavors and many more. It is yet too early to speculate about what might be learned from the experiences we – all of us – are currently making in exceptional times. Soundscape research provides vital clues to understanding the perception and design of the multimodal environment – how humans are psychologically, physiologically, physically, and socially affected by sound; and also, how other living creatures are likewise affected. It makes a constructive and oftentimes undervalued contribution. Will researchers, sound designers, and architects be part of a discussion with urbanists, policy makers, politicians, businesses, and indeed the general public, in seeking solutions to the mounting challenges to urban living conditions in the future?

We believe that the tangible–intangible duality can be applied as a holistic approach to include both objective and subjective indicators to the evaluation of urban soundscape. At the start of this article, we briefly recalled the origins of the soundscape movement and its influence on urban ecology and urban planning. Still, public endeavors such as the Noise White Paper of Getxo exemplify the fact that "landscape architecture and related disciplines have not fully recognized the possibilities of considering sound issues in design projects" (Cerwén, 2017). The Noise White Paper of Getxo is written in development of the Acoustic Pollution Decree of the Basque Government (Decreto 213/2012, de 16 de octubre, de contaminación acústica de la Comunidad Autónoma del País Vasco) that, undeniably, only focuses on the negative aspects of sound and/or noise in order to detect problematic points in the urban landscape and mitigate the effects and the subsequent discomfort they may cause. Furthermore, the project for the Law for the Protection of Landscape of the Basque Government (1998) does not make any specific mention of soundscape as one of the elements of urban planning. This is all the more remarkable given that for more than two decades the European Commission has had policies in place aimed at reducing noise exposure. Indeed, noise is still "the ignored pollutant" (King and Murphy, 2016), but one way forward is to focus less on noise in general and more on how to promote specific and positively valenced sounds in the environment (e.g., Davies et al., 2013; Aletta et al., 2016a). Our study aims to contribute to an understanding of the relationship between sound sources and holistic soundscape evaluation. Through this study, we employed a mixed methodology that aims to measure both tangible (such as loudness) and intangible (such as perception of quality) aspects of the environment of a specific neighborhood of the city of Getxo. We believe that it would be possible – and advisable – to introduce this approach to the analysis of the urban soundscape with the ultimate goal to include the attention to the audible landscape of the city in the procedures established by the law in terms of landscape protection and ultimately in the local urban planning policies.

Specifically, we seek to continue the development of such a mixed methodology in terms of both qualitative and quantitative research. On the one hand, we will further develop the collection of field notes and direct observations of the soundscape as a complement to the collection of field recordings. In circumstances other than those allowed by the restrictions imposed to mobility during lockdown (that affected the development of this study), we recommend that such field notes and observations are complemented by interviews to residents and other participatory activities.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The collection of diary notes and soundscape recordings did not require an ethics approval from the institution of the first author. The procedures for data colloction for soundscape evaluation and annotation were approved by the Research Ethics Committee of City University of Hong Kong (ref. 13-2020-08-E). The expert group members provided their written informed consent to participate.

AUTHOR CONTRIBUTIONS

SL and JS conceived the study. SL collected audio, photos, and diary notes. PL conducted annotations and evaluations by the expert group and analyzed the data. SL, JS, and PL wrote the manuscript. JS made illustrations and drawings. All authors contributed to the article and approved the submitted version.

FUNDING

This research was conducted with funding from Politcnico di Milano - Design (Lenzi), EKOPOL (Sábada), and CityU StUp Grant 7200671 (Lindborg).

ACKNOWLEDGMENTS

The authors thank each and all of the experts who volunteered their time and effort to annotate and evaluate the soundscape recordings.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.570741/full#supplementary-material

Supplementary Data Sheet 1 | Audio Selection.

Supplementary Data Sheet 2 | Diary notes.

Supplementary Data Sheet 3 | Phases of the lockdown regulations.

Supplementary Data Sheet 4 | Instructions to the expert group for soundscape annotations.

Supplementary Data Sheet 5 | Pleasantness-Eventfulness plots for all the days in the study.

Supplementary Data Sheet 6 | Instructions to the expert group for soundscape evaluations.

Supplementary Data Sheet 7 | Annotations and levels of the sound source taxonomy (in pdf format).

Supplementary Data Sheet 8 | Annotations and levels of the sound source taxonomy (in csv format).

REFERENCES

AENA (2013). *Mapas Estratégicos De Los Grandes Aeropuertos. Aeropuerto De Bilbao*. Madrid: AENA.

Agencia Estatal de Meteorología [AEMET] (2020). *Análisis Estacional: Bilbao Aeropuerto*. Available online at: http://www.aemet.es/es/serviciosclimaticos/vigilancia_clima/analisis_estacional?w=2&l=1082&datos=temp (accessed February 23, 2021).

Aletta, F., Kiang, J., and Östen, A. (2016a). Soundscape descriptors and a conceptual framework for developing predictive soundscape models. *Landsc. Urban Plann.* 149, 65–74. doi: 10.1016/j.landurbplan.2016.02.001

Aletta, F., Lepore, F., Kostara-Konstantinou, E., Kang, J., and Astolfi, A. (2016b). An experimental study on the influence of soundscapes on people's behaviour in an open public space. *Appl. Sci.* 6:276. doi: 10.3390/app6100276

Alonso, A. G. (2020). *La Paloma Gestionó Tan Solo 1.415 Pasajeros En Abril*. Available online at: https://www.deia.eus/bizkaia/2020/05/13/paloma-gestiono-1415-pasajeros-abril/1038267.html (accessed February 23, 2021).

Alvarez, M. (2020). *El Covid-19 Tiene Un Impacto Moderado En El Tráfico Del Puerto De Bilbao*. Available online at: https://www.elcorreo.com/economia/covid19-impacto-moderado-20200525133838-nt.html (accessed February 23, 2021).

Anikin, A. (2017). *Package 'Soundgen'*.

Arana, M. (2010). Are urban noise pollution levels decreasing? *J. Acoust. Soc. Am.* 127, 2107–2109. doi: 10.1121/1.3337228

Asociación Nacional de Empresas del Sector de Dos Ruedas [ANESDOR] (2020)). *Las Matriculaciones de Vehículos Ligeros Crecieron un 3,2% en Septiembre, Press Release*. Available online at: https://www.anesdor.com/category/notas-de-prensa/ (accessed February 23, 2021).

Atkinson, P., Bauer, M. W., and Gaskell, G. (2000). *Qualitative Researching with Text, Image and Sound: A Practical Handbook for Social Research*. Thousand Oaks, CA: SAGE Publications.

Axelsson, Ö, Nilsson, M. E., and Berglund, B. (2010). A principal components model of soundscape perception. *J. Acoust. Soc. Am.* 128, 2836–2846. doi: 10.1121/1.3493436

Banerjee, P. (2020). *Life in the Time of Covid-19: When Distancing Draws You Closer*. Available online at: https://www.hindustantimes.com/lifestyle/life-in-the-time-of-covid-19-when-distancing-draws-you-closer/story-k6kqSf3CaQb45yP0i6oEPJ.html (accessed February 23, 2021).

Berglund, B., and Nilsson, M. E. (2007). "Summary of the studies in soundscape perception," in *Soundscape Support to Health*, ed. K. Spång (Stockholm: Swedish Foundation for Strategic Environmental Research (Mistra)), 14–23.

Blanchar, C. (2020). *Las Bicis Bajan Del Trastero Y De Los Balcones*. Available online at: https://elpais.com/espana/catalunya/2020-06-02/las-bicis-bajan-del-trastero-y-de-los-balcones.html (accessed February 23, 2021).

Bui, Q., and Badger, E. (2020). *The Coronavirus Quieted City Noise. Listen to What's Left*. Available online at: https://www.nytimes.com/interactive/2020/05/22/upshot/coronavirus-quiet-city-noise.html (accessed February 23, 2021).

Cain, R., Jennings, P., Adams, M., Bruce, N., Carlyle, A., Cusack, P., et al. (2008). Sound-scape: a framework for characterising positive urban soundscapes. *J. Acoust. Soc. Am.* 123, 3394–3394. doi: 10.1121/1.2934071

Cerwén, G. (2017). *Sound in Landscape Architecture. A Soundscape Approach to Noise*. Ph.D. thesis, Swedish University of Agricultural Sciences.

Chion, M., and Gorbman, C. (2009). *Film, a Sound Art*. New York, NY: Columbia University Press.

Chrobak, U. (2020). *Bird Songs Got Sexier During the COVID-19 Shutdown*. Available online at: https://www.popsci.com/story/environment/more-bird-song-pandemic/ (accessed February 23, 2021).

Cockburn, H. (2020). *Birdsong Became More Beautiful During Lockdown Because Humans Weren't Drowning It Out, Scientists Reveal*. Available online at: https://www.independent.co.uk/independentpremium/birdsong-lockdown-impact-animals-coronavirus-nature-b599282.html (accessed February 23, 2021).

Consejos de Ministros de España (2020). *Plan De Desescalada*. La Moncloa: Consejos de Ministros de España.

Davies, W. J., Adams, M. W., Bruce, N. S., Cain, R., Carlyle, A., Cusack, P., et al. (2013). Perception of soundscapes: an interdisciplinary approach. *Appl. Acoust.* 74, 224–231. doi: 10.1016/j.apacoust.2012.05.010

Dirección General de Tráfico (2020). Available online at: http://www.dgt.es/es/covid-19 (accessed February 23, 2021).

EUSTAT (2020). *Instituto Vasco de. Getxo. Realidad Estadística Actual Y Evolución Histórica*. Available online at: https://www.eustat.eus/municipal/datos_estadisticos/getxo_c.html (accessed February 23, 2021).

Fernandez, M. (2020). *Los Retos De Las Ciudades En El Escenario Post-Coronavirus*. Available online at: https://elpais.com/elpais/2020/04/29/seres_urbanos/1588155518_790132.html (accessed February 23, 2021).

Flick, U. (2014). *An Introduction to Qualitative Research*, 5th Edn. Thousand Oaks, CA: Sage.

Garcia, J. (2020). *El Gobierno Vasco Reduce Entre Un 40 Y Un 60% La Oferta De Transporte Público*. Available online at: https://www.elcorreo.com/sociedad/salud/gobierno-vasco-reduce-20200314183512-nt.html (accessed February 23, 2021).

Gehl, J., and Svarre, B. (2013). *How to Study Public Life*. Washington: Island Press.

Basque Government (1998). *Ley 3/1998, de 27 de febrero, General de Protección del Medio Ambiente del País Vasco*. Alonsotegi: Basque Government.

Goffman, E. (2018). *How to Manage the Chaotic 21st Century Curb*. Available online at: https://mobilitylab.org/2018/06/08/managing-the-chaos-of-the-21st-century-curb/ (accessed February 23, 2021).

Gutierrez, A. (2020). *La Bicicleta, La Gran Favorita*. Available online at: http://revista.dgt.es/es/noticias/nacional/2020/06JUNIO/0603dia-mundial-de-la-bicicleta.shtml# (accessed February 23, 2021).

Henley, J. (2020). *Lockdown Living: How Europeans Are Avoiding Going Stir Crazy*. Available online at: https://www.theguardian.com/world/2020/mar/28/lockdown-living-europe-activities-coronavirus-isolation (accessed February 23, 2021).

Hong, J. Y., and Jin, Y. J. (2015). Influence of urban contexts on soundscape perceptions: a structural equation modeling approach. *Landsc. Urban Plann.* 141, 78–87. doi: 10.1016/j.landurbplan.2015.05.004

ISO (2014). *Acoustics—Soundscape—Part 1: Definition and Conceptual Framework*. Geneva: International Organization for Standardization.

Jacobs, J. (1961). *The Death and Life of Great American Cities*. New York: Random House.

Jacyna, M., Wasiak, M., Lewczuk, K., and Karoń, G. (2017). Noise and environmental pollution from transport: decisive problems in developing ecologically efficient transport systems. *J. Vibroeng.* 19, 5639–5655. doi: 10.21595/jve.2017.19371

Jefatura del Estado. (2003). Ley 37/2003, de 17 de noviembre, del Ruido. ≪BOE≫ núm. 276, de 18 de noviembre de 2003, 40494–40505.

Jones, M., Beardmore, A., Biddle, M., Gibson, A., and Umar Ismail, S. (2020). Apart but not Alone? A cross-sectional study of neighbour support in a major UK urban area during the COVID-19 lockdown. *Emerald Open Res.* 2:37. doi: 10.35241/emeraldopenres.13731.1

Kang, J., Aletta, F., Oberman, T., Erfanian, M., Kachlicka, M., Lionello, M., et al. (2019). "Towards soundscape indices," in *Proceedings of the International Congress on Acoustics—ICA*, Aachen.

Kassam, A., and Burgen, S. (2020). *Spain's Streets Turn Eerily Quiet Following Coronavirus Lockdown*. Available online at: https://www.theguardian.com/world/2020/mar/15/spain-city-streets-turn-quiet-coronavirus-lockdown-madrid-barcelona (accessed February 23, 2021).

Kasten, E. P., Gage, S. H., Fox, J., and Joo, W. (2012). The remote environmental assessment laboratory's acoustic library: an archive for studying soundscape ecology. *Ecol. Inform.* 12, 50–67. doi: 10.1016/j.ecoinf.2012.08.001

King, E. A., and Murphy, E. (2016). Environmental noise–'Forgotten'or 'Ignored'pollutant? *Appl. Acoust.* 112, 211–215. doi: 10.1016/j.apacoust.2016.05.023

Kipnis, J. (1996). *El Ultimo Koolhas . El Croquis 79, no*. Rotterdam: OMA/Rem Koolhas.

Krause, B. (2008). Anatomy of the soundscape: evolving perspectives. *J. Audio Eng. Soc.* 56, 73–80.

Kulauzović, B., Pejanović Nosaka, T., and Jamnik, J. (2020). "Relationship between weight of the heavy trucks and traffic noise pollution in the viewpoint of feasibility of fines for exceeded noise–a case study," in *Proceedings of 8th Transport Research Arena TRA*, Helsinki.

Lenzi, S. (2020). *Data Sonification Blog*. Available online at: https://www.saralenzi.com/datasonificationblog (accessed January 16, 2021).

Lindborg, P. M. (2015). Psychoacoustic, physical, and perceptual features of restaurants: a field survey in singapore. *Appl. Acoust.* 92, 47–60. doi: 10.1016/j.apacoust.2015.01.002

Lindborg, P. M. (2016). Taxonomy of sound sources in restaurants. *Appl. Acoust.* 110, 297–310. doi: 10.1016/j.apacoust.2016.03.032

Lindborg, P. M. (2019). How do we listen? *Emille Journal of the Korean Electro-Acoustic Society* 16, 43–49.

Liu, F., and Kang, J. A. (2016). Grounded theory approach to the subjective understanding of urban soundscape in sheffield. *Cities* 50, 28–39. doi: 10.1016/j.cities.2015.08.002

Lucas, B. (2020). *Los Niños Podrán Salir Una Hora Al Día, a Un Kilómetro Y Llevar Patinete O Bicicleta*. Available online at: https://elpais.com/sociedad/2020-04-23/los-ninos-podran-pasear-a-un-kilometro-de-su-casa-con-un-adultocon-tres-ninos-y-podran-llevar-su-patinete.html (accessed February 23, 2021).

Maharaj, N. (2016). Using field notes to facilitate critical reflection. *Reflect. Pract.* 17, 114–124. doi: 10.1080/14623943.2015.1134472

Maki, F. (1964). *Investigations in Collective Form. The City as a Pattern of Events*. Saint Louis: Washington University.

Mallo, J. (2020a). *Euskadi Entra En La Fase 3 Y Recuperará El Lunes La Movilidad Entre Sus Provincias*. Available online at: https://www.elcorreo.com/sociedad/salud/euskadi-pasa-fase-20200605112710-nt.html (accessed February 23, 2021).

Mallo, J. (2020b). *Euskadi Ya Está En La Fase 2: Todo Lo Que Se Puede Hacer Desde Este Lunes*. Available online at: https://www.elcorreo.com/sociedad/salud/fase-acerca-podre-20200517084256-nt.html (accessed February 23, 2021).

McVeigh, K. (2020). *Silence Is Golden for Whales as Lockdown Reduces Ocean Noise*. Available online at: https://www.theguardian.com/environment/2020/apr/27/silence-is-golden-for-whales-as-lockdown-reduces-ocean-noise-coronavirus (accessed February 23, 2021).

Mendoza Pérez, S. (2020). *José María Ezquiaga: La Ciudad Pos-Covid Será La Ciudad Posautomóvil*. Available online at: https://www.eldiario.es/madrid/Entrevista-Ezquiaga_0_1024348147.html (accessed February 23, 2021).

Morss, A. (2020). *Lockdown Yields First Global Sound Map of Spring Dawn Chorus*. Available online at: https://www.theguardian.com/environment/2020/may/30/lockdown-yields-first-global-sound-map-spring-dawn-chorus-birds (accessed February 23, 2021).

Municipal de Getxo (2016). *Municipal Act 4601/2016*. Available online at: https://www.getxo.eus/es/medio-ambiente/atmosfera-y-ruido (accessed February 23, 2021).

Ortega Dolz, P. (2020). *El Cierre De Fronteras Se Amplía También a Puertos Y Aeropuertos*. Available online at: https://elpais.com/espana/2020-03-22/el-cierre-de-fronteras-se-amplia-tambien-a-puertos-y-aeropuertos.html (accessed February 23, 2021).

Pieretti, N., Farina, A., and Morri, D. (2011). A new methodology to infer the singing activity of an avian community: the acoustic complexity index (ACI). *Ecol. Indic.* 11, 868–873. doi: 10.1016/j.ecolind.2010.11.005

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Remote Environmental Assessment Laboratory [REAL] (2020). *Ecosystem Assessment*. Available online at: https://www.remoteenvironmentalassessmentlaboratory.com/overview/ecoassessment.php (accessed February 23, 2021).

Ro, C. (2020). *Is Coronavirus Reducing Noise Pollution?*. Available online at: https://www.forbes.com/sites/christinero/2020/04/19/is-coronavirus-reducing-noise-pollution/#6f2de8f7766f (accessed February 23, 2021).

Sádaba, J. (2019). "Expanded/exploded mobility. matching alternative mobility research with spatial and city planning needs," in *Paper presented at the IX CIOT 2019*, (Santander: Congréso Internacional de Ordenación del Territorio).

Sandford, A. (2020). *Coronavirus: Half of Humanity Now on Lockdown as 90 Countries Call for Confinement*. Available online at: https://www.euronews.com/2020/04/02/coronavirus-in-europe-spain-s-death-toll-hits-10-000-after-record-950-new-deaths-in-24-hou (accessed February 23, 2021).

Schafer, R. M. (1977). *The Soundscape: Our Sonic Environment and the Tuning of the World*, 1994 Edn. Merrimac, MA: Destiny books.

Scott-Ram, N. R. (1990). *Transformed Cladistics, Taxonomy and Evolution*. Cambridge, MA: Cambridge University Press.

Sueur, J. (2018). *Sound Analysis and Synthesis with R*. Cham: Springer.

Sueur, J., Aubin, T., Simonis, C., Lellouch, L., Brown, E. C., Depraetere, M., et al. (2019). *Package 'Seewave' Version 2.1.6*.

The Economist (2020). *After Hearing Birdsong During Lockdown, French Cities Vote Green*. Available online at: https://www.economist.com/europe/2020/07/02/after-hearing-birdsong-during-lockdown-french-cities-vote-green (accessed February 23, 2021).

Truax, B. (2019). "Acoustic Ecology and the World Soundscape Project," in *Sound, Media, Ecology*, eds M. Droumeva and R. Jordan (Cham: Palgrave Macmillan), 21–44. doi: 10.1007/978-3-030-16569-7_2

Weston, P. (2020). *'The Bliss of a Quiet Period': Lockdown Is a Unique Chance to Study the Nature of Cities*. Available online at: https://www.theguardian.com/environment/2020/apr/27/the-bliss-of-a-quiet-period-lockdown-is-a-unique-chance-to-study-the-nature-of-cities-aoe (accessed February 23, 2021).

Wickham, H., Chang, W., and Hadley Wickham, M. (2016). *Package 'Ggplot2'. Create Elegant Data Visualisations Using the Grammar of Graphics. Version 2.* 1–189.

Zwicker, E., and Fastl, H. (2013). *Psychoacoustics: Facts and Models*, Vol. 22. Cham: Springer Science & Business Media.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

REVIEW
published: 26 April 2021
doi: 10.3389/fpsyg.2021.570563

Check for updates

Sound and Soundscape in Restorative Natural Environments: A Narrative Literature Review

*Eleanor Ratcliffe**

Faculty of Health and Medical Sciences, School of Psychology, University of Surrey, Guildford, United Kingdom

Acoustic experiences of nature represent a growing area in restorative environments research and are explored in this narrative literature review. First, the work surveyed indicates that nature is broadly characterized by the sounds of birdsong, wind, and water, and these sounds can enhance positive perceptions of natural environments presented through visual means. Second, isolated from other sensory modalities these sounds are often, although not always, positively affectively appraised and perceived as restorative. Third, after stress and/or fatigue nature sounds and soundscapes can lead to subjectively and objectively improved mood and cognitive performance, as well as reductions in arousal, although some inconsistencies in findings are observed. Fourth, theoretical frameworks of restorative environments would benefit from inclusion of acoustic environmental properties such as sound intensity or frequency. Fifth, findings regarding positive, learned semantic associations with nature have arisen as a result of recent work on sounds and restoration. This represents another important area of potential theoretical development for broader restorative environments research.

Keywords: soundscape, nature sounds, restorative environments, attention restoration, stress recovery

OPEN ACCESS

Edited by:
Francesco Aletta, University College London, United Kingdom

Reviewed by:
Neil Spencer Bruce, School of Electronic Music (SEM), United Kingdom
Giuseppe Carrus, Roma Tre University, Italy

****Correspondence:***
Eleanor Ratcliffe
eleanor.ratcliffe@surrey.ac.uk

Specialty section:
This article was submitted to Environmental Psychology, a section of the journal Frontiers in Psychology

Received: *08 June 2020*
Accepted: *04 March 2021*
Published: *26 April 2021*

Citation:
Ratcliffe E (2021) Sound and Soundscape in Restorative Natural Environments: A Narrative Literature Review. Front. Psychol. 12:570563. doi: 10.3389/fpsyg.2021.570563

INTRODUCTION

There is an abundance of literature regarding the ability of certain settings, termed "restorative environments," to facilitate recovery from everyday cognitive fatigue, negative mood, and stress (Collado et al., 2017). Much attention has been paid to the restorative value of natural environments in particular (Hartig et al., 2014). Studies on these topics tend to focus on visuo-spatial experience of environments, utilizing stimuli such as photographs, videos, and slideshows, but environments are not experienced through vision alone. There is growing interest in and call for study of non-visual aspects of restorative environments, including sound, smell, and touch (Conniff and Craig, 2016; Iyendo, 2016; Franco et al., 2017; Aletta and Kang, 2019; Sona et al., 2019; Schebella et al., 2020). Such work is important to ensure that the research field remains relevant to individuals with visual impairment (Shaw et al., 2015; Bell, 2019a,b) and to maximize extended reality presentations of environments, e.g., through virtual or augmented reality (Depledge et al., 2011).

While research on touch and smell in restoration remains limited, nature sounds and natural soundscapes are increasingly identified as important ecosystem services that can aid psychological restoration as well as well-being more broadly (Francis et al., 2017). Here soundscape is defined as the acoustic environment as perceived, understood, and/or experienced by people, in context (see International Organization for Standardization, 2014). However, the theories that seek to explain

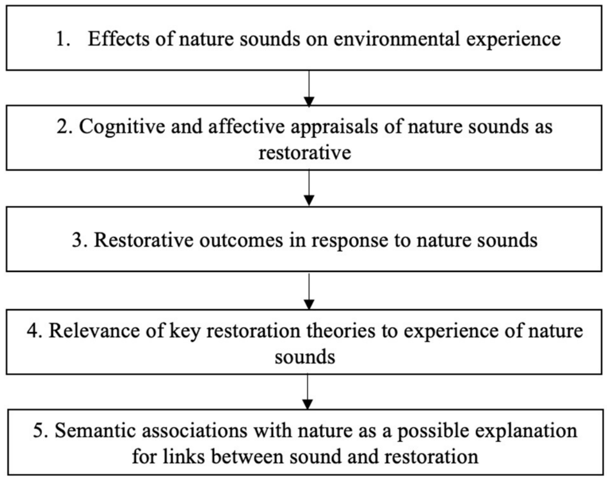


FIGURE 1 | Flowchart summarizing the key topics of interest in this literature review.

why certain environments facilitate restoration focus primarily on visual experience (see Ulrich, 1983; Kaplan and Kaplan, 1989). The first step in better integrating sound and soundscape into our theoretical understanding is to examine and review the available literature.

A systematic review by Aletta et al. (2018) has identified links between positive urban soundscapes (which may also include nature sounds) and health and well-being, including stress recovery. Given the emphasis on nature within restorative environments (see Hartig et al., 2014), the present narrative literature review focuses on evidence for positive psychological experiences of nature sounds and soundscapes specifically, and in particular how listening to these can generate perceptions and outcomes of restoration from stress and fatigue. This review has five key objectives, summarized in **Figure 1**. First, it explores literature regarding the impact of nature sounds on perceptions and experiences of wider natural environments. Second, it examines evidence regarding cognitive and affective appraisals of nature sounds and their contributions to overall perceptions of restorative environments. Third, literature regarding restorative outcomes in response to nature sounds is assessed. Fourth, the relevance of key restoration theories to this topic is examined and areas where these theories are limited are identified. Fifth, a possible new theoretical area of interest—semantic associations with nature—is discussed and exemplified by recent acoustics research.

SOUNDS ARE IMPORTANT FOR ENVIRONMENTAL EXPERIENCES OF NATURE

Peace and quiet are important aspects of being in nature but this does not mean the presence of complete silence—rather, it can relate to the concept of relative tranquility, or reduction in sounds from the built environment and the opportunity to hear pleasant sounds of nature (De Coensel and Botteldooren, 2006; Pheasant et al., 2008).

Qualitative studies describe exposure to natural environments as a positively regarded, multi-sensory experience, whereas a lack of such multi-sensory aspects is regarded negatively. For example, following qualitative interviews with 20 wildlife tourists, Curtin (2009, p. 461) reported that participants experienced a heightened sensory awareness after wilderness trips to locations in Spain and USA: "I have seen and heard things in the natural world that I didn't know even existed. It was as if my senses were coming alive…" Curtin describes the sensory dominance of vision in the wildlife tourism experience but notes that it is experienced in the context of other sensory modalities such as sound and smell. In their qualitative study, Fredrickson and Anderson (1997, p. 31) found that a sample of 12 women reported direct experience of the sounds of nature as a particularly meaningful aspect of wilderness trips to Minnesota and Arizona, USA. As one participant observed, "It was so incredible being able to hear the birds, yeah, and just the crunching of animals all around us… The sounds of the forest, the snapping of the twigs, hearing the tiny sigh of the wind through the treetops at night."

In a study of participants awaiting treatment at a stress clinic in Sweden, Kjellgren and Buhrkall (2010, p. 470) qualitatively explored differences in restoration after direct exposure to Swedish woodland and exposure to the same environment mediated through photographs. In the mediated exposure condition themes regarding an absence of sensory input were prevalent; e.g., "Missing the smells and sounds." The absence of auditory input was related to potentially negative affective states such as loneliness ("I feel a lonely quietness") although another participant framed the lack of sound in a more positive way: "Peace and quiet." In contrast, themes from the direct exposure condition reflected increased sensory awareness ("After awhile I hear more and more sounds of nature"… "My senses feel heightened now;" Kjellgren and Buhrkall, 2010, p. 469). These data suggest that experiencing the mediated natural environment, lacking in sound, was unsatisfactory for some participants to the extent that it caused varying perceptions of stress, boredom, and lack of concentration. Kjellgren and Buhrkall (2010) suggest that this may be due to a lack of presence in the mediated environment.

This perspective is supported by qualitative participant comments in an otherwise quantitative study conducted by Annerstedt et al. (2013) regarding virtual reality experience of a forest with and without nature sounds. When the sounds were absent the forest was regarded as unsettling, as though something was missing. In quantitative analysis of data provided by Swedish residents, Grahn and Stigsdotter (2003, p. 7) observed that areas of green space such as quiet parks, rated as helpful when feeling stressed or worried, do not lack sounds completely but feature "sounds of the wind, birds, water, etc." Similarly, Björk et al. (2008, p. 3) note that serenity and lushness are desirable characteristics of natural environments, where serenity is defined as "sounds of wind, water, birds, and insects" and lushness as "a place rich in species."

The sounds of nature are an integral part of environmental experience and appreciation (Mace et al., 2004) and quantitative studies also show that they play an important role in the way natural environments are perceived. For example, supplying nature sounds alongside visuospatial nature stimuli can significantly enhance positive appraisals of the setting, including preference and perceived restorativeness (e.g., Anderson et al., 1983; Jahncke et al., 2015; Franěk et al., 2018; Zhao et al., 2018; Zhu et al., 2020]. This may be due to an increased sense of presence in the environment generated by greater sensory input and awareness as a result of the presence of sound. Support for this argument comes from a body of qualitative work, described below, in which the experience of natural sounds is expressed as a desirable and immersive aspect of being in nature.

Overall, exposure to aspects of nature beyond the purely visual—including sounds—appears related to a greater sensory awareness, immersion in, and sense of presence within nature. This immersion is described in positive terms by participants in qualitative studies such as Curtin (2009) and Kjellgren and Buhrkall (2010), whereas the lack of immersion offered by visual experience of nature only is seen as less positive in comparison (Annerstedt et al., 2013). These findings suggest that natural sounds may offer benefits to restorative perceptions and experiences by affording a greater sense of realism and immersion in nature.

APPRAISALS OF NATURE SOUNDS AS PLEASANT, RELAXING, AAND POTENTIALLY RESTORATIVE

Perhaps the largest body of literature on human experiences of natural sounds relates to how they are affectively and cognitively appraised. This literature is both qualitative and quantitative, and these two bodies of work are discussed separately here.

Qualitative Approaches to Appraisals of Natural Sounds

Qualitative research indicates a relationship between the presence of natural sounds and a state of positive affect. In semi-structured interviews with rural-dwelling Portuguese participants, Pereira et al. (2005, p. 26) revealed a theme of "the feeling of joy provided by bird songs." Similarly, Modelmog (2002) interviewed farmers' wives in Ammerland, Germany, about their relationships with nature. A participant associated listening to birdsong with a positive affective state: "In my garden there blooms a sunflower. [...] Sometimes a bird sits on it and sings. This is happiness to me (Modelmog, 2002, p. 120)." Curtin (2009) also reported that participants associated wildlife sounds with changes in psychological states. For one participant, birdsong was associated with a shift from negative to positive affect: "When you have not been sleeping and you wake up very early and you hear the dawn chorus and you hear the birds, you can suddenly in seconds feel uplifted..." (Curtin, 2009, p. 469).

In a series of semi-structured interviews, Ratcliffe et al. (2013) found that members of the British public generally associated the sounds of nature (e.g., water, wind, and birdsong) with perceived restorative experiences such as pleasure, relaxation, and escape from everyday concerns. Kjellgren and Buhrkall (2010, p. 469) reported that participants in their study responded to the sounds of nature with positive affective appraisals and perceptions of reduced arousal, with one participant noting, "The singing of the birds makes me feel relaxed" and another describing "calming sounds" heard in nature. Similarly, in Cerwén's et al. (2016) qualitative study, Swedish patients recovering from stress perceived nature sounds in a rehabilitation garden as a source of pleasure, relaxation, and restoration. While much research on restoration focuses on green space, Nicolosi et al. (2020) identify coastal soundscapes as positive predictors of perceived restoration.

The restorative experiences of blind and visually impaired individuals in nature has been largely neglected in environmental restoration literature, perhaps due to the strong visuo-spatial focus of existing studies, but this body of work can tell us a great deal about perceptions of natural sounds. Shaw et al. (2015) specifically examined the experience of visually impaired individuals in nature via semi-structured interviews. Thematic analysis revealed perceived restoration arising from sounds as a key theme of experiences in nature. One participant, Helen, noted that, "...you hear a lot of birds. That, that gives you a tremendous feeling of well-being [...] a much more peaceful feeling than you have when you are at home" (Shaw et al., 2015, p. 8). In the context of her wider project "Sensing Nature," Bell (2019a,b) also reported that individuals living with sight impairment used sound as means of connecting with nature, and particularly with wildlife, and experienced positive affective states such as pleasure, freedom, and reduced vulnerability as a result.

Quantitative Approaches to Appraisals of Natural Sounds

The work referenced above indicates that natural sounds are often related to affective states of pleasure and relaxation. Quantitative evidence suggests a similar story and contrasts these positive appraisals with more negative evaluations of anthropogenic sounds. For example, Kariel (1980) recruited individuals from the general public and a mountaineering population and found that both samples considered nature-based sounds of wind, water, wild native fauna (including birds and insects) pleasing or agreeable, whereas the sounds of people and technology were considered neutral or acceptable at best and annoying at worst. Both samples rated the top three sounds (wind, water, and wild animals) equally pleasant. Similarly, Anderson et al. (1983) observed that sounds such as wind, insects, and birdsong were most preferred amongst a range of natural, human, and mechanical sounds.

Assessments of these sounds as pleasant has implications for how beneficial they may be to listeners. Medvedev et al. (2015) integrated subjective ratings of environmental sounds and objective measures of stress recovery to show that ratings of natural sounds as pleasant were related to their ability to aid recovery from stress. In a questionnaire study of Swedish residents, Hedblom et al. (2017) found that women and older participants in particular reported finding nature sounds (such as birdsong and wind in leaves) calming, suggesting potential interactions between sound appraisal and demographics or individual differences.

Using the purpose-developed Perceived Restorativeness Soundscape Scale (PRSS), Payne (2013) differentiated between the perceived restorativeness of urban, urban park, and rural soundscapes in a lab setting, with the rural soundscape (comprising birds, water, and wind) scoring most highly. Similarly, Emfield and Neider (2014) observed that natural sounds of the sea and seagulls were rated as more relaxing than sounds from the urban environment. Even when differences in positive appraisals were controlled for, Kryzwicka and Byrka (2017) found that nature soundscapes were perceived as more restorative. These findings indicate that typical sounds and soundscapes of nature are considered more restorative than those from the built environment, echoing the distinction found between visuo-spatial natural and urban environments (Hartig et al., 2014).

Not All Sounds in Nature Are Perceived as Pleasant

There is, however, evidence to suggest that not all nature sounds are regarded equally positively. In a ratings study of fifteen natural sounds, Björk (1985) found that the songs of chaffinches and other songbirds were rated as more pleasant than the calls of lapwings or gulls. Bradley and Lang (2007) measured 167 sounds on scales of pleasure, arousal, and dominance, of which 21 sounds were from natural sources such as animals (including birds), water, and wind. Some natural sounds, such as water and birds, scored relatively high on pleasure while others, such as growling, were rated as less pleasant, indicating that although natural sounds may generally be perceived as pleasant there is variation depending on the type of sound and its source. Similar findings are reported by Hume and Ahtamad (2013), in which wave sounds and birdsong were rated as very pleasant but the sound of foxes was not. Work by Ratcliffe et al. (2013, 2016, 2020) shows that there is variation even within a single category of nature sound (bird songs and calls): songbirds are qualitatively and quantitatively regarded as more pleasant, relaxing, and potentially restorative than birds which make rough, noisy, and simple calls, or those which have negative meanings or associations. Zhao et al. (2020) have linked crow sounds specifically to lower evaluations of the perceived restorativeness of park soundscapes, while woodpeckers and sparrows are related to more positive evaluations. These findings suggest that variations in preference and perceived restorative value exist even between types of nature sound within the same category. Moreover, combinations of natural sound that reflect biodiversity are also positively regarded. Hedblom et al. (2014) observed that combinations of bird sounds were rated as more pleasant than the sounds of a single species, which may be linked to positive perceptions of biodiversity. This is supported by findings that locations judged to be rich in bird sound are also perceived as more restorative (Fisher et al., 2021), including when such sounds are experimentally manipulated (Ferraro et al., 2020).

NATURE SOUNDS CAN LEAD TO RESTORATIVE OUTCOMES

In many experimental studies that examine restorative outcomes, sounds have been included as part of the experimental stimuli (audible *in situ* or through audio-visual recordings) but their contributions to the restorative experience were not specifically examined (e.g., Ulrich et al., 1991; Hartig et al., 2003; van den Berg et al., 2003; Berman et al., 2008). A growing body of literature has set out to address this. In the following section this is reviewed in two parts: research relating to subjectively measured restoration, and that relating to objective measures (i.e., change in physiological state and/or performance on cognitive tasks).

Subjectively Measured Restoration

Jahncke et al. (2011) examined the restorative effect of a 7-min exposure to audio-visual media of a river, audio media of a river only, silence, or high office noise. Participants who experienced audio-visual media of the river self-reported having more energy than those who experienced only river only or high noise conditions. Both audio-visual and audio exposure to the river media resulted in higher self-reported motivation to work than exposure to office noise. This suggests that experience of nature sounds contextualized by visuals may produce self-reported restorative outcomes, although Ma and Shu (2018) have reported restorative effects of nature sounds independently of visual stimuli.

Studies exploring the restorative effects of natural sounds, separate from visual experience, have until recently been relatively limited. Goel and Etwaroo (2006) observed that exposure to a recording of birdsong combined with classical music significantly reduced self-reported depression and anger in a sample of University students, both depressed and non-depressed. While the findings suggest that listening to birdsong, among other sounds, can have rapid effects on self-reported mood, the study does not dissociate the effects of birdsong from the effects of music, a stimulus which is well-known to induce affective change (see McDermott, 2012, for a review). In a laboratory experiment, Benfield et al. (2014) exposed participants to a stress- and negative affect-inducing video and then to one of four conditions: natural sounds (birdsong and rustling leaves); natural sounds plus traffic; natural sounds plus voices; or a control condition with no audio present. Only in the natural sounds condition did participants show improvements in mood, while participants in the other three conditions showed either declines or non-significant increases. Effects on arousal were not investigated in this study; however, Ma and Shu (2018) examined responses to nature sounds within a simulated open-plan office and found that water and birdsong sounds significantly aided recovery from self-reported annoyance as well as fatigue.

Objectively Measured Restoration

In line with physiological effects of soundscapes more broadly (see Erfanian et al., 2019, for a review), studies that objectively measure the effects of nature sounds on stress recovery reveal mixed results. On the one hand, Annerstedt et al. (2013) observed that experiencing a virtual reality forest environment with birds and water sounds aided recovery from a social stress task (measured via change in heart rate variability) to a greater extent than experiencing the forest environment without sounds or no environmental experience. Participants who listened to nature sounds for 7 min in a waiting room setting showed significantly reduced pulse rate and muscle tension, whereas those who listened to classical music or silence did not (Largo-Wight et al., 2016). Alvarsson et al. (2010) found that stress recovery, as measured by change in skin conductance level (SCL), was significantly greater when participants listened to birdsong and water sounds mixed together vs. loud traffic noise. Recovery from stress tended to be faster, although not significantly greater, in the nature condition than in response to quiet traffic noise or ambient environmental noise. Hedblom et al. (2019) reported no significant differences in stress recovery between three sound conditions (birdsong, traffic, and birdsong + traffic).

Alvarsson et al. (2010) reported faster recovery in the nature condition than the low noise condition even though these were presented at the same sound pressure level (50dB LAeq, 4 min), suggesting that differences in the loudness vs. quietness of an acoustic environment may not be completely responsible for stress recovery. Instead, they suggest that the perceived pleasantness of the sounds may also be relevant and could pertain to their semantic content rather than merely their acoustic properties. In an extension of this work where sound pressure levels were controlled at an average of 64 dB SPL[1] across conditions, Medvedev et al. (2015) observed faster decreases in skin conductance level following stress when participants were exposed to bird and water sounds, vs. sounds from the built environment.

Similar results are reported in studies of objective psychophysiological responses to natural sounds, even in the absence of a prior stress/fatigue condition. For example, Gould van Praag et al. (2017) found that participants who listened to familiar nature sounds showed better attentional monitoring and increased parasympathetic nervous activity than those who listened to artificial sounds. Jo et al. (2019) found that participants who experienced sounds of the forest displayed reduced signs of physiological arousal (i.e., reduced sympathetic nervous system activity) as compared to those who experienced urban sounds. Li and Kang (2019) found that listening to 5-min nature sound recordings (birdsong, ocean waves) led to reductions in certain signs of physiological arousal, including heart rate and respiration frequency and depth, whereas street and traffic soundscapes did not. Contrastingly, Hume and Ahtamad (2013) observed small but significant reductions in heart rate after listening to short (8-s) clips of unpleasant sounds.

Literature regarding effects of natural sounds on objective measures of cognitive performance, or cognitive restoration after fatigue, is also somewhat contradictory. On one hand, Emfield and Neider (2014) reported no significant differences in change in cognitive performance (as measured via pre- and post-exposure administration of a battery of cognitive tasks) as a result of listening to ocean and bird sounds, vs. urban sounds. Abbott et al. (2016) reported only marginally significant restorative effects of nature sounds on cognitive performance as measured via a backwards digit span task (BDST). On the other hand, Van Hedger et al. (2019a) reported significant improvements in cognitive performance (as measured by a composite dual n-back task and BDST) among participants exposed to nature sounds, as opposed to urban sounds, although surprisingly no such effects were found on change in affect. Among samples of school children, Shu and Ma (2019) found that listening to nature sounds (birdsong, water sounds) led to faster responses on a sustained attention to response task (SART) and increased performance on a digit span task (DST). In an *in situ* study in China where sound recordings were experimentally manipulated, Zhang et al. (2017) found that participants exposed to nature-based sounds showed greater attention restoration (as measured via performance on a mental arithmetic task) than those exposed to traffic or machinery sounds.

The studies reviewed above indicate that natural sounds can, in some cases, generate restorative outcomes in terms of affect, psychophysiological arousal, and cognition. These can occur separately from visual exposure to nature but may be enhanced by the presence of visual stimuli.

THEORETICAL APPROACHES TO SOUNDS AND RESTORATION

Since the 1980s two key theories have attempted to explain why certain environments, and particularly nature, can facilitate restoration. These are attention restoration theory (ART; Kaplan and Kaplan, 1989; Kaplan, 1995) and stress reduction theory (SRT; Ulrich, 1983; Ulrich et al., 1991). As mentioned at the start of this review these theories focus predominantly on visual experience of natural environments. Ulrich (1983, p. 86) observes that "many sounds and smells in natural settings surely also influence our feelings," but ultimately focuses on visuo-spatial and aesthetic properties of the environment that can influence affective appraisals and reductions in arousal. These include visual complexity, pattern, depth of scene, surface texture, deflected vistas, and affordances of resources (e.g., water) and threats (e.g., predatory animals). A more recent processing fluency account (PFA; Joye and van den Berg, 2011) challenges some of the psycho-evolutionary principles behind SRT but this too is framed in terms of ease of processing of visual environmental properties.

ART (Kaplan and Kaplan, 1989; Kaplan, 1995) proposes that restoration is driven by cognitive experiences of soft fascination or effortless attention to the environment, a sense of psychological escape or being away, spatial extent, and person-environment compatibility. While these concepts should, in principle, apply to different types of sensory experience, in practice the theory relies heavily on visual examples of such experiences (e.g., experiencing fascination by looking at natural phenomena) to illustrate relevant concepts. Sounds are not mentioned in the original theoretical work, yet visually complex scenes can be represented through acoustically complex

[1]Further information on the acoustic presentation, e.g., A-weighted SPL, is not given in the cited paper and therefore cannot be commented upon here.

soundscapes (Andringa and Lanser, 2013) and in developing a measure of perceived restorativeness of soundscapes Payne (2013) has shown that ART can be applied to acoustic experiences. Even the acoustic and aesthetic properties of individual sounds (bird songs and calls) are related to assessments of perceived restorativeness (Ratcliffe et al., 2020). Work by Qiu et al. (2021) on natural soundscapes during the COVID-19 pandemic suggests that acoustic features of an environment can impact directly on appraisals of ART constructs of extent and fascination, while being away and compatibility may be indirect products of these appraisals. It seems timely to evaluate and update key theories of restoration in order to include acoustic properties in the same way that low-level visual features of environments are considered (see, e.g., Schertz and Berman, 2019).

ASSOCIATIONS WITH NATURE: A NEW AVENUE FOR THEORETICAL DEVELOPMENT?

A concept that is alluded to in SRT (Ulrich, 1983) but is otherwise not greatly explored is the potential relationship between semantic properties specific to an individual and restorative perceptions and outcomes in response to nature. Such properties might include associations, memories, or meanings otherwise linked to the environment (Stigsdotter et al., 2017). Researchers studying the restorative effects of natural sounds and soundscapes have argued for a more interpretative, constructionist approach to how individuals perceive and respond to these environments; i.e., that individuals experience natural soundscapes as a result of both bottom-up, perceptually driven processes and those that are top-down, based on existing preferences, attitudes, and cognitions (Payne, 2008; Ratcliffe et al., 2016). Compatibility between individuals and their soundscapes is also emphasized as a key predictor of perceived restoration by Qiu et al. (2021).

On listening to nature sounds individuals may visualize their own imagined natural environments (Ratcliffe et al., 2016; Bates et al., 2020). There is experimental evidence regarding influences of imagination and learned association on perceptions of the restorativeness of sounds. Listening to pink and white noise reduced self-reported feelings of exhaustion when participants were told that it was the sound of a waterfall, as opposed to a machine, despite the sound itself remaining objectively the same (Haga et al., 2016). Similarly, Van Hedger et al. (2019b) found that preference for natural over urban sounds was dependent on the sounds being recognizable as from these respective categories, and that when this was not recognizable the acoustic properties that characterized these sounds were not in themselves predictors of preference. This recent body of work emphasizes the contribution that semantic, associative interpretations of nature make to restorative perceptions and outcomes, over and above any such effects resulting from perceptual, sensory experiences. As a next step, restoration researchers could incorporate this constructivist approach into theoretical models. This may be achieved by focusing on individual differences in environmental identities, implicit associations with environments or environmental stimuli, perceived rather than objective sources of environmental stimuli, and aspects of individual bonds with place.

DISCUSSION AND CONCLUSIONS

A growing body of work demonstrates restorative perceptions of and outcomes associated with listening to nature sounds, in line with wider evidence that visuo-spatial experiences of nature can benefit psychological well-being (Hartig et al., 2014). As outlined in this review, birdsong, wind, and water are often considered characteristic of pleasant, tranquil natural environments. The presence of these sounds can enhance immersion and sense of presence in visual or other virtually mediated environments and increase positive appraisals of these settings. These sounds are typically perceived as pleasant and calming although variation in such appraisals exists between different *types* of natural sounds. While evidence for restorative *perceptions* of nature sounds is broadly consistent, the evidence for restorative *outcomes* (both cognitive and affective) arising from such exposure is somewhat inconsistent. This may be a result of the different methodologies used in these studies, which are themselves still limited in number in comparison to research on visual experience of nature.

This review has also considered the attention given to acoustic stimuli by key restorative environments theories; i.e., ART (Kaplan and Kaplan, 1989) and SRT (Ulrich, 1983). It is notable that vision and visual examples from nature dominate the original works in which these theories were set out. It is evidently the case that natural sounds can lead to some form of restoration, be it subjectively perceived or objectively measured, and therefore the key theories will need to change to better accommodate acoustic environmental factors in the same way that they do visual properties of environments. For example, in addition to including surface texture and depth of scene as SRT does, a modified theory or model might also include sound intensity and frequency. Finally, work by Haga et al. (2016) and Van Hedger et al. (2019a) suggests that the semantic value of natural sounds (i.e., as natural, and therefore positive) may inform restorative perceptions, beyond mere evaluation of perceptual properties. Greater focus within restoration theory on the meanings that people associate with environments is overdue and may serve to better explain how and why different settings—within and beyond nature and experienced through a variety of sensory means—can support psychological well-being.

AUTHOR CONTRIBUTIONS

ER conducted all aspects of this literature review.

ACKNOWLEDGMENTS

Some of the content of this manuscript has been published as part of ER's Ph.D. thesis (Ratcliffe, 2015). I am grateful to the two reviewers for their comments on earlier versions of this manuscript.

REFERENCES

Abbott, L. C., Taff, D., Newman, P., Benfield, J. A., and Mowen, A. J. (2016). The influence of natural sounds on attention restoration. *J. Park Recreation Adm.* 34, 5–15. doi: 10.18666/JPRA-2016-V34-I3-6893

Aletta, F., and Kang, J. (2019). Promoting healthy and supportive acoustic environments: going beyond the quietness. *Int. J. Environ. Res. Public Health* 16:4988. doi: 10.3390/ijerph16244988

Aletta, F., Oberman, T., and Kang, J. (2018). Associations between positive health-related effects and soundscapes perceptual constructs: a systematic review. *Int. J. Environ. Res. Public Health* 15:2392. doi: 10.3390/ijerph15112392

Alvarsson, J. J., Wiens, S., and Nilsson, M. E. (2010). Stress recovery during exposure to nature sound and environmental noise. *Int. J. Environ. Res. Public Health* 7, 1036–1046. doi: 10.3390/ijerph7031036

Anderson, L. M., Mulligan, B. E., Goodman, L. S., and Regen, H. Z. (1983). Effects of sounds on preferences for outdoor settings. *Environ. Behav.* 15, 539–566. doi: 10.1177/0013916583155001

Andringa, T. C., and Lanser, J. J. L. (2013). How pleasant sounds promote and annoying sounds impede health: a cognitive approach. *Int. J. Environ. Res. Public Health* 10, 1439–1461. doi: 10.3390/ijerph10041439

Annerstedt, M., Jönsson, P., Wallergård, M., Johansson, G., Karlson, B., Grahn, P., et al. (2013). Inducing physiological stress recovery with sounds of nature in a virtual reality forest — results from a pilot study. *Physiol. Behav.* 118, 240–250. doi: 10.1016/j.physbeh.2013.05.023

Bates, V., Hickman, C., Manchester, H., Prior, J., and Singer, S. (2020). Beyond landscape's visible realm: recorded sound, nature, and wellbeing. *Health Place* 61:102271. doi: 10.1016/j.healthplace.2019.102271

Bell, S. L. (2019a). Experiencing nature with sight impairment: seeking freedom from ableism. *Environ. Plan. E* 2, 304–322. doi: 10.1177/2514848619835720

Bell, S. L. (2019b). Nurturing sociality with birdlife in the context of life with sight impairment: a role for nonhuman charisma. *Soc. Cult. Geogr.* doi: 10.1080/14649365.2019.1667018. [Epub ahead of print].

Benfield, J. A., Taff, B. D., Newman, P., and Smyth, J. (2014). Natural sound facilitates mood recovery. *Ecopsychology* 6, 183–188. doi: 10.1089/eco.2014.0028

Berman, M. G., Jonides, J., and Kaplan, S. (2008). The cognitive benefits of interacting with nature. *Psychol. Sci.* 19, 1207–1212. doi: 10.1111/j.1467-9280.2008.02225.x

Björk, E. A. (1985). The perceived quality of natural sounds. *Acustica* 57, 185–188.

Björk, J., Albin, M., Grahn, P., Jacobsson, H., Ardö, J., Wadbro, J., et al. (2008). Recreational values of the natural environment in relation to neighbourhood satisfaction, physical activity, obesity and wellbeing. *J. Epidemiol. Commun. Health* 62:e2. doi: 10.1136/jech.2007.062414

Bradley, M. M., and Lang, P. J. (2007). *The International Affective Digitized Sounds (2nd Ediiton, IADS-2): Affective ratings of sounds and instruction manual.* Technical Report, University of Florida, Gainesville, FL, USA.

Cerwén, G., Pedersen, E., and Pálsdóttir, A.-M. (2016). The role of soundscape in nature-based rehabilitation: a patient perspective. *Int. J. Environ. Res. Public Health* 13:1229. doi: 10.3390/ijerph13121229

Collado, S., Staats, H., Corraliza, J. A., and Hartig, T. (2017). "Restorative environments and health," in *International Handbooks of Quality-of-Life*, Handbook of Environmental Psychology and Quality of Life Research, eds G. Fleury-Bahi, E. Pol, and O. Navarro (Cham: Springer International Publishing), 127–148. doi: 10.1007/978-3-319-31416-7_7

Conniff, A., and Craig, T. (2016). A methodological approach to understanding the wellbeing and restorative benefits associated with greenspace. *Urban Forestry Urban Greening* 19, 103–109. doi: 10.1016/j.ufug.2016.06.019

Curtin, S. (2009). Wildlife tourism: the intangible, psychological benefits of human-wildlife encounters. *Curr. Issues Tourism* 12, 451–474. doi: 10.1080/13683500903042857

De Coensel, B., and Botteldooren, D. (2006). The quiet rural soundscape and how to characterize it. *Acta Acustica United Acustica* 92, 887–897. Available online at: https://www.ingentaconnect.com/content/dav/aaua/2006/00000092/00000006/art00006

Depledge, M. H., Stone, R. J., and Bird, W. J. (2011). Can natural and virtual environments be used to promote improved human health and wellbeing? *Environ. Sci. Technol.* 45, 4660–4665. doi: 10.1021/es103907m

Emfield, A. G., and Neider, M. B. (2014). Evaluating visual and auditory contributions to the cognitive restoration effect. *Front. Psychol.* 5:548. doi: 10.3389/fpsyg.2014.00548

Erfanian, M., Mitchell, A. J., Kang, J., and Aletta, F. (2019). The psychophysiological implications of soundscape: a systematic review of empirical literature and a research agenda. *Int. J. Environ. Res. Public Health* 16:3533. doi: 10.3390/ijerph16193533

Ferraro, D. M., Miller, Z. D., Ferguson, L. A., Taff, B. D., Barber, J. R., Newman, P., et al. (2020). The phantom chorus: birdsong boosts human well-being in protected areas. *Proc. R. Soc. B* 287:20201811. doi: 10.1098/rspb.2020.1811

Fisher, J. C., Irvine, K. B., Bicknell, J. E., Hayes, W. M., Fernandes, D., Mistry, J., et al. (2021). Perceived biodiversity, sound, naturalness and safety enhance the restorative quality and wellbeing benefits of green and blue space in a neotropical city. *Sci. Total Environ.* 755:143095. doi: 10.1016/j.scitotenv.2020.143095

Francis, C. D., Newman, P., Taff, B. D., White, C., Monz, C. A., Levenhagen, M., et al. (2017). Acoustic environments matter: synergistic benefits to humans and ecological communities. *J. Environ. Manage.* 203, 245–254. doi: 10.1016/j.jenvman.2017.07.041

Franco, L. S., Shanahan, D. F., and Fuller, R. A. (2017). A review of the benefits of nature experiences: more than meets the eye. *Int. J. Environ. Res. Public Health* 14:864. doi: 10.3390/ijerph14080864

Franěk, M., Režný, L., Šefara, D., and Cabal, J. (2018). Effect of traffic noise and relaxations sounds on pedestrian walking speed. *Int. J. Environ. Res. Public Health* 15:752. doi: 10.3390/ijerph15040752

Fredrickson, L. M., and Anderson, D. H. (1997). A qualitative exploration of the wilderness experience as a source of spiritual inspiration. *J. Environ. Psychol.* 19, 21–39. doi: 10.1006/jevp.1998.0110

Goel, N., and Etwaroo, G. R. (2006). Bright light, negative air ions and auditory stimuli produce rapid mood changes in a student population: a placebo-controlled study. *Psychol. Med.* 36, 1253–1263. doi: 10.1017/S0033291706008002

Gould van Praag, C. D., Garfinkel, S. N., Sparasci, O., Mees, A., Philippides, A. O., Ware, M., et al. (2017). Mind-wandering and alterations to default mode network connectivity when listening to naturalistic versus artificial sounds. *Sci. Rep.* 7:45273. doi: 10.1038/srep45273

Grahn, P., and Stigsdotter, U. K. (2003). Landscape planning and stress. *Urban Forestry Urban Greening* 2, 1–18. doi: 10.1078/1618-8667-00019

Haga, A., Halin, N., Holmgren, M., and Sörqvist, P. (2016). Psychological restoration can depend on stimulus-source attribution: a challenge for the evolutionary account? *Front. Psychol.* 7:1831. doi: 10.3389/fpsyg.2016.01831

Hartig, T., Evans, G. W., Jamner, L. D., Davis, D. S., and Gärling, T. (2003). Tracking restoration in natural and urban field settings. *J. Environ. Psychol.* 23, 109–123. doi: 10.1016/S0272-4944(02)00109-3

Hartig, T., Mitchell, R., de Vries, S., and Frumkin, H. (2014). Nature and health. *Ann. Rev. Public Health* 35, 207–228. doi: 10.1146/annurev-publhealth-032013-182443

Hedblom, M., Gunnarsson, B., Schaefer, M., Knez, I., Thorsson, P., and Lundström, J. N. (2019). Sounds of nature in the city: no evidence of bird song improving stress recovery. *Int. J. Environ. Res. Public Health* 16:1390. doi: 10.3390/ijerph16081390

Hedblom, M., Heyman, E., Antonsson, H., and Gunnarsson, B. (2014). Bird song diversity influences young people's appreciation of urban landscapes. *Urban Forestry Urban Greening* 13, 469–474. doi: 10.1016/j.ufug.2014.04.002

Hedblom, M., Knez, I., Ode Sang, Å., and Gunnarsson, B. (2017). Evaluation of natural sounds in urban greenery: potential impact for urban nature preservation. *R. Soc. Open Sci.* 4:170037. doi: 10.1098/rsos.170037

Hume, K., and Ahtamad, M. (2013). Physiological responses to and subjective estimates of soundscape elements. *Appl. Acoust.* 74, 275–281. doi: 10.1016/j.apacoust.2011.10.009

International Organization for Standardization (2014). *Acoustics – Soundscape – Part 1: Definition and Conceptual Framework (ISO 12913-1: 2014).* Available online at: https://www.iso.org/standard/52161.html (accessed March 01, 2021).

Iyendo, T. O. (2016). Exploring the effect of sound and music on health in hospital settings: a narrative review. *Int. J. Nurs. Stud.* 63, 82–100. doi: 10.1016/j.ijnurstu.2016.08.008

Jahncke, H., Eriksson, K., and Naula, S. (2015). The effects of auditive and visual settings on perceived restoration likelihood. *Noise Health* 17, 1–10. doi: 10.4103/1463-1741.149559

Jahncke, H., Hygge, S., Halin, N., Green, A. M., and Dimberg, K. (2011). Open-plan office noise: cognitive performance and restoration. *J. Environ. Psychol.* 31, 373–382. doi: 10.1016/j.jenvp.2011.07.002

Jo, H., Song, C., Ikei, H., Enomoto, S., Kobayashi, H., and Miyazaki, Y. (2019). Physiological and psychological effects of forest and urban sounds using high-resolution sound sources. *Int. J. Environ. Res. Public Health* 16:2649. doi: 10.3390/ijerph16152649

Joye, Y., and van den Berg, A. (2011). Is love for green in our genes? A critical analysis of evolutionary assumptions in restorative environments research. *Urban Forestry Urban Greening* 10, 261–268. doi: 10.1016/j.ufug.2011.07.004

Kaplan, R., and Kaplan, S. (1989). *The Experience of Nature: A Psychological Perspective.* New York, NY: Cambridge University Press.

Kaplan, S. (1995). The restorative benefits of nature: toward an integrative framework. *J. Environ. Psychol.* 15, 169–182. doi: 10.1016/0272-4944(95)90001-2

Kariel, H. G. (1980). Mountaineers and the general public: a comparison of their evaluation of sounds in a recreational environment. *Leis. Sci.* 3, 155–167. doi: 10.1080/01490408009512932

Kjellgren, A., and Buhrkall, H. (2010). A comparison of the restorative effect of a natural environment with that of a simulated natural environment. *J. Environ. Psychol.* 30, 464–472. doi: 10.1016/j.jenvp.2010.01.011

Kryzwicka, P., and Byrka, K. (2017). Restorative qualities of and preference for natural and urban soundscapes. *Front. Psychol.* 8:1705. doi: 10.3389/fpsyg.2017.01705

Largo-Wight, E., O'Hara, B. K., and Chen, W. W. (2016). The efficacy of a brief nature sound intervention on muscle tension, pulse rate, and self-reported stress: nature contact micro-break in an office or waiting room. *Health Environ. Res. Design J.* 10, 45–51. doi: 10.1177/1937586715619741

Li, Z., and Kang, J. (2019). Sensitivity analysis of changes in human physiological indicators observed in soundscapes. *Landsc. Urban Plan.* 190:103593. doi: 10.1016/j.landurbplan.2019.103593

Ma, H., and Shu, S. (2018). An experimental study: the restorative effect of soundscape elements in a simulated open-plan office. *Acta Acustica United Acust.* 104, 106–115. doi: 10.3813/AAA.919150

Mace, B. L., Bell, P. A., and Loomis, R. J. (2004). Visibility and natural quiet in national parks and wilderness areas: psychological considerations. *Environ. Behav.* 36, 5–31. doi: 10.1177/0013916503254747

McDermott, J. H. (2012). "Auditory preferences and aesthetics: music, voices, and everyday sounds," in *The Neuroscience of Preference and Choice: Cognitive and Neural Mechanisms*, eds R. J. Dolan, and T. Sharot (London: Academic Press), 227–256. doi: 10.1016/B978-0-12-381431-9.00020-6

Medvedev, O., Shepherd, D., and Hautus, M. J. (2015). The restorative potential of soundscapes: a physiological investigation. *Appl. Acoust.* 96, 20–26. doi: 10.1016/j.apacoust.2015.03.004

Modelmog, I. (2002). 'Nature' as a promise of happiness: farmers' wives in the area of Ammerland, Germany. *Sociol. Ruralis* 38, 109–122. doi: 10.1111/1467-9523.00066

Nicolosi, V., Wilson, J., Yoshino, A., and Viren, P. (2020). The restorative potential of coastal walks and implications of sound. *J. Leis. Res.* 52, 41–61. doi: 10.1080/00222216.2020.1741329

Payne, S. R. (2008). Are perceived soundscapes within urban parks restorative? *Proc. Acoust.* 2008, 5521–5526. doi: 10.1121/1.2935525

Payne, S. R. (2013). The production of a perceived restorativeness soundscape scale. *Appl. Acoust.* 74, 255–263. doi: 10.1016/j.apacoust.2011.11.005

Pereira, E., Queiroz, C., Pereira, H. M., and Vicente, L. (2005). Ecosystem services and human well-being: a participatory study in a mountain community in Portugal. *Ecol. Soc.* 10, 14–36. doi: 10.5751/ES-01353-100214

Pheasant, R., Horoshenkov, K., Watts, G., and Barrett, B. (2008). The acoustic and visual factors influencing the construction of tranquil space in urban and rural environments tranquil spaces-quiet places? *J. Acoust. Soc. Am.* 123, 1446–1457. doi: 10.1121/1.2831735

Qiu, M., Sha, J., and Utomo, S. (2021). Listening to forests: comparing the perceived restorative characteristics of natural soundscapes before and after the COVID-19 pandemic. *Sustainability* 13:293. doi: 10.3390/su13010293

Ratcliffe, E. (2015). *Restorative perceptions and outcomes associated with listening to birds* (Doctoral thesis). University of Surrey, Guildford, United Kingdom. Available online at: http://epubs.surrey.ac.uk/808249/ (accessed March 01, 2021).

Ratcliffe, E., Gatersleben, B., and Sowden, P. T. (2013). Bird sounds and their contributions to perceived attention restoration and stress recovery. *J. Environ. Psychol.* 36, 221–228. doi: 10.1016/j.jenvp.2013.08.004

Ratcliffe, E., Gatersleben, B., and Sowden, P. T. (2016). Associations with bird sounds: how do they relate to perceived restorative potential? *J. Environ. Psychol.* 47, 136–144. doi: 10.1016/j.jenvp.2016.05.009

Ratcliffe, E., Gatersleben, B., and Sowden, P. T. (2020). Predicting the perceived restorative potential of bird sounds: the roles of acoustics and aesthetics. *Environ. Behav.* 52, 371–400. doi: 10.1177/0013916518806952

Schebella, M. F., Weber, D., Schultz, L., and Weinstein, P. (2020). The nature of reality: human stress recovery during exposure to biodiverse, multisensory virtual environments. *Int. J. Environ. Res. Public Health* 17:56. doi: 10.3390/ijerph17010056

Schertz, K. E., and Berman, M. G. (2019). Understanding nature and its cognitive benefits. *Curr. Dir. Psychol. Sci.* 28, 496–502. doi: 10.1177/0963721419854100

Shaw, B., Coyle, A., Gatersleben, B., and Ungar, S. (2015). Exploring nature experiences of people with visual impairments. *Psyecology* 6, 287–327. doi: 10.1080/21711976.2015.1026086

Shu, S., and Ma, H. (2019). Restorative effects of classroom soundscapes on children's cognitive performance. *Int. J. Environ. Res. Public Health* 16:293. doi: 10.3390/ijerph16020293

Sona, B., Dietl, E., and Steidle, A. (2019). Recovery in sensory-enriched break environments: integrating vision, sound and scent into simulated indoor and outdoor environments. *Ergonomics* 62, 521–536. doi: 10.1080/00140139.2018.1491643

Stigsdotter, U. K., Corazon, S. S., Sidenius, U., Refshauge, A. D., and Grahn, P. (2017). Forest design for mental health promotion—using perceived sensory dimensions to elicit restorative responses. *Landsc. Urban Plan.* 160, 1–15. doi: 10.1016/j.landurbplan.2016.11.012

Ulrich, R. S. (1983). "Aesthetic and affective response to natural environment," in *Human Behavior and Environment*, eds I. Altman, and J. F. Wohlwill (New York, NY: Plenum Press), 85–125. doi: 10.1007/978-1-4613-3539-9_4

Ulrich, R. S., Simons, R. F., Losito, B. D., Fiorito, E., Miles, M. A., and Zelson, M. (1991). Stress recovery during exposure to natural and urban environments. *J. Environ. Psychol.* 11, 201–230. doi: 10.1016/S0272-4944(05)80184-7

van den Berg, A. E., Koole, S. L., and van der Wulp, N. Y. (2003). Environmental preference and restoration: (how) are they related? *J. Environ. Psychol.* 23, 135–146. doi: 10.1016/S0272-4944(02)00111-1

Van Hedger, S. C., Nusbaum, H. C., Clohisy, L., Jaeggi, S. M., Buschkuehl, M., and Berman, M. G. (2019a). Of cricket chirps and car horns: the effect of nature sounds on cognitive performance. *Psychon. Bull. Rev.* 26, 522–530. doi: 10.3758/s13423-018-1539-1

Van Hedger, S. C., Nusbaum, H. C., Heald, S. L. M., Huang, A., Kotabe, H. P., and Berman, M. B. (2019b). The aesthetic preference for nature sounds depends on sound object recognition. *Cogn. Sci.* 43:e12734. doi: 10.1111/cogs.12734

Zhang, Y., Kang, J., and Kang, J. (2017). Effects of soundscape on the environmental restoration in urban natural environments. *Noise Health* 19, 65–72. doi: 10.4103/nah.NAH_73_16

Zhao, Y., Li, H., Zhu, X., and Ge, T. (2020). Effect of birdsong soundscape on perceived restorativeness in an urban park. *Int. J. Environ. Res. Public Health* 17:5659. doi: 10.3390/ijerph17165659

Zhao, Y., Xu, W., and Ye, L. (2018). Effects of auditory-visual combinations on perceived restorative potential of urban green space. *Appl. Acoust.* 141, 169–177. doi: 10.1016/j.apacoust.2018.07.001

Zhu, X., Gao, M., Zhao, W., and Ge, T. (2020). Does the presence of birdsongs improve perceived levels of mental restoration from park use? Experiments on parkways of Harbin Sun Island in China. *Int. J. Environ. Res. Public Health* 17:2271. doi: 10.3390/ijerph17072271

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership