Computers, Environment and Urban Systems 110 (2024) 102112

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Exploring emergent soundscape profiles from crowdsourced audio data

Aura Kaarivuo [a,b,*], Jonas Oppenländer [c], Tommi Kärkkäinen [a], Tommi Mikkonen [a]

[a] Faculty of Information Technology, University of Jyväskylä, P.O. Box 35 (Agora), Jyväskylän yliopisto FIN-40014, Finland
[b] School of Media, Design and Conservation, Metropolia University of Applied Sciences, P.O. Box 4072, Metropolia FIN-00079, Finland
[c] Elisa, Ratavartijankatu 5, Helsinki FIN-00520, Finland

ARTICLE INFO

Keywords:
Soundscapes
Mobile crowdsensing
Machine learning
Emotional information
Perception
Urban planning

ABSTRACT

The key component of designing sustainable, enriching, and inclusive cities is public participation. The soundscape is an integral part of an immersive environment in cities, and it should be considered as a resource that creates the acoustic image for an urban environment. For urban planning professionals, this requires an understanding of the constituents of citizens' emergent soundscape experience. The goal of this study is to present a systematic method for analyzing crowdsensed soundscape data with unsupervised machine learning methods. This study applies a crowdsensed sound- scape experience data collection method with low threshold for participation. The aim is to analyze the data using unsupervised machine learning methods to give insights into soundscape perception and quality.

For this purpose, qualitative and raw audio data were collected from 111 participants in Helsinki, Finland, and then clustered and further analyzed. We conclude that a machine learning analysis combined with accessible, mobile crowdsensing methods enable results that can be applied to track hidden experiential phenomena in the urban soundscape.

1. Introduction

Citizens' experience of the surrounding *soundscape* in rapidly growing, increasingly populous cities is strongly connected to well-being, comfort, and contentment (Kang, 2023; van Kamp, Leidelmeijer, Marsman, & de Hollander, 2003). Characterizing soundscapes of urban areas and defining when they are pleasing to the public has been a long-term goal of many soundscape research projects (Gontier, Aumond, Lagrande, Lavandier, & Petiot, 2018; Kang, 2023; Raimbault & Dubois, 2005; Xiao, Lavia, & Kang, 2018). The project of understanding and developing the quality of a soundscape dates back to R. Murray Schafer's "World Soundscape Project" in the 1960s (Schafer, 1977). In this international multidisciplinary project, Schafer aimed to find a sound- scape in which human society and the acoustic environment were in balance (Schafer, 1977). The term *acoustic environment* refers to the combination of sounds of a place or space that are modified by the environment (ISO, 2014) and can be heard (Brown, Gjestland, & Dubois, 2015). *A soundscape* is a person's perceptual concept (ISO, 2014) of the acoustic environment in question (Brown et al., 2015).

Urban soundscapes are living, multi-layered, and composed of an ongoing flow of events, (Arkette, 2004). As many of previous studies have concluded, it is difficult and highly problematic to describe the experience of a sound- scape using single words such as "eventful" or "pleasant" (Aletta et al., 2020; Axelsson, Guastavino, & Payne, 2019; Kang, 2023). This is due to the nature of sound and human perception. Sound is time bound and variable, and its percep- tion is dependent on individual and context-related judgment (Raimbault & Dubois, 2005; Schafer, 1977). Momentary changes in the soundscape can drastically change evaluation of it (Axelsson et al., 2019). It is also known that hedonistic judgment affects this evaluation and that individ- ual assessment is often based on semantic evaluation rather than solely on the perception of sound (Dubois, Guastavino, & Raimbault, 2006; Niessen, Cance, & Dubois, 2010). There- fore, due to the individual nature of the auditory perception, one person can evaluate the same sound sources differently than another (Guastavino, 2007; Mitchell, Aletta, & Kang, 2022). Citizen's needs, context, perceptions and experiences affect their evaluation of the soundscape (Yan, Meng, Yang, & Li, 2024). Soundscape experience is also affected by other sensations (smells, visuals, etc.), and the reporting of different perceptions might be conflated (Calleri et al., 2019; Engel, Paas, Schneider, Pfaffenbach, & Fels, 2018; Shao, Hao, Yin, Meng, & Xue, 2022; Wang, Zhang, Xie, Yang, & He, 2022). Several related studies have suggested that there should be

* Corresponding author at: Faculty of Information Technology, University of Jyväskylä, P.O. Box 35 (Agora), Jyväskylän yliopisto FIN-40014, Finland.
E-mail address: aura.kaarivuo@metropolia.fi (A. Kaarivuo).

https://doi.org/10.1016/j.compenvurbsys.2024.102112
Received 12 September 2023; Received in revised form 24 March 2024; Accepted 25 March 2024
Available online 8 April 2024
0198-9715/

more international and interdisciplinary collaboration in sound- scape research, as well as the development of new tools and methodological approaches, as traditional approaches and tools are not sufficient to holisti- cally represent and evaluate soundscapes (Axelsson et al., 2019; Aletta et al., 2020; Mitchell et al., 2022, Song, Meng, Kang, Yang, & Li, 2023). In particular, it would be necessary to develop collection methods and indicators that assess the health-related quality of sound- scapes (Kang, 2023).

According to Potts (2020), the design of cities and their soundscapes should involve a wide range of interest groups and create consensus through social interaction, facilitated by city planners. In recent years, more inter- action, knowledge sharing, and debate between decision makers and inter- ested parties has resulted from the development of communication and mobile technologies (Potts, 2020). *Crowdsourcing* is an example of a method of fa- cilitating participation and democratic decision-making with a large group of dispersed people via the Internet (Brabham, 2013). Crowdsourcing pro- vides a means to gather extensive amounts of situated intelligence (Brabham, 2013) using smart and efficient methods (Liao et al., 2019). Urban crowdsourcing (Steils, Hanine, Rochdane, & Hamdani, 2021) can be used to inform the design of smart cities, using participatory design approaches (Mueller, Lu, Chirkin, Klein, & Schmitt, 2018) instead of traditional, expensive, labor-intensive methods, such as questionnaires or public hearings (Liao et al., 2019). Situated crowdsourcing has an enormous potential in soundscape design, as it allows participants to provide both qual- itative and quantitative information in-situ, in everyday life situations, and in larger groups (Craig, Moore, & Knox, 2017). *Crowdsensing* here refers to col- labo- rations with citizens in which both people and their mobile devices act as sensors (Brambilla & Pedrielli, 2020; Cardone et al., 2013; Lefe- vre, Agarwal, Issarny, & Mallet, 2021). Crowdsensing (or mobile crowdsensing) has been utilized especially for noise monitoring and mapping soundscape quality (Craig et al., 2017; Li, Liu, & Haklay, 2018; Orio, De Carolis, & Liotard, 2021).

Crowdsensed data can provide more diverse information for sound- scape research (Brambilla & Pedrielli, 2020; Gontier et al., 2018; Nieto-Mora, Rodríguez-Buritica, Rodríguez-Marín, Martínez- Vargaz, & Isaza-Narváez, 2023; Zappatore, Longo, & Bochicchio, 2017). According to recent studies, the most common analysis methods consist of manual labeling of data by listening to record- ings or visually inspecting spec- trograms, summarizing variations in acoustic energy, or automatically recognizing sound sources or insides using machine learning algorithms (Benocci, Afify, Potenza, Roman, & Zambon, 2023; Nieto-Mora et al., 2023). However, big audio data cannot be manually labeled and analyzed, due to its time- consuming nature (Benocci et al., 2023; Nieto-Mora et al., 2023). Automatic recognition of acoustic insides and sound sources is sensitive to noise and the sound sources may vary depending on the specific environment being studied. Machine learning methods have been used to identify geographic patterns (Quinn et al., 2022), to evaluate urban spaces (Yu & Kang, 2009), and to classify species and other acoustic features (Dias, Ponti, & Minghim, 2022). Both supervised and unsupervised techniques have offered promising results, but again supervised machine learning is labor intensive and time consuming (Nieto-Mora et al., 2023).

The goal of this paper is to present a systematic method for analyzing crowdsensed soundscape data with unsupervised machine learning methods. We will apply unsupervised machine learning methods to the results of man- ual qualitative data analysis of soundscapes, and observe the resulting clus- ters to obtain information about the perceived quality of the soundscape.

These aims are addressed through the following research questions:

RQ1. How can crowdsensed soundscape data be analyzed using unsupervised machine learning methods?

RQ2. What kind of soundscape profiles emerge from the analysis and how could their interpretation be linked to improve our understanding of urban soundscape experiences?

The rest of this paper is structured as follows. We will present an analy- sis that employs manual labeling, qualitative analysis, and machine learning methods (see Fig. 1) for soundscape data which is collected with participa- tory crowdsensing method. We use methodological triangulation to augment the findings of different analysis methods (Denzin, 1970). First, in Section 2, we describe the data collection, manual labeling and automated analysis of soundscape data, which was based on a combination of unsupervised ma- chine learning and feature selection methods and the results of the qualitative analysis. Second, in Section 3, we provide details of the identified clusters and analyze the groups and profiles of the soundscape experience from the crowdsensed audio data and manual qualitative analysis. We compare the results of the manual qualitative analysis with the results of the unsupervised machine learning approach and, finally, present the general characterization of the emerging soundscape experience. In Section 4, we discuss the inter- pretation and key findings of the research. Finally, in Section 5, we draw conclusions and suggest implications and ideas for future work.

2. Material and methods

Various research and methodological approaches, solutions, and frame- works for soundscape data collection and analysis have been presented over the past five decades at an accelerating pace (Aletta, Kang, & Axelsson, 2016; Guastavino, 2007; Jiang et al., 2022; Kang, 2010, 2023; Kang & Aletta, 2018; Schafer, 1977). The current standardized method is presented in ISO standard 12,913 parts 1–3, which contain a definition of soundscape and a conceptual frame- work, data collection, reporting, and analysis requirements (ISO, 2014, 2018, 2019) for research. According to this ISO standard, a soundscape study should be holistic and contain several investigative methods to ensure that the study considers different viewpoints, such as the human perception, the acoustic environment, and the context in question. The standard does not give a single answer or a clear research approach but recommends a collection of methods because a consensus could not be reached regarding a protocol (Mitchell et al., 2022). Qualitative data analysis is recommended to be done with a chosen coding method to generalize the observations. Quantitative analysis is recommended but is considered less important, especially in cases of qualitative and explorative methods. The analysis of responses about the perceived quality of a soundscape is presented in the following dimension (ISO, 2019):

- pleasant – unpleasant
- calm – chaotic
- vibrant – monotonous
- eventful – uneventful

The following data collection and analysis method loosely follows the ISO standards. With the ISO standard, the fundamental question is that the definitions for dimensions are presented in English, and as Aletta et al. (2020) state in their article, sounds are described in a different way in different languages (Guastavino, 2007). According to Axelsson et al. (2019) context and person-related factors create great variance, which leads to difficulties in interpretation of the results. These and other limitations and perspectives of the critique toward the ISO standards (Aletta et al., 2020; Jo, Seo, & Jeon, 2020; Mitchell et al., 2022) were considered when designing this method. According to the ISO standard the choice of indicators depends on the people, acoustic environment and context.

The data set contained 111 one-minute-long raw audio files and question- naire answers related to them. The data collection method used here follows a method developed and tested by Kaarivuo, Salo, and Mikkonen (2021). The aim of this method was to develop an accessible, mobile, and participatory method that would produce live recordings of a soundscape in addition to traditional written descriptions and questionnaires. The purpose of this approach was to observe emerging pleasant soundscapes that citizens pass through in their everyday lives.

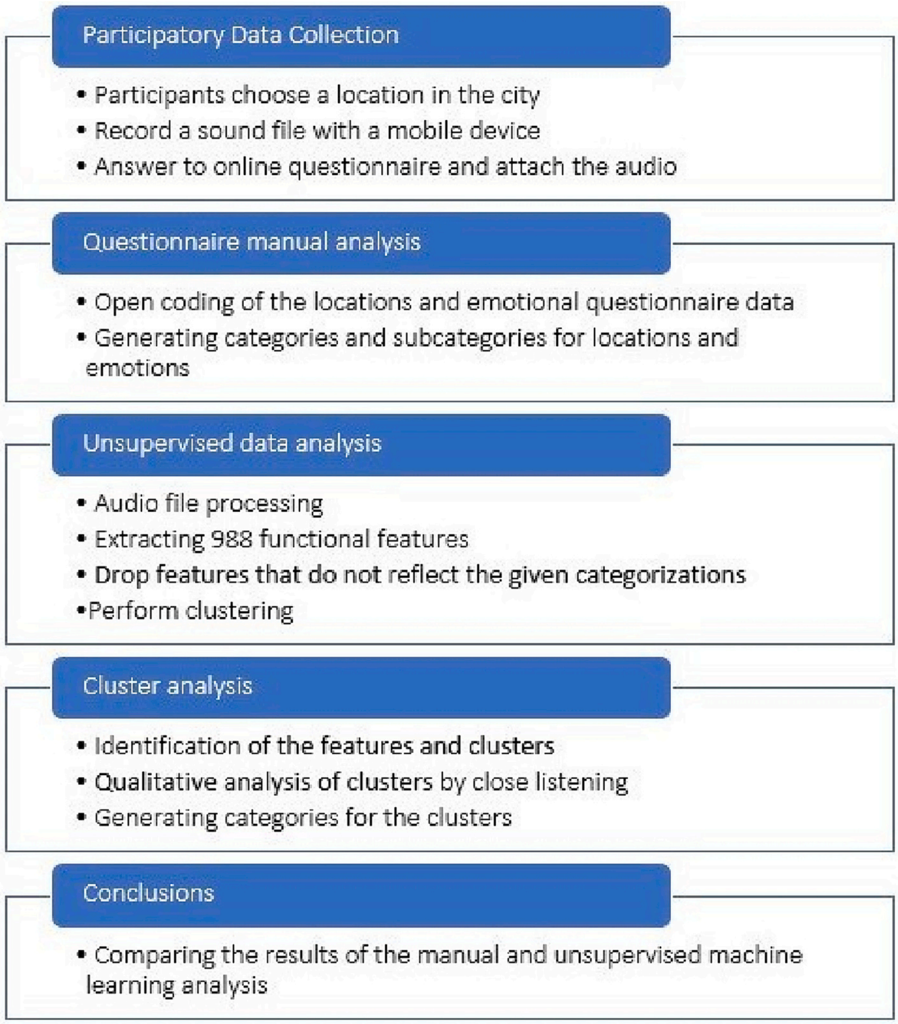


Fig. 1. Data collection and analysis method.

According to the evaluation of technical procedure and the functionality of the mobile data collection method it seemed that the recording with mobile tools and sharing the audio is easy and does not require any specific applications or even technical instructions. The evaluation also showed that this particular method identifies pleasant and easily accessible places in the city in which the participants enjoy in their surroundings. The study concluded that with this method it would be possible to collect training data for machine learning. (Kaarivuo et al., 2021).

2.1. *Participants, context, and data collection*

The research participants were first-year university media production stu- dents at a university of applied sciences located in Helsinki. The experiential soundscape data was collected in three workshops in August and September of 2020–2022 in the greater Helsinki area, which is the home environment for the participants. There were 111 participants in total, 35 to 38 participants per year. Most of the participants (68.5%) were 18 to 25 years old, 28.8% 26 to 35 years old, and 1.8% 36 to 45 years old.

The motivation of the media students to complete the assignment was most likely higher than average due to their motivation and interest in audio and sound design, but their technical competencies or listening and analyzing skills at the beginning of the studies were quite diverse. Most of the students were not familiar with soundscapes, urban planning, or analytical listening.

Participants received a short introduction lesson about the surrounding acoustic environment and a listening and soundscape recording assignment. To strengthen the engagement of the participants, the assignment was designed so that it connected to the participants' personal experiences about the urban soundscape (Neuvonen, 2019). In the assignment, the participants were asked to choose a location in the city in which they found the sound- scape pleasant and comfortable. They were asked to focus and listen to the soundscape for 20 min and record it using any kind of recording device and application for one minute. Next, they were asked to share the recording via an online form and answer questions concerning the soundscape. The questions in the online form were as follows:

- What is the name of the location?
- List the sounds you heard.
- What sounds would you add to the soundscape to make it more pleasant?
- What sounds would you remove or reduce?
- In your own words, describe how the soundscape feels and sounds and justify why. What in the soundscape evokes these feelings?

The online questionnaire was designed to be a combination of a questionnaire and an interview, both of which are mentioned as data collection methods in the ISO 12913-2 standard (ISO, 2018). As the participants were not describing the same locations, it was necessary to collect more detailed information about the soundscape, such as sounds heard in-situ. The list of sounds provided a reference point for comparing the recordings, and the question about emotion provided information about the emotions and fea- tures experiences, such as pleasantness, calmness, vibrancy, eventfulness, and loudness.

As our approach aimed to lower the threshold of participation, self-reporting was made easy. We aimed to design the questions so that they were easy to answer and would produce detailed data about the physical and psycho-acoustic features of the soundscape. The idea was to lead the par- ticipator to first observe their surrounding soundscape in a focused manner, to recognize the elements in the soundscape, and then to create associations between emerging emotions and sounds and feelings. The aim was to create a procedure that can be repeated with any group of people, regardless of their age, education, prior knowledge, or sonological competence.

2.2. Manual qualitative analysis method

The self-reported emotional perceptions of the participants and locations of the recorded soundscapes were manually coded and labeled, drawing on categorizations from the related literature.

The emotional answers were coded under naturally emerging categories, following a grounded theory approach (Glaser & Strauss, 1967), rather than strictly applying the ISO standard labels. The qualitative analysis of the questionnaire was conducted in the following steps:

Step 1. Open coding: recognizing key terms concerning the emotions associated with sounds.

Step 2. Eliminating unnecessary and irrelevant information that is not directly related to the soundscape in question.

Step 3. Identifying repeated words and expressions.

Step 4. Identifying concepts: comparing the emerging terms and expressions to the ISO 12913-3:2019 standard for perceived soundscape affective quality.

Step 5. Generating categories: grouping similar expressions and concepts.

Step 6. Generating subcategories: modifying the chosen framework to illustrate the emerging phenomena.

Step 7. Drawing conclusions from the results.

The testing of the manual analysis indicated that the labeling of freely written Finnish answers with the original ISO standard English dimensions is problematic. The free-form lyrics did not distinguish between, e.g., vibrancy and eventfulness because there is no Finnish translation which would trans- late similarly. Also, the clustering of a small sample requires that the number of evaluation axes is reasonably small. We decided to test the analysis methods on the basis of what emerges from the data. Therefore, the dimensions were narrowed down to three:

pleasant – unpleasant calm – chaotic vibrant – monotonous

According to the reported locations of the recordings, we identified the recording locations and categorized them. The seven identified location cat- egories are as follows:

1. Sports/activity,
2. Street,
3. Social activity,
4. Neighborhood,
5. Station,
6. Park,
7. Miscellaneous.

2.3. Manual qualitative analysis results

In all three rounds, the participants chose locations mainly in the Helsinki metropolitan area in Finland. It seems that the selected locations are close to the places where the students live, commute between home and university, or spend their free time.[1]

The participants recorded mainly street locations, such as bus stops and other places where it is convenient to stay for a while to listen. About one quarter of the participants (23%) selected a park to represent a comfortable soundscape. Residential areas, train and metro stations, sports venues, and cafe terraces were mentioned <10 times each. The miscellaneous category contained recordings that did not meet the requirements of the assignments, and were recorded in indoor spaces such as shopping centers, vehicles, and indoor metro stations. The distribution of the created categories is presented in Fig. 2.

The answers to the question *"In your own words, describe how the sound- scape feels and sounds and justify why. What in the soundscape evokes these feelings?"* produced a variety of thoughts and opinions about the soundscape and the participants' memories, associative thoughts, and emotions and rela- tion toward the sounds and the place. It is well known that people describe their experience of an environment affectively (ISO, 2019). However, the an- swers contained expressions of the pleasantness, calmness, and vibrancy of the places in question, or the opposite.

The pleasant and unpleasant soundscapes were described, for exam- ple, as "homelike," "safe," "cozy," "comfortable," or with words like "gloomy," "restless," "disturbing," and "inharmonious." As the precondition of the task was to go to a place where the soundscape was comfortable, 77.5% of the soundscapes were labeled as pleasant and 22.5% unpleasant.

The calmness and chaos of the places could also be characterized as quiet and loud. As the task concerned urban environments, the word "quietness" did not appear in the answers. These impressions were expressed with words such as "relaxing," "carefree," "serene" or "smooth" and "noisy," "hectic," "busy," or "stentorian." The distribution was fairly even, with 51.4% of the soundscapes being described as calm.

The vibrancy and monotony of the soundscapes were expressed within various contexts. A monotonous soundscape was a place in which partici- pants could pick up "quiet sounds" and "be with your own thoughts" and a vibrant one was "multi-layered" or "eventful" with "continuous stimuli." A soundscape was "morning-like, with only small sounds" or "ordinary and bor- ing." In contrast the soundscapes were "speedy" and had "sounds of life" and "there [was] a lot going on around". Over half (58.6%) of the places were described as vibrant and 41.4% as monotonous.

The self-reported written descriptions of the emotions related to the soundscapes were categorized under three label pairs (see Table 1).

[1] In 2020, the Covid-19 pandemic affected our lives, including social behavior. However, in August–September 2020, the Covid-19 situation in Finland was fairly stable, allowing students to study on campus, use public transportation, and freely move outdoors. Restau- rants and other leisure activities were available, with certain limitations (Ministry of Social Affairs and Health and the National Institute for Health and Welfare, 2020). The circumstances might have affected the participants, choices of recording locations. As the main aim of our study was to develop a method for deriving insights from recorded locations, the circumstances in 2020 did not compromise the collected data and the development of the method.

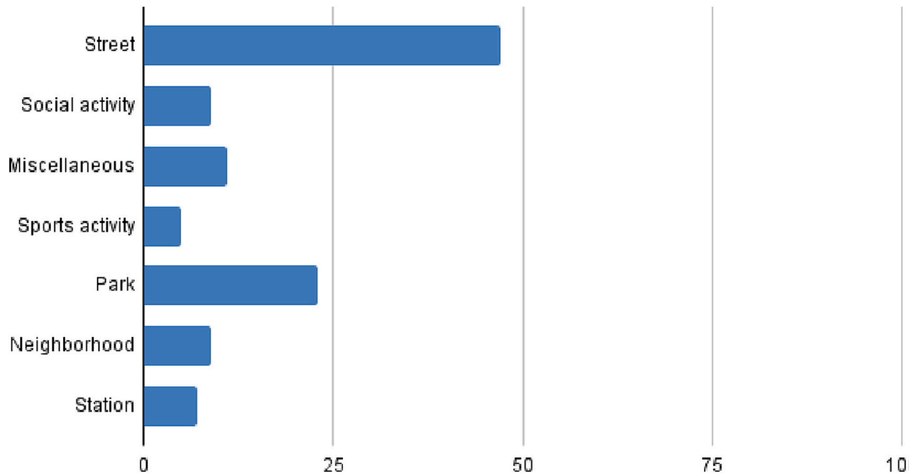


Fig. 2. Distribution of the created location categories per 111 audio samples.

Table 1
Categorization of the 111 soundscapes into three categories according to labeling of the self-reported expressions.

Pleasant		Unpleasant		Quiet		Loud		Monotonous		Vibrant	
86	(77.5%)	25	(22.5%)	57	(51.4%)	54	(48.6%)	46	(41.4%)	65	(58.6%)

2.4. Unsupervised machine learning based analysis

For automatic profiling of the soundscape experience, we applied a four- step procedure (see below Section 2.4) to identify the most important audio features based on the manual qualitative categorization. These most important features are used to link the manually produced knowledge to the raw audio recordings, thereby indicating which features are primarily related to different categories and which features contribute the most to the classification.

The recorded audio files were preprocessed as follows. The audio files were first converted to 16-bit with a sampling rate of 44.2 kHz and two channels with normalized volume, using ffmpeg. Each audio file was then truncated to the median length of all audio files (61.69 s). Files below this minimum length were padded with silence at the end of the audio file. Audio features were extracted using OpenSMILE 3.0.1 (Eyben, Wöllmer, & Schuller, 2010). In total, we extracted 988 functional features using the specification file *emobase.conf*. As summarized in Appendix A, this set of acoustic features that are com- monly used in emotion recognition research (Schuller, Steidl, & Batliner, 2009) contains statistical transformations (e.g., maximum, minimum, range, mean, stddev, skewness, kurtosis, and quartiles) as well as first- and second-order deriva- tives of the following basic groups of audio descriptors: intensity, loudness, spectral envelope, zero crossing, speech probability, fundamental frequency, pitch, and Mel-frequency cepstral coefficients (MFCC). While many of these feature sets relate to the paralinguistic analysis of a voiced speech, emobase has been applied in various other contexts of affective computing, including soundscape analysis (Lionello, Aletta, & Kang, 2020).

We followed a four-step procedure to use the extracted audio features to identify a small set of similar groups of soundscape experiences based on the audio recordings and their qualitative analysis. The first three steps perform a filter-type feature selection (Linja, Hämäläinen, Nieminen, & Kärkkäinen, 2023), and the last step establishes the division into soundscape clusters (Niemelä, Äyrämö, & Kärkkäinen, 2021).

Step 1. The range, *Rng*, of the original 988 emobase features varied in 0–2.14e+4. A range of zero means a constant, noninformative feature. Therefore, features whose range is close to zero are treated as non-informative. There were slightly >100 features with ranges of around 1e-3 or less, so we decided to drop the 102 features whose range was below this threshold. The basis for this decision is illus- trated in Fig. 3 (left).

Step 2. As defined in Cord, Ambroise, and Cocquerez (2006) and applied in, for example, Saarela, Hämäläinen, and Kärkkäinen (2017) and Jääskelä, Heilala, Kärkkäinen, and Häkkinen (2021), the H statistics of the non- parametric Kruskal-Wallis (or Mann-Whitney U for binary labelling) test (Kruskal & Wallis, 1952) can be used to evaluate how well a certain feature signifies a given classification. We computed these values with respect to the three soundscape categorizations that were derived in Section 2.2 (see Table 1). To unify the scale of statistics, all three sets were individually normalized by division of the largest value, resulting in the uniform range [0,1].

Step 3. To ensure that a feature can separate all three of the qualitative categories, we computed the minimum H statistics value over the normalized sets and sorted this vector into decreasing order. These values were then given to the knee point detection algorithm (Kaplan, 2023), which estimated the location where the curve "turns" (the "knee," see Thorndike (1953)). This point provided us the in- dex (351) and the tolerance level (0.05) that identified the point at which additional features signified less correspondence to the three manual classifications. Therefore, these 536 non-strongly separating features on the tail were removed, and we ended up with 350 features that were used in the consequent clustering step. This selection is illustrated in Fig. 3 (right).

Step 4. Because of the non-Gaussian distribution of the features to be ana- lyzed, the robust k-spatmeds++ clustering algorithm (Hämäläinen, Jauhiainen, & Kärkkäinen, 2017) with 1000 repetitions for the number of clusters ranging from $k = 2 \ldots 10$ was applied using the toolbox given in in the study by Niemelä et al. (2021).

The Wemmert-Gancarski (WG) cluster validation index, which was the best performing one in the comparisons of large-dimensional datasets with hundreds of features performed in Niemelä et al. (2021), was applied to es- timate the number of clusters. As depicted in Fig. 4, the best division into nondisjoint clusters is given with three or five clusters. These results are analyzed next.

3. Results

This section first presents the results of the machine learning based

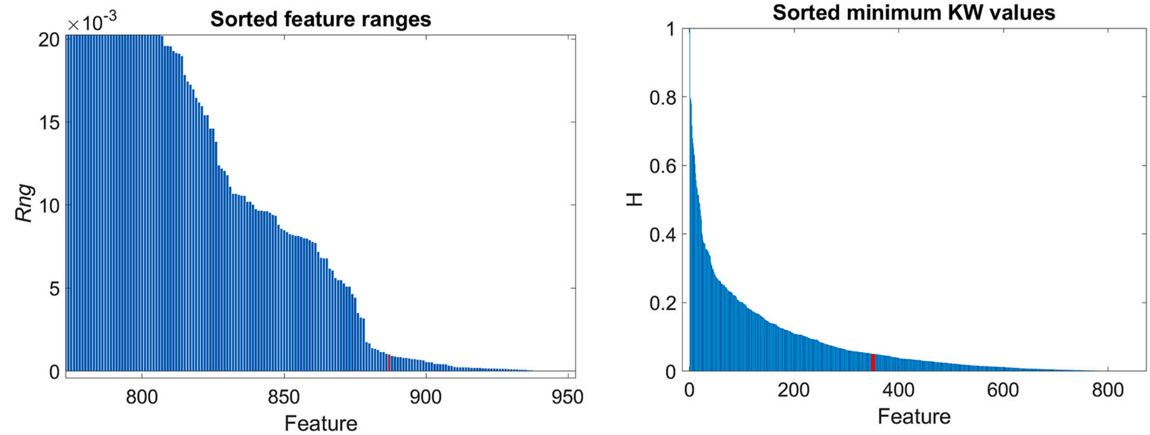


Fig. 3. Identification of noninformative, almost constant features (left). Selection of features using minimal H statistics values and the knee point detection (right).

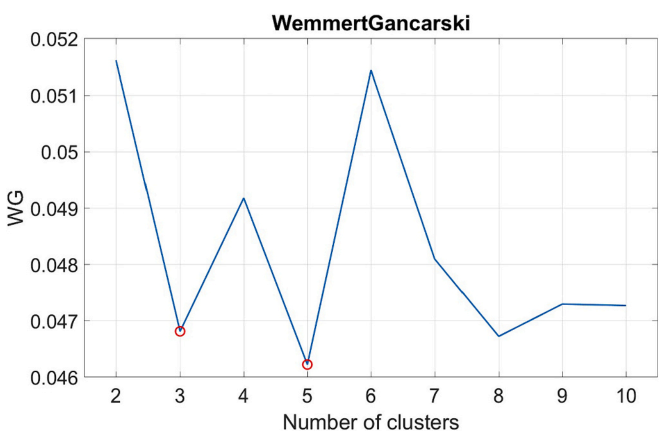


Fig. 4. Behavior of the Wemmert-Gancarski index identifies three- and five-cluster solutions for further analysis.

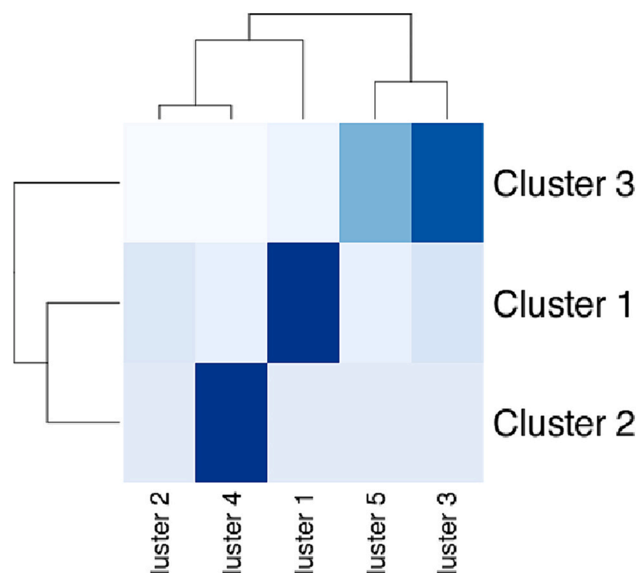


Fig. 5. Cross-tabulation comparing three versus five-cluster solutions.

anal- ysis. Then, soundscape experience profiles are analyzed both quantitatively and qualitatively. We also compare the results of the manual qualitative analysis with the results of the unsupervised machine learning approach and. Finally, we present the general characterization of the emerging soundscape experience.

3.1. Identification of the features and clusters

As depicted in Fig. 4, the clustering of 111 soundscapes represented with the qualitative separation of audio features resulted in two potential solutions: one with three and one with five clusters.

The Pearson's χ^2-test between the two clustering results shows that there is a strong similarity between the two solutions ($\chi^2 = 197$, $p = 0.000$; see Fig. 5).

Given that clusters 1 and 5 in the five-cluster solution contain few observations (8 and 2, respectively), we focused on analyzing the three-cluster solution with respect to which audio features depicting the soundscape can explain the formation of the three groups.

Based on Step two of the four-step procedure and the feature groups in Appendix A, the five features that most strongly separate the three clusters correspond to the smoothed version of the fundamental frequency of the audio signal (F0env_sma), which captures the overall pitch contour of the signal. Interestingly, Raimbault and Dubois (2005) also note that pitch can be related to the non-expert experiences of soundscapes.

We also analyzed which features mostly separated the qualitative clas- sifications, as developed in Section 1. For the pleasant/unpleasant catego- rization, the two best-separating features ("lspFreq_sma_de[3]_kurtosis", "mfcc_sma_de[6]_max") were related to the spectral envelope (i.e., sound quality) and MFCC coefficients (i.e., how people hear sounds). For the quiet/loud division, the three best separating-features were also all related to the spectral envelope. For the vibrant/monotonous categorization, the five most dominant features were again all related to sound quality ("lspFreq_sma" oriented features).

We further analyzed differences in the loudness between the three clusters. A pairwise comparisons using a Wilcoxon rank sum test with continuity correction found significant differences in loudness between cluster 1 ($M = -25.08$) and cluster 3 ($M = -33.91$), $\chi^2 = 41.812$, $df = 2$, $p < 0.0000$; see Fig. 6.

3.2. Analysis of emergent soundscape profiles

As summarized in Table 2, the three-cluster result contained two larger clusters (41 and 64 audios each). The third largest cluster was too small to be analyzed (6 audios) therefor we focus on the comparison of the two main clusters.

As summarized in Appendix A, the set of 988 functional features of the *emobase* configuration from OpenSMILE can be grouped into more general categories. Within the set of 350 features which were included in the cluster analysis in Section 2.4 (Step 3 of the four-step procedure), the

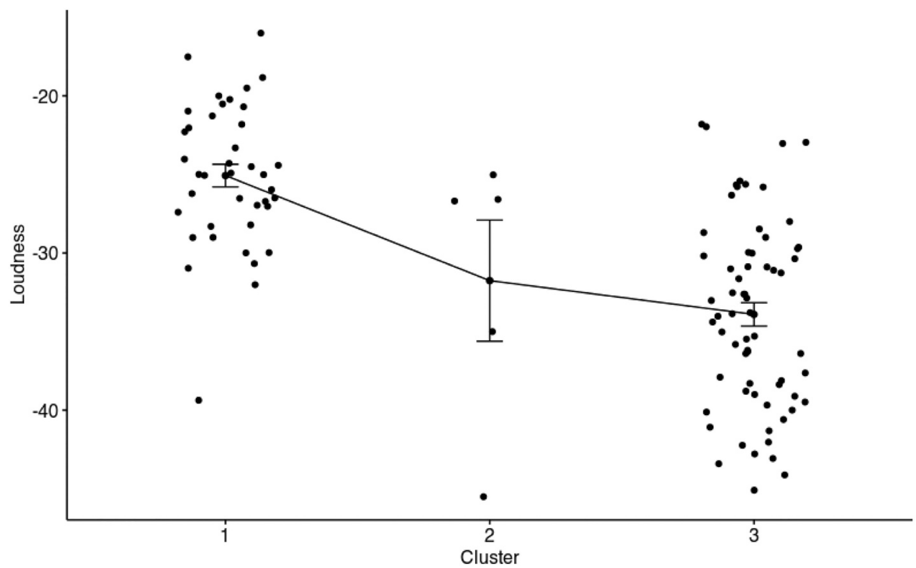


Fig. 6. Loudness of observations in the three clusters.

Table 2
Three-cluster solution.

Cluster	#	Pleasant	Unpleasant		Quiet		Loud	Monotonous		Vibrant	
1	41	28 (68%)	13	(32%)	29	(71%)	12 (29%)	30	(73%)	11	(27%)
2	6	6 (100%)	0	(0%)	2	(33%)	4 (67%)	3	(50%)	3	(50%)
3	64	52 (81%)	12	(19%)	23	(36%)	41 (64$)	33	(52%)	31	(48%)

numbers of features from different categories are given in Table 3. In the table, the name of the feature group, the string that is used to refer to these features in OpenSMILE, number of the selected features (of 350), and, finally, number of the Δ, i.e., difference-based features are given.

For the statistical analysis of the difference between the two main clusters, we first created 14 aggregated variables of the feature groups as given in Ta- ble 3 (7 groups of basic features +7 groups of Δ features) by computing the groupwise means (means in the two columns # and Δ-# in each row). These variables of the 105 observations from the two clusters were then analyzed using again the Kruskal-Wallis test and the corresponding test statistics H. This analysis yielded to the following order of the most separating aggre- gated variables: 1) Pitch, 2) Loudness delta, 3) Fundamental frequency, and 4) Loudness. This shows that highness/lowness, loudness and its changes, and the existence of natural voices (human, bird etc.) most importantly differentiate the soundscape experience in the two main clusters.

The identified two main clusters system audio files were analyzed with spectrum analysis and LUFS (Loudness Unit Full Scale) measuring. Spectrum analysis visualizes the dominant features of the clusters, and LUFS measuring provides a reference for the overall loudness of the audio. As the audio was recorded with lo-fi consumer quality mobile device microphones, it is not prudent to draw conclusions about noise level or any other physical features of the sounds.

Table 3
Numbers of selected features.

Group	OS-name	#	Δ-#
Loudness	'pcm_loudness'	15	14
Mfcc	'mfcc'	79	80
Spectral Envelope	'lspFreq'	62	54
Zero-crossing	'pcm_zcr'	4	6
Voice	'voiceProb'	6	4
Fundamental frequency	'F0'	4	10
Pitch	'F0env'	5	7

The spectrograms of the two main clusters (Fig. 7a and b) also reveal significant differences in the overall loudness of the sound files. The average integrated LUFS levels were − 25 LUFS in cluster 1 and − 34,2 LUFS in cluster 3 which is aligned with the visual observation from the spectrogram.

3.3. *Qualitative analysis of clusters*

The clusters were then manually analyzed by close listening to the audio files and identifying details, such as sound sources and analyzing the struc- ture of the soundscapes (foreground-background structure, dominant sounds, variations of sounds events, context of sounds, and possible recording errors).

The close listening was conducted with the following procedure:

Step 1. Listening through the audio files in each cluster to derive an overview of the material.

Step 2. Listening to each audio file individually and coding the sound sources.

Step 3. Second listening to observe the context and relations of the sounds in each recording.

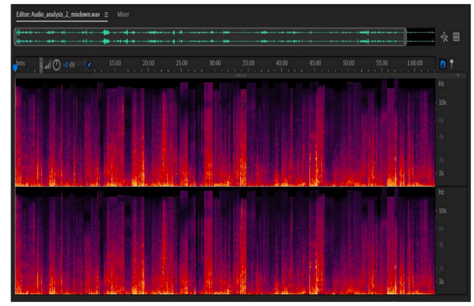
Step 4. Modifying the chosen framework and generating a suitable catego- rization for the research context.

Step 5. Drawing conclusions from the results.

The resulting sound source framework was modified based on the ISO/TS 12913–2:2018 framework (ISO, 2018) in which the urban acoustic environment is divided into *anthropophonic* sounds which are generated by human activity, and *geophonic* and *biophonic* sounds, which are not generated by human activity. In addition, we applied a contextual framework, that defines a hierarchical method that distinguishes background and foreground, disruptive and supportive sounds, and calming and stimulating soundscapes (Sun et al., 2019).

The modification of the ISO (2018) framework aimed to visualize the spa- tial differences and distances in the soundscape structures, so we separated the motorised transport sounds from the anthropophony

(a) In cluster 1 the overall loudness is high



(b) In cluster 3 the overall loudness is lower

Fig. 7. Spectrograms of all audio files of cluster 1 and 3.

category and created sub-categories for loud and distant traffic sounds (Fig. 8). This made it possible to present the hierarchy of the soundscape, as presented in the Sun et al. (2019) framework.

The reported locations (Fig. 9) were mainly streets (53.7%) in cluster 1, and in cluster 3, they were streets (38.5%), parks (24.6%), and neighborhoods (10.8%).

Close listening to the two clusters showed that cluster 1 contained louder motorised vehicle sounds (48.8%) than cluster 3 (30.8%) (Fig. 10). This observation is in line with the loudness observations presented in Fig. 6. Anthropophonic sounds, such as human movement and voices were present in most of the soundscapes, but they were covered by the loud motorised sounds in cluster 1 and therefore were less recognizable and noticeable. Only 17% of the cluster 1 soundscapes contained biophonic wildlife sounds, such as birds, or geophonic sounds such as water and wind. Most likely, they were covered by the traffic and technical sounds. In both categories, over 30% of the sound files contained recording errors, such as wind noise or sounds of handling the recording device.

The sound files in cluster 3 had more perceptible human movement and voices, and sound sources were easier to separate. Motorised transport sounds were present, but in 58% of the recordings, they were less loud, were distant or appeared only from time to time. More delicate sounds, such as human voices, bicycles, birds, breezes, and footsteps, could be heard. Due to the lesser presence of traffic sounds, bird sounds and nature sounds could be heard in this cluster. This supports the finding in the most separating vari- ables, “Pitch” and “Fundamental

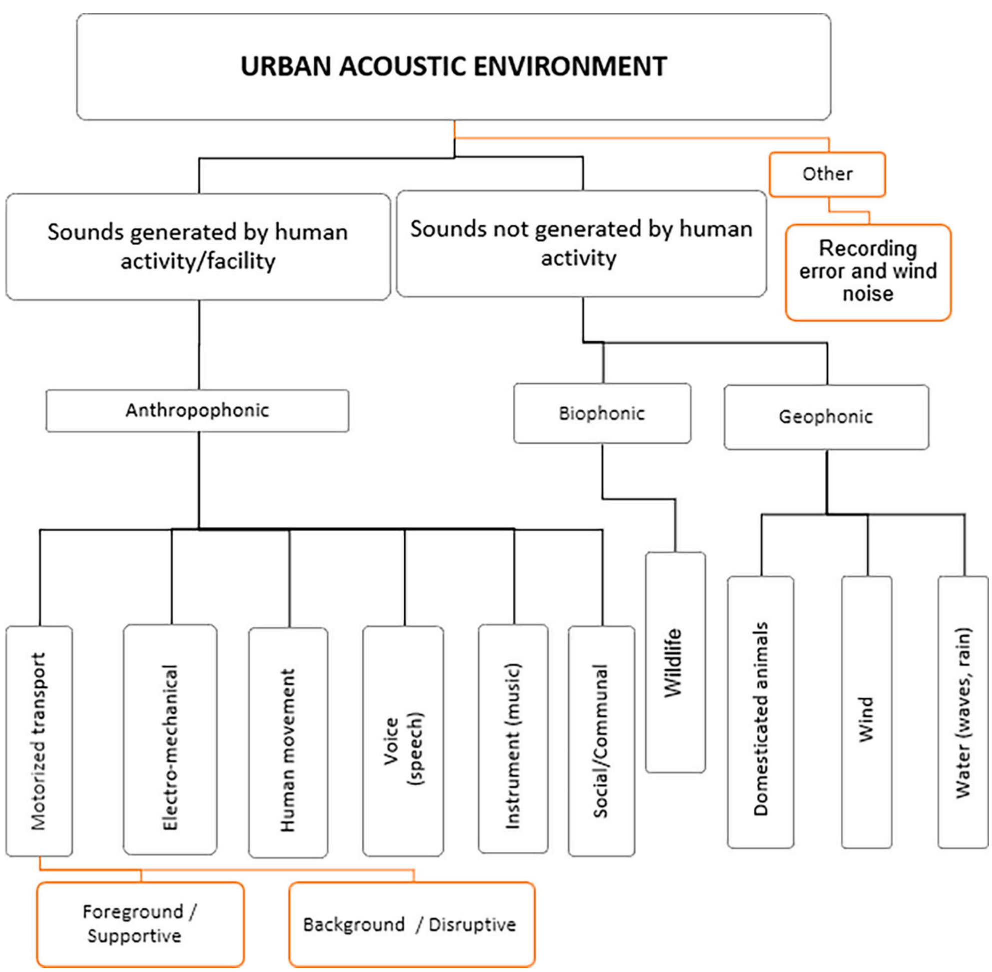


Fig. 8. Modified framework for sound source identification. Modifications to the original TS 12913–2:2018 framework highlighted with color.

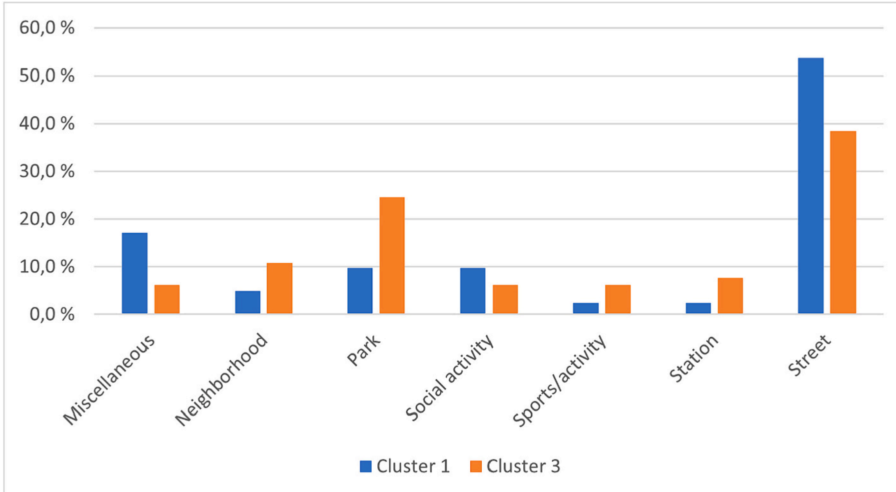


Fig. 9. Reported locations of the two main clusters show that in cluster 1 most of the recordings are from street areas and cluster 3 a mixture of street, park and neighborhood locations.

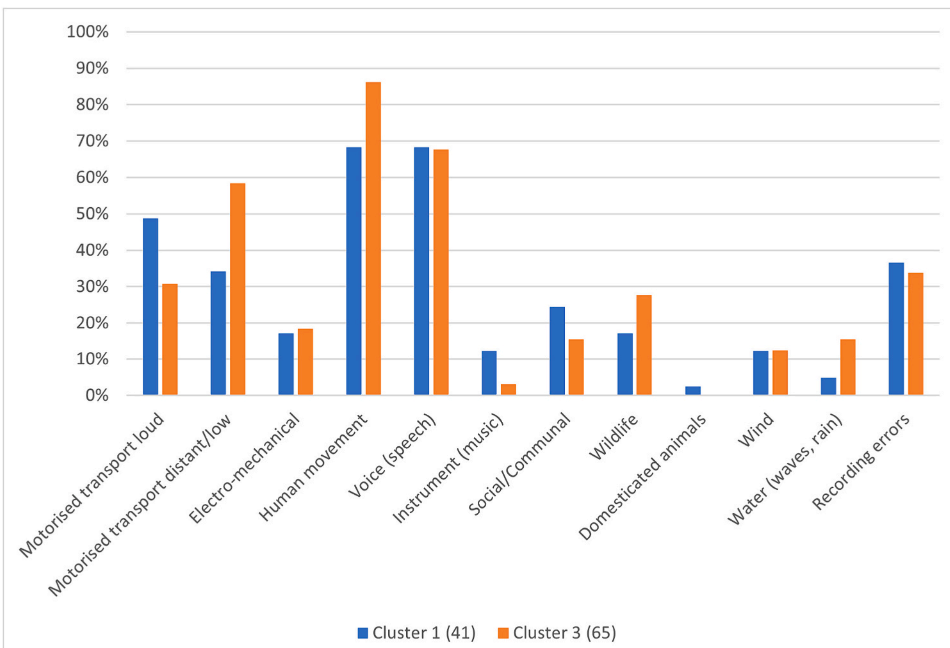


Fig. 10. Recognized sound sources of the three-cluster system's two main clusters in- dictate that Cluster 1 is more loud due to the presence of loud motorised transport. In cluster 2 motorised transport sound are more quiet or distant and therefore human and nature sound can be heard.

frequency" are both significantly higher in cluster 3. Human movement and voices were strongly present in both main clusters, but it is note-worthy that the occurrence of human movement was significantly more frequent in the less noisy cluster. The audibility of bio- phonic sounds increased in line with the distance of traffic from the recording location and other technical sounds.

We conclude that the manual analysis coincides with the results of the statistical analysis of the mostly separating feature groups. The loudness, pitch and overall frequency reflect the difference of auditory observations. The observations can be summarized as follows: cluster 1 is louder and lower frequency due to the presence of traffic and cluster 3 is quieter and has a higher frequency because of human and natural sounds. It can be said that the emerging profiles in the main clusters resemble Schafer's original main categorization: hi-fi and lo-fi. Hi-fi soundscapes are "natural sound- scapes" with a favorable signal-to-noise ratio. Urban soundscapes represent lo-fi soundscapes, where individual sounds blend into the dense mass of city noise (Schafer, 1977) (See Fig. 11). The emerging soundscape profiles found in this research could be conceptualized according to Table 4.

4. Discussion

In this research, it became obvious that labor-intensive methods, such as close listening or manual labeling, are insufficient and too time consuming for larger amounts of data. The application of machine learning methods for data analysis becomes relevant when handling masses of data. To create trustworthy and optimal training data for machine learning, data needs to be analyzed and gathered very carefully to avoid misinterpretations. Raw audio data is interesting and surprising but challenging due to its variability. According to our study, it seems

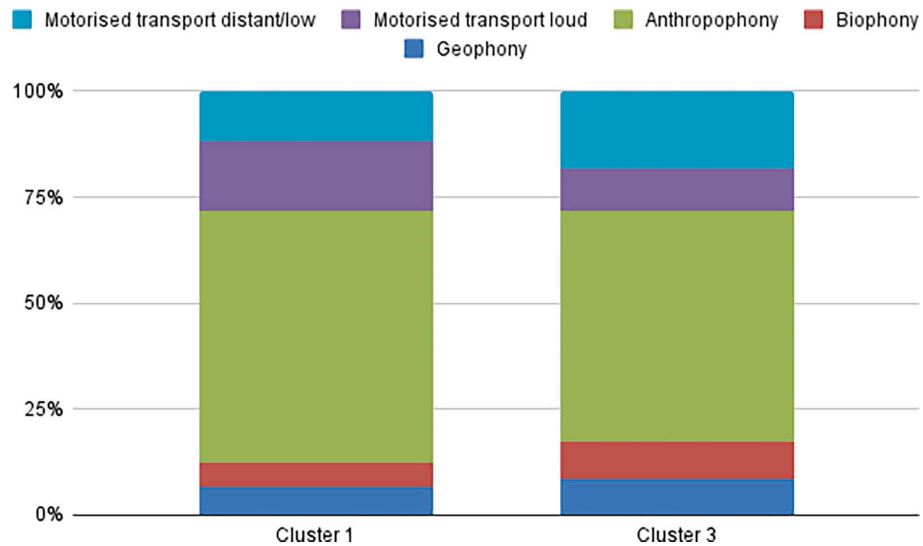


Fig. 11. Sound source profiles of the two main clusters.

Table 4
Characteristics of the two main clusters.

Cluster	Location	Loudness	Dominant category
Cluster 1: City buzz and activity	Street	Loud	Vibrant
Cluster 3: Calming havens and privacy	Park, Street	Quiet	Monotonous

that machine learning analysis separated sounds according to some kind of loudness, intensity and frequency which resulted from different distances of the recording devices from the sound sources, and especially, noise. The dataset and the clusters created with machine learning indicated that it is possible to find at least at a rough level soundscapes that are emotionally similarly labeled. In this case study the soundscapes enjoyed by young adults in Helsinki can be divided into two main categories: "places for calming down" and "places for belonging." The meaning of these phrases varies among individuals, but the data indicates that while vibrancy and social interaction make the city feel like a city, the participants also felt a need to gain distance from the city buzz.

Such findings are promising if we wish to screen large amounts of data for phenomena of interest. Labeling and data collection requires future research, and it appears that it is necessary to carefully consider the methods according to the people, soundscape and context. If the research design is precisely defined, machine learning analysis can help to find clusters from audio data which could give indications of interesting phenomena and silent signals. However, it is difficult to predict the outcome of the AI analysis and interpret the overall logic behind it.

5. Conclusion

This study reinforced the finding of our previous research that cities en- compass locations and soundscapes that researchers and professionals cannot find without crowdsensing and the help of the local people (Kaarivuo et al., 2021). Unsupervised machine learning opens possibilities to efficiently anal- yse large volumes of data that is collected with participatory methods. The interesting finding is that these methods can be applied to emotional and ex- periential data analysis, as well as for species identification or noise mapping. The end-to-end method presented in this paper opens possibilities to study soundscapes in different contexts. Due to its accessibility and efficiency, it could be applied to serve different research objectives by fine-tuning the tasks and the method.

In this research, we sought comfortable locations. From this dataset we found that surprising and commonplace locations can feel comfortable and suit the individual needs of a given citizen in a given particular moment. These locations might not be beautiful or unique, but they offer a pleasant sensory experience in everyday life situations. They might seem meaningless to designers, but they are nevertheless valuable to some citizens. The method could equally well be used to identify scary places, safe places, or places that require development, for example.

Urban environments and thereby their soundscapes are rapidly changing. To understand the context-related individual experience of a soundscape, it is necessary to broaden the framework for assessing urban soundscapes. This would also require a redefinition of balance between human society and acoustic environment. With real crowdsensing, a sufficient amount of data, and carefully developed analysis method, it would be possible to recognize emerging soundscape phenomena from cities. Mobile technology and IoT, combined with machine learning methods provide an opportunity to study large entities such as cities and even megalopolis. In constantly redevelop- ing urban areas, smart technologies would help to maintain a dialogue and understanding between stakeholders and decision makers in urban areas.

Author statement

The data collection was carried out by Mrs. Aura Kaarivuo. The analysis method and research layout were designed in collaboration with co-author Doctor Jonas Oppenländer and Professor Tommi Kärkkäinen. Dr. Oppenländer was responsible for feature selection and and Prof. Mikkonen for clustering. The main author designed and performed the manual analysis of the questionnaire data and the qualitative analysis of the clusters. The final paper was written mainly by the main author. Co-writers Dr. Oppenläder wrote the parts concerning feature selection and Prof. Kärkkäinen the machine learning and cluster identification perspectives. The paper was reviewed by the 3rd and 4th co-authors Professors Tommi Kärkkäinen and Tommi Mikkonen.

Credit authorship contribution statement

Aura Kaarivuo: Writing – review & editing, Writing – original draft, Visualization, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jonas Oppenländer:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation. **Tommi**

Kärkkäinen: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Methodology, Formal analysis. **Tommi Mikkonen:** Writing – review & editing, Writing – original draft, Supervision.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was supported by Finnish Cultural Foundation/Central Finland Regional fund, Ellen and Artturi Nyyssönen Foundation and City of Helsinki Research Grants.

Appendix A. OpenSMILE audio features

- pcm_intensity_sma (38 features): the overall intensity or volume of the audio signal
- pcm_loudness_sma (38 features): the loudness of the audio, taking into account the frequency-dependent sensitivity of human hearing
- lspFreq_sma (304 features): the spectral envelope of the audio signal using line spectral pairs (LSPs)
- pcm_zcr_sma (38 features): the rate at which the audio signal crosses the zero axis, which is related to the amount of high-frequency noise in the signal
- voiceProb_sma (38 features): the probability that the audio signal contains voiced speech (i.e., speech produced with vibration of the vocal cords).
- F0_sma (38 features): the fundamental frequency of the audio signal (F0), which is the lowest frequency component that is periodic.
- F0env_sma (38 features): a smoothed version of F0 that is intended to capture the overall pitch contour of the signal.
- mfcc_sma (456 features): Mel-frequency cepstral coefficients (MFCC), a set of features that are calculated from short-time Fourier transfor- mations of the audio signal.

References

Aletta, F., Axelsson, O., Xie, H., Zhang, Y., Lau, S. K., & Tang, S. K. (2020). *Soundscape assessment: Towards a validated translation of perceptual at- tributes in different languages. Technical report*.

Aletta, F., Kang, J., & Axelsson, O. (2016). Soundscape descrip- tors and a conceptual framework for developing predictive sound- scape models. *Landscape and Urban Planning, 149*, 65–74. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169204616000074 files/25/Aletta et al. - 2016 - Soundscape descriptors and a conceptual framework.pdf https://doi.org/10.1016/j.landurbplan.2016.02.001.

Arkette, S. (2004). Sounds like city. In *, 21*. *The- ory, culture & society* (pp. 159–168). http://journals.sagepub.com/doi/10.1177/0263276404040486. https://doi.org/10.1177/0263276404040486.

Axelsson, O., Guastavino, C., & Payne, S. R. (2019). Editorial: Soundscape assessment. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.02514

Benocci, R., Afify, A., Potenza, A., Roman, H. E., & Zambon, G. (2023). Toward the definition of a soundscape ranking index (SRI) in an Urban Park using machine learning techniques. *Sensors, 23*, 4797. https://doi.org/10.3390/s23104797

Brabham, D. C. (2013). *Crowdsourcing.* The MIT Press. http://www.jstor.org/stable/j.ctt5hhk3m.

Brambilla, G., & Pedrielli, F. (2020). Smartphone-based participatory sound- scape mapping for a more sustainable acoustic environment. *Sustain- Ability, 12*, 7899. URL: https://www.mdpi.com/2071-1050/12/19/7899 files/23/Brambilla and Pedrielli - 2020 - Smartphone-Based Participatory Soundscape Mapping.pdf https://doi.org/10.3390/su12197899.

Brown, L., Gjestland, T., & Dubois, D. (2015). Acoustic environments and soundscapes. *Soundscape and the Built Environment*, 1–16. https://doi.org/10.1201/b19145-2

Calleri, C., Astolfi, A., Pellegrino, A., Aletta, F., Shtrepi, L., Bo, E., … Orecchia, P. (2019). The effect of soundscapes and Lightscapes on the perception of safety and social presence analyzed in a laboratory experiment. *Sustainability, 11*, 3000. https://doi.org/10.3390/su11113000

Cardone, G., Foschini, L., Bellavista, P., Corradi, A., Borcea, C., Talasila, M., & Curtmola, R. (2013). Fostering participaction in smart cities: A geo- social crowdsensing platform. *IEEE Communications Magazine, 51*, 112–119. https://doi.org/10.1109/MCOM.2013.6525603

Cord, A., Ambroise, C., & Cocquerez, J. P. (2006). Feature selection in robust clustering based on Laplace mixture. *Pattern Recognition Letters, 27*, 627–635.

Craig, A., Moore, D., & Knox, D. (2017). Experience sampling: Assessing urban soundscapes using in-situ participatory methods. *Applied Acoustics, 117*, 227–235. https://doi.org/10.1016/j.apacoust.2016.05.026

Denzin, N. K. (1970). *The research act : A theoretical introduction to socio- logical methods. Methodological perspectives, Aldine, Chicago*.

Dias, F. F., Ponti, M. A., & Minghim, R. (2022). A classification and quan- tification approach to generate features in soundscape ecology using neural networks. *Neural Computing and Applications, 34*, 1923–1937. https://doi.org/10.1007/s00521-021-06501-w

Dubois, D., Guastavino, C., & Raimbault, M. (2006). Multilingual food descriptors from a sociocognitive perspective. In *View project sound in space view project A CTA A CUSTICA UNITED WITH A CUS- TICA ACognitive ApproachtoUrban soundscapes: Using verbal data to access EverydayLife AuditoryCategories. Technical report*. https://www.researchgate.net/publication/200045136.

Engel, M. S., Paas, B., Schneider, C., Pfaffenbach, C., & Fels, J. (2018). Per- ceptual studies on air quality and sound through urban walks. *Cities, 83*, 173–185. https://doi.org/10.1016/j.cities.2018.06.020

Eyben, F., Wöllmer, M., & Schuller, B. (2010). Opensmile. In *Proceedings of the 18th ACM international conference on multimedia* (pp. 1459–1462). New York, NY, USA: ACM. https://doi.org/10.1145/1873951.1874246.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Aldine de Gruyter.

Gontier, F., Aumond, P., Lagrande, M., Lavandier, C., & Petiot, J. F. (2018). Towards perceptual soundscape characterization using event detection al- gorithms. In *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2018 Workshop. DCASE2018., Surrey, UK* (p. 6).

Guastavino, C. (2007). Categorization of environmental sounds. *Canadian Journal of Experimental Psychology / Revue Cana- Dienne de Psychologie Expérimentale, 61*, 54–63. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/cjep2007006 files/40/Guastavino - 2007 - Categorization of environmental sounds.pdf https://doi.org/10.1037/cjep2007006.

Hämäläinen, J., Jauhiainen, S., & Kärkkäinen, T. (2017). Comparison of internal clustering validation indices for prototype-based clustering. *Algorithms, 10*, 105. https://doi.org/10.3390/a10030105

ISO. (2014). *ISO 12913-1:2014 soundscape part 1: Definition and conceptual framework. Technical report.* Geneva: International Organization for Stan-dardization.

ISO. (2018). *ISO/DIS 12913–2:2017 acoustics soundscape part 2: Data collection and reporting requirements. Technical report.* Geneva: Interna- tional Organization for Standardization.

ISO. (2019). *ISO/TS 12913–2:2018 acoustics soundscape part 2: Data collection and reporting requirements. Technical report.* Geneva: International Or- ganization for Standardization.

Jääskelä, P., Heilala, V., Kärkkäinen, T., & Häkkinen, P. (2021). Student agency analytics: Learning analytics as a tool for analysing student agency in higher education. *Behaviour & Information Technology, 40*, 790–808.

Jiang, L., Bristow, A., Kang, J., Aletta, F., Thomas, R., Notley, H., Thomas, A., & Nellthorp, J. (2022). Ten questions concern- ing soundscape valuation. *Building and Environment, 219*, Article 109231. https://doi.org/10.1016/j.buildenv.2022.109231

Jo, H. I., Seo, R., & Jeon, J. (2020). *Soundscape assessment methods: Compat- ibility of questionnaires and narrative interview based on ISO 12913-2*.

Kaarivuo, A., Salo, K., & Mikkonen, T. (2021). From sonic experiences to urban planning innovations. *European Planning Studies*, 1–18. https://www.tandfonline.com/doi/full/10.1080/09654313.2021.1988062. https://doi.org/10.1080/09654313.2021.1988062.

van Kamp, I., Leidelmeijer, K., Marsman, G., & de Hollander, A. (2003). Ur- ban environmental quality and human well-being. *Landscape and Urban Planning, 65*, 5–18. https://doi.org/10.1016/S0169-2046(02)00232-3

Kang, J. (2010). From understanding to designing soundscapes. *Fron- tiers of Architecture and Civil Engineering in China, 4*, 403–417. https://doi.org/10.1007/s11709-010-0091-5

Kang, J. (2023). Soundscape in city and built environment: Current de- velopments and design potentials. *City and Built Environment, 1*, 1. https://doi.org/10.1007/s44213-022-00005-6

Kang, J., & Aletta, F. (2018). The impact and outreach of soundscape research. *Environments, 5*, 58. URL: http://www.mdpi.com/2076-3298/5/5/58 files/21/Kang and Aletta - 2018 - The Impact and Outreach of Soundscape Research.pdf https://doi.org/10.3390/environments5050058.

Kaplan, D. (2023). Knee point. https://www.mathworks.com/matlabcentral/fileexchange/35094-knee-point.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47*, 583–621.

Lefevre, B., Agarwal, R., Issarny, V., & Mallet, V. (2021). Mobile crowd-sensing as a resource for contextualized urban public policies: A study using three use cases on noise and soundscape monitoring. *Cities & Health, 5*, 179–197. https://doi.org/10.1080/23748834.2019.1617656

Li, C., Liu, Y., & Haklay, M. (2018). Participatory sound- scape sensing. *Landscape and Urban Planning, 173*, 64–69. https://doi.org/10.1016/j.landurbplan.2018.02.002

Liao, P., Wan, Y., Tang, P., Wu, C., Hu, Y., & Zhang, S. (2019). Applying crowdsourcing techniques in urban planning: A bibliomet- ric analysis of research and practice prospects. *Cities, 94*, 33–43. https://doi.org/10.1016/j.cities.2019.05.024

Linja, J., Hämäläinen, J., Nieminen, P., & Kärkkäinen, T. (2023). Feature se- lection for distance-based regression: An umbrella review and a one-shot wrapper. *Neurocomputing, 518*, 344–359.

Lionello, M., Aletta, F., & Kang, J. (2020). A systematic review of prediction models for the experience of urban soundscapes. *Applied Acoustics, 170*, Article 107479. https://doi.org/10.1016/j.apacoust.2020.107479

Ministry of Social Affairs and Health and the National Institute for Health and Welfare. (2020). *Coronavirus epidemic remained stable in August*.

Mitchell, A., Aletta, F., & Kang, J. (2022). How to analyse and rep- resent quantitative soundscape data. *JASA Express Letters, 2*, 037201. URL: https://asa.scitation.org/doi/10.1121/10.0009794 files/18/Mitchell et al. - 2022 - How to analyse and represent quantitative soundsca.pdf https://doi.org/10.1121/10.0009794.

Mueller, J., Lu, H., Chirkin, A., Klein, B., & Schmitt, G. (2018). Citizen design science: A strategy for crowd-creative urban design. *Cities, 72*, 181–188. https://doi.org/10.1016/j.cities.2017.08.018

Neuvonen, A. (2019). Experiencing the soundscape with mobile mixing tools and participatory methods. *International Journal of Electronic Gover- nance, 11*(44). https://doi.org/10.1504/IJEG.2019.098811. URL: http://www.inderscience.com/link.php?id=98811 files/148/Neuvonen - 2019 - Experiencing the soundscape with mobile mixing too.pdf.

Niemelä, M., Äyrämö, S., & Kärkkäinen, T. (2021). Toolbox for distance esti- mation and cluster validation on data with missing values. *IEEE Access, 10*, 352–367.

Niessen, M., Cance, C., & Dubois, D. (2010). *Categories for soundscape: Toward a hybrid classification*.

Nieto-Mora, D., Rodríguez-Buritica, S., Rodríguez-Marín, P., Martínez- Vargaz, J., & Isaza-Narváez, C. (2023). Systematic review of machine learning methods applied to ecoacoustics and soundscape monitoring. *Heliyon, 9*, Article e20275. https://doi.org/10.1016/j.heliyon.2023.e20275

Orio, N., De Carolis, B., & Liotard, F. (2021). Locate your soundscape: Inter- acting with the acoustic environment. *Multimedia Tools and Applications, 80*, 34791–34811. https://doi.org/10.1007/s11042-021-10683-9

Potts, R. (2020). Is a new planning 3.0 paradigm emerging? Exploring the relationship between digital technologies and plan- ning theory and practice. *Planning Theory & Practice, 21*, 272–289. https://doi.org/10.1080/14649357.2020.1748699

Quinn, C. A., Burns, P., Gill, G., Baligar, S., Snyder, R. L., Salas, L., Goetz, S. J., & Clark, M. L. (2022). Soundscape classification with convolutional neural networks reveals temporal and geographic pat- terns in ecoacoustic data. *Ecological Indicators, 138*, 108831. URL: https://linkinghub.elsevier.com/retrieve/pii/S1470160X22003028 files/27/Quinn et al. - 2022 - Soundscape classification with convolutional neura.pdf https://doi.org/10.1016/j.ecolind.2022.108831.

Raimbault, M., & Dubois, D. (2005). Urban soundscapes: Ex- periences and knowledge. *Cities, 22*, 339–350. https://linkinghub.elsevier.com/retrieve/pii/S0264275105000557. https://doi.org/10.1016/j.cities.2005.05.003.

Saarela, M., Hämäläinen, J., & Kärkkäinen, T. (2017). Feature ranking of large, robust, and weighted clustering result. In *, 21. Advances in knowledge discovery and data mining: 21st Pacific-Asia conference, PAKDD 2017, Jeju, South Korea, May 23-26, 2017, proceedings, part I* (pp. 96–109). Springer.

Schafer, R. M. (1977). *The tuning of the world* (1st ed ed.). New York: A. A. Knopf.

Schuller, B., Steidl, S., & Batliner, A. (2009). The INTERSPEECH 2009 emotion challenge. In *Interspeech 2009, ISCA* (pp. 312–315). ISCA. https://doi.org/10.21437/Interspeech.2009-103.

Shao, Y., Hao, Y., Yin, Y., Meng, Y., & Xue, Z. (2022). Improving soundscape comfort in urban green spaces based on aural-visual interaction at- tributes of landscape experience. *Forests, 13*, 1262. https://doi.org/10.3390/f13081262

Song, J., Meng, Q., Kang, J., Yang, D., & Li, M. (2023). Effects of planning variables on urban traffic noise at different scales. *Sustainable Cities and Society, 100*, Article 105006. https://doi.org/10.1016/j.scs.2023.105006

Steils, N., Hanine, S., Rochdane, H., & Hamdani, S. (2021). Urban crowd- sourcing: Stakeholder selection and dynamic knowledge flows in high an low complexity projects. *Industrial Marketing Management, 94*, 164–173. https://doi.org/10.1016/j.indmarman.2021.02.011

Sun, K., De Coensel, B., Filipan, K., Aletta, F., Van Renterghem, T., De Pessemier, T., Joseph, W., & Botteldooren, D. (2019). Classification of soundscapes of urban public open spaces. *Landscape and Urban Planning, 189*, 139–155. https://doi.org/10.1016/j.landurbplan.2019.04.016

Thorndike, R. L. (1953). Who belongs in the family? *Pyschometrika, 18*, 267–276.

Wang, P., Zhang, C., Xie, H., Yang, W., & He, Y. (2022). Perception of Na- tional park soundscape and its effects on visual aesthetics. *Interna- tional Journal of Environmental Research and Public Health, 19*, 5721. https://doi.org/10.3390/ijerph19095721

Xiao, J., Lavia, L., & Kang, J. (2018). Towards an agile partici- patory urban soundscape planning framework. *Journal of En- vironmental Planning and Management, 61*, 677–698. URL: https://www.tandfonline.com/doi/full/10.1080/09640568.2017.1331843 files/45/Xiao et al. - 2018 - towards an agile participatory urban soundscape pl.pdf https://doi.org/10.1080/09640568.2017.1331843.

Yan, W., Meng, Q., Yang, D., & Li, M. (2024). Developing a theory of tranquility in urban public open spaces for future designs. *Applied Acoustics, 217*, Article 109824.

Yu, L., & Kang, J. (2009). Modeling subjective evaluation of soundscape quality in urban open spaces: An artificial neural network approach. *The Journal of the Acoustical Society of America, 126*, 1163–1174. https://doi.org/10.1121/1.3183377

Zappatore, M., Longo, A., & Bochicchio, M. A. (2017). Crowd-sensing our smart cities: A platform for noise monitoring and acoustic urban planning. *Journal of Communications Software and Systems, 13*, 53. URL: https://jcoms.fesb.unist.hr/10.24138/jcomss.v13i2.373/ files/47/Zappatore et al. - 2017 - Crowd-sensing our Smart Cities a Platform for Noi.pdf 10.24138/jcomss.v13i2.373.