From sonic experiences to urban planning innovations

Aura Kaarivuo[1[0000-0002-4514-3879]], Kari Salo[2[0000-0002-6158-1131]],
Tommi Mikkonen[3[0000-0002-8540-9918]]

[1]School of Media, Design and Conservation,
Metropolia University of Applied Sciences Helsinki, Finland

[2]School of ICT Metropolia University of Applied Sciences Espoo, Finland

[3]Department of Computer Science, University of Helsinki, Helsinki, Finland,

Word count: 6176

Abstract. It is widely accepted that personal responses to soundscapes are more dependent on listeners' emotions and attitudes, than on sounds or their physical features alone. Fast-growing cities have catalyzed the importance of designing urban spaces that citizens find pleasant and homely and that support a communal style of living.

Unfortunately, there are no standardized methods or techniques to translate sonic experiences into measurable and reliable data, which urban planning professionals or the building industry could turn into innovations and solutions. Most of the data pertaining to noise pollution and city soundscapes is still based on predictive acoustic models and rarely takes any real-life experiences or physical measurements into consideration.

This paper presents the concept of a smart and participatory approach for gathering sonic experiences that could be translated into measurable values. The aim is to search for data collection methods to provide data to train deep learning. With machine learning methods, it is possible to find patterns in both desirable and undesirable urban soundscapes. The aim of this concept is to create crowdsourced data collection methods and improve the understanding and communication between citizens and planning processes by producing more accurate and comparable experiential data.

Keywords: Sound, design, soundscape research, communicative planning, smart cities, urban planning, tool support, crowdsourcing

1 Introduction

Urbanization and fast-growing cities have catalyzed the importance of designing urban spaces that citizens find pleasant and homely, and that support a communal style of living. People are the key element of cities, and they all have an effect on the unique sonic experience of the city. It is widely accepted that a personal response to a soundscape is more dependent on the listener's emotions and attitudes than on the sounds or their physical features alone.

Unfortunately, there are no standardized methods, tools, or techniques to translate sonic experiences into measurable and reliable data, which urban planning professionals, manufacturers, and the building industry could turn into innovations and solutions. Much of the noise pollution and city soundscape data is still based on predictive acoustic models, which rarely consider any real-life experience or physical measurements.

This paper presents a smart and participatory approach for gathering and analyzing soundscape experiences and translating them into measurable values. The intention of the approach is to search for patterns in desirable and undesirable urban soundscapes. Although the gathering of verbal and written data has created a significant knowledge base about soundscapes, there is a lack of research about their design. The aim of the approach is to transform subjective experiences into objective measures. In this paper, we discuss how to improve understanding and communication among citizens, planners and building professionals.

The remainder of the paper is structured as follows: In Section 2, we present the background and motivation for this work. In Section 3, we present the problem of gathering sonic memories together with a study of how mobile technology and automated analyzing techniques could deal with this problem. In Section 4, we discuss the possibilities offered by recorded and analyzed data. In Section 5, we discuss our findings. Finally, the conclusions are presented in Section 6.

2 The Everyday Urban Soundscape Context

The urban environment is experienced through our senses. Sight has been the dominant sense in the Western Culture, but every building or space has it's characteristic sound, as well a visual shape. (Pallasmaa, 2005) The term soundscape was introduced in the 60's but the attention to it has mainly been paid during the past few decades, especially in the field of noise prevention. In soundscape research, there is still a need for standards, harmonization, documentation, etcetera, but also for experiential approach and technological innovations (Kang, 2010).

If we wish to understand the complexity of this issue, we first have to understand the key elements of the urban soundscape research:

- Acoustic environment and soundscape
- Urban soundscapes
- Soundscape design, noise prevention and quiet areas
- Tracking the everyday soundscapes

Acoustic environment and soundscape.

The acoustic environment is a combination of all sounds from multiple sources after being modified by the environment. The environment can be natural or artificial, experienced live or in memory (ISO, 2014). The term soundscape refers to the surrounding composition of sounds as perceived by humans. Although soundscape studies have emerged from various fields of science, the common research interest is the relationship between people and their sonic environment (Schafer, 1994).

Schafer classifies soundscapes as either hi-fi or lo-fi according to their signal-to-noise ratio. According to Schafer, rural soundscapes (hi-fi) have more favorable ratios due to lower ambient noise levels. Discrete sounds can be heard clearly even from a long distance, and there is less overlapping or masking of sounds. In lo-fi soundscapes, individual sounds become masked by broadband noise, and quieter sounds are not easily heard unless artificially amplified. This poor signal-to-noise ratio leads to a loss of acoustic perspective and an environment that is mainly heard in urban areas (Schafer, 1994).

Urban soundscapes

Urban soundscapes are acoustic phenomena that are often described as uncontrolled, undefinable, dislocating, complicated, complex, and multi-layered. The concept of soundscape was created to study all of its aspects and to determine how changes might affect thinking and social activities. However, rapid and ongoing changes are the main reason why we are trying to study the urban soundscape.

There are similarities in urban soundscapes and sound in general. As Schafer writes: "The world of sound is primarily one of sensation rather than reflection. It is a world of activities rather than artifacts, and whenever one writes about sound or tries to graph it, he departs from its essential reality, often in absurd ways" (Schafer, 1994). The complexity of urban soundscapes derives from the fact that every city, including all of its parts, differs from each other. The key element of cities is the people, all of whom have an effect on and a unique soundscape experience of the city. It is widely accepted that personal responses to soundscapes are more dependent on the listeners' emotions and attitudes than on the sounds or their physical features alone (Raimbault & Dubois, 2005).

Soundscape design, noise prevention and quiet areas

Although various researchers and projects have raised questions about soundscape design (Kang, 2010) and planning (Adams, et al., 2009) (Kang, et al., 2018) (Xiao, et al., 2018), it is not a standardized part of urban planning processes. The soundscape approach, which considers environmental sound as a resource, is a widely referenced and established method within which different research and measurement methods are combined to produce knowledge about how to improve urban acoustic environments (Aletta & Kang, 2018) (Jennings & Cain, 2013).

Public discussions on (and the practice of) urban soundscape planning have often been limited to two topics:

- Noise and how it could be limited and prevented

- Preserving quiet and original soundscapes ~~sonic environments~~

Environmental noise is defined in the Environmental Noise Directive (END) as ‘unwanted or harmful outdoor sound created by human activities’ (EU, 2002). Environmental noise and noise pollution are undoubtedly raising health issues in rapidly growing cities (World Health Organisation, 2018). Further, the 7th Environment Action Programme (EAP) presented an important objective that noise pollution in the EU should be decreased significantly by 2020 (European Environment Agency, 2018). In the report and subsequent monitoring reports, this objective was predicted to fail. It was unlikely that efforts to reduce noise pollution would succeed, due to transport demands, air traffic, and the number of city inhabitants increase annually. The aim of the EU noise policy was to limit harmful and polluting sounds (European Commission, 2002), and the target was to enhance the soundscape, albeit at a very approximate level.

The European Environment Agency (EEA) has recognized the need to preserve areas that are currently unaffected by noise pollution. These “quiet areas” are seen as an important component of the European soundscape, acting as balancing spaces for inhabitants who suffer from noise (European Environment Agency, 2016). In the EU’s quiet area protection plan, the approach is to ensure that citizens have access to quiet areas that are not affected by human-based noise (European Environment Agency, 2018) (European Environment Agency, 2016). As a result, enjoying a quiet and calm soundscape often means intentionally having to travel to places with this type of sonic feature. The provided definition of a quiet area is actually misleading because absolute silence can be perceived as frightening and unpleasant (European Environment Agency, 2018). Accordingly, the precise EEA definition of a quiet area is a calm and relaxing environment that is not affected by noise.

All these considerations pose important and fundamental questions pertaining to soundscape design. It is important to understand that noise is a form of pollution, and that soundscapes (and landscapes) need to be preserved and maintained. From this perspective, Schafer’s “World Soundscape project” (Schafer, 1994) reached its goal. However, the aim "to find solutions for an ecologically balanced soundscape where the relationship between the human community and its sonic environment is in harmony" remains unachievable. Moreover, if discussions about soundscapes in urban planning are limited to noise pollution and preserving quiet areas, the human experience is forgotten.

Tracking the everyday soundscapes

Places and situations we encounter on a daily basis are the interesting and rarely observed parts of urban soundscapes. The concept of something being aesthetically pleasing is traditionally interpreted as something that is exceptional, beautiful, or special and therefore gives us sensory pleasure (Naukkarinen, 2011). In crowded and busy urban areas, the soundscape normally contains several sound sources, each with a different loudness. Although these environments might not be defined as beautiful or exceptional, equally, they should not be noisy or unpleasant. Most people accept that a pleasant urban environment contains sounds of living (Aletta & Kang, 2018), which often manifests as unpredictability and overlapping sounds. However, people find some of these places cozy and comfortable, raising questions about where these places can be found and how they can be produced intentionally.

The key issue is that the quality of urban soundscapes is often understood and measured numerically (using the decibel [dB] scale). Once again, this limits the discussion to a noise–quiet dichotomy. It is important to note that noise maps are often based on predictive acoustic models and rarely take any real-life experiences or physical measurements into consideration (Gontier, et al., 2018). Moreover, the metrics do not necessarily correlate with experiential data (Jennings & Cain, 2013) (Gontier, et al., 2018). If the intention is to create an environment for people to enjoy, the primary tool for measuring acoustic environments should be what they hear (Schafer, 1994). Various pilots and projects have studied possibilities for defining, detecting, and predicting the perception of urban soundscapes (Gontier, et al., 2018) (Aletta & Kang, 2018) (Schulte-Fortkamp & Jordan, 2016). Schulte-Fortkamp et al. concluded that if we intend to understand the experiential side of a built environment, a standardized framework for physical and experiential data is required. A soundscape approach could provide this framework for combining and understanding the connection between subjective experiences and objective measures (Schulte-Fortkamp & Jordan, 2016). Further, Gontier et al. suggest that by creating perceptual parameters and sensor networks, it would be possible to employ machine learning for assessing and characterizing urban soundscapes (Gontier, et al., 2018).

In the introduction of his famous book on soundscapes, Schafer states that soundscape studies must be taken out of the laboratory and move into the living environment. Soundscape studies have always been interdisciplinary. However, the benefits of combining for example technological development, the humanities, and social sciences have only become a reality within the past two decades. It can also be stated that within this period, the relationship between science, citizenship, and technology has become increasingly important, especially when solving environmental issues (Irwin, 1995) such as improving soundscapes. If we wish to develop successful long-term solutions for sustainable environments, this cannot be accomplished without the knowledge and participation of the population (Irwin, 1995). If we wish to understand citizens' experiences of their everyday soundscapes, data collection should be conducted by the citizens themselves.

Oral or written descriptions have been the most common methods for gathering sonic experiences for soundscape research. However, in this type of citizen or crowd science, the challenge is to overcome the barriers between sonic experiences and verbalizing sounds. In essence, we cannot hear sounds in our minds correctly based on verbal descriptions. Given the development of mobile technology and mobile data over the past two decades, the recording and online sharing of sounds have become everyday functions. Further, although recording a soundscape is only a "digital copy" of the original acoustic event, it is more easily translated into objective measures compared to written data.

Crowdsourcing is a participative online activity where citizens voluntarily bring their knowledge together for individuals or an organization. (Brabham, 2013) In soundscape research, the benefit of using crowdsourcing and gathering recorded sounds instead of text is the opportunity to create training data for machine learning. Machine learning, especially deep learning, has been actively researched in recent years. It first gained success in image processing, since then it has been widely applied in the audio domain, such as in speech processing, music, and environmental sound processing (Purwins, et al., 2019). Supervised learning has typically been applied to acoustic scene classification and the detection of sound events. Acoustic scene classification describes the environment in general terms and is based on the idea that it is possible to provide a textual label as a general characterization of a particular environment. By contrast, sound event detection involves detecting individual sound events from an acoustic environment (Mesaros, et al., 2018). Unsupervised learning is typically used for clustering to reveal patterns in the data. In the audio domain, this translates into extracting features from acoustic scenes, without additional metadata, and evaluating them, such as finding common patterns (Freitag, et al., 2017).

Even after years of research, in the ongoing urbanization and changing environments, the designing process of the soundscape will most likely be ignored. If we wish to improve soundscapes, we need to identify everyday city havens, gather their experiential and characteristic data, and then translate this data into objective measures. With sufficient data and using machine learning, we believe it would be possible to predict and design more pleasant and healthy urban soundscapes.

3 Case study description and results

We arranged a test with a group of young adults in order to create an auditive and mobile methodology for gathering and sharing these experiences, and to gather more authentic data about everyday urban soundscapes.

Smartphones have rendered it possible (and easy) to record audio, and messaging services have allowed this to become an everyday function, especially among young people. By using the recording features of mobile phones and a simple online questionnaire, we wanted to provide answers to the following topics:

- How easily can a young adult record and share an audio file without instructions?
- What kind of environments would they record?
- What kind of data could be gathered from the audio files?
- How the sonic experiences could be described verbally compared to a sensory evaluation of the recordings?

3.1 Steps of the Methodological Approach

Participants. The test group comprised 36 university students from the Film and Television Department of the Metropolia University of Applied Sciences. The students were aged between 19 and 45, with the majority (58.3%) being between 18 and 25 years old (Fig. 1).

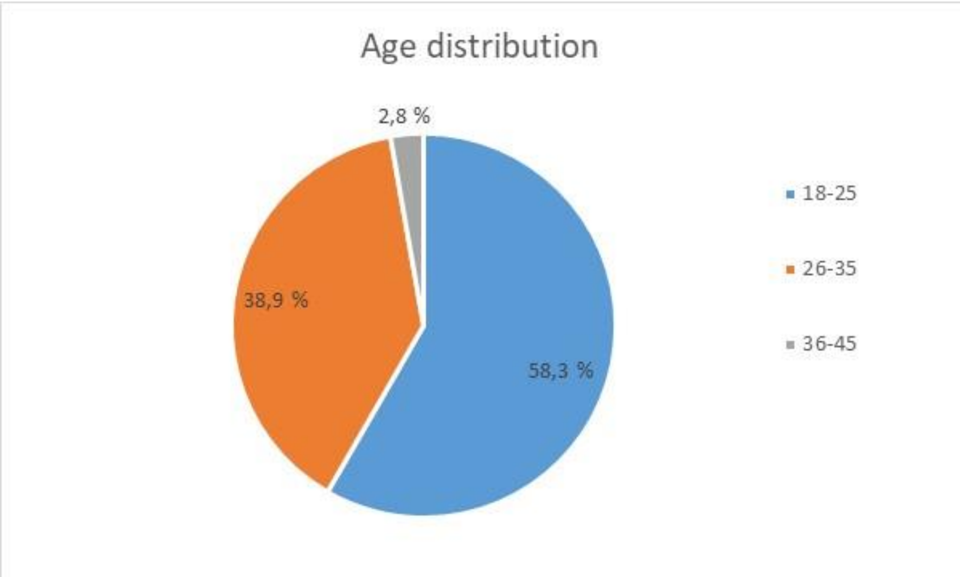

Figure 1. Age distributions of the participants.

Procedure. The participants were asked to find a place in the city (Helsinki area) that had a pleasant city soundscape and then record it with their mobile devices. They were asked to concentrate on listening to the soundscape and to answer the questions in an e-form (either subsequently or on-location). The e-form contained the following questions:

- Name of the location
- List of the sounds you heard
- What sounds would you add to the soundscape to make it more pleasant?
- What sounds would you remove or reduce?

The participants were asked to upload or share the audio file they had recorded. Finally, they were asked to write with their own words how the soundscape felt, what it sounded like, and what raised these emotions.

We did not provide any devices or detailed instructions for recording or sharing data, with the intention of simulating crowd-sourcing or participatory monitoring situations. If the intention of public participation in scientific research (or a crowd-sourcing project) is to provide a low participation threshold and amass a large amount of data, the research design should not contain training, the use of specific audio equipment, or limitations in audio file formats. Accordingly, we simply asked participants to record and share the soundscape of their choice. Further, the e-form instructed participants to send the audio file via email if sharing was complicated. The participants were given a week to accomplish the task, meaning they had time to visit their chosen location.

The test group completed the task within the allocated time, and none of them requested further instruction. Only six participants sent the audio file via email, with the remainder uploading the audio directly via the e-form. Although we did not ask how the audio was recorded, from the file names and file types it could be determined that they used mobile audio recorders, video recorders, and mobile messaging applications. Most of the audio files were m4a files, the second most popular were .mp3 files, and the remainder were either .wav or other file formats. This refers to the use of Android or iOS mobile devices.

3.2 Data analysis

We analyzed the chosen locations and written information about the soundscapes. The audio recordings were sensory evaluated and ran through a sound spectrum analyzer. By locating the recordings, it was possible to determine the places that students found the most pleasant and whether there were any similarities in the choice of locations. The city of Helsinki has mapped any areas that citizens find quiet and pleasant, which we compared to the locations chosen by the students. With the written information, the aims were to compare traditional soundscape research data collecting to digital methods and to observe any differences between them. Sensory evaluations were also conducted to compare impressions between the written descriptions and recordings.

Sound spectrum analysis is an approximate, automatic, and graphical method of translating recorded experiences into loudness levels (expressed in dB) and frequencies (expressed in Hz). This offers a visual image of differences between recordings and a visual impression of the signal-to-noise ratios of recorded soundscapes.

Locations. The participants were asked to go to a place in the city where they found the soundscape the most pleasant. The concept was that they would rely on their memories of pleasant environments and then closely listen to them. Most of the locations were in the greater Helsinki area, although two participants chose a location in another city in southern Finland (Fig. 2)

The participants chose 29 different locations in the Helsinki area. Only two locations appeared more than once in the recordings: a new shopping center in

Helsinki and a famous amusement park. The remainder were common everyday locations where people would either pass by or stop at regularly (Fig 2).

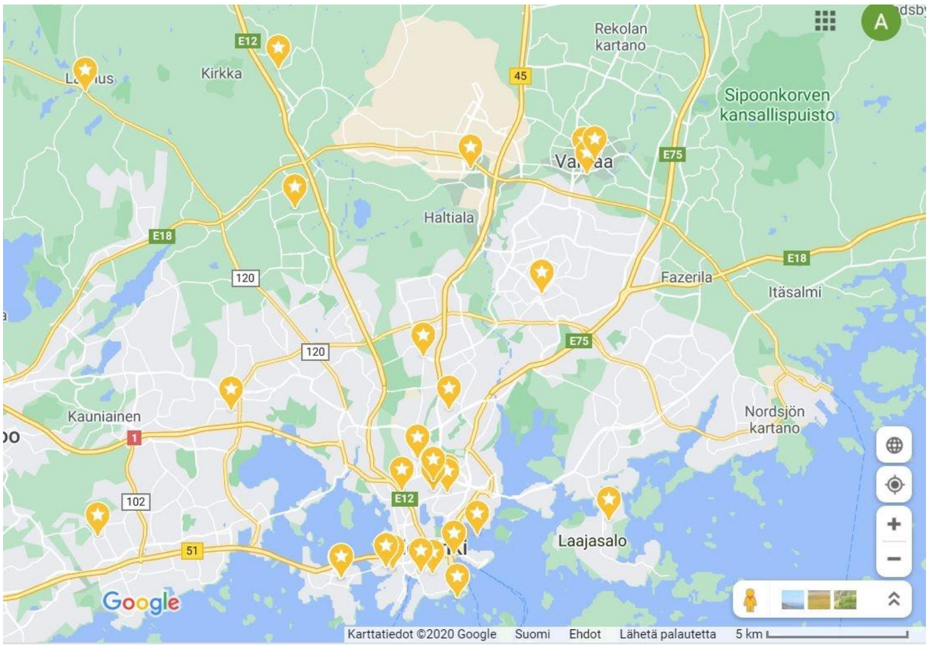

Figure 2 Recording locations in Helsinki city (mainly in the downtown area).

The majority of locations were in the downtown area of Helsinki (Fig. 3). Even though quiet areas, such as natural environments, are commonly referred to as comfortable and relaxing, most of the recording locations were in areas where the measured decibel level during the day was greater than 45–50 dB (Fig. 4). The recording locations were not seemingly in the "green areas" of the map. Conversely, they appeared to be chosen from areas that were nearby everyday routes or accommodation locations instead of outdoor or natural areas.

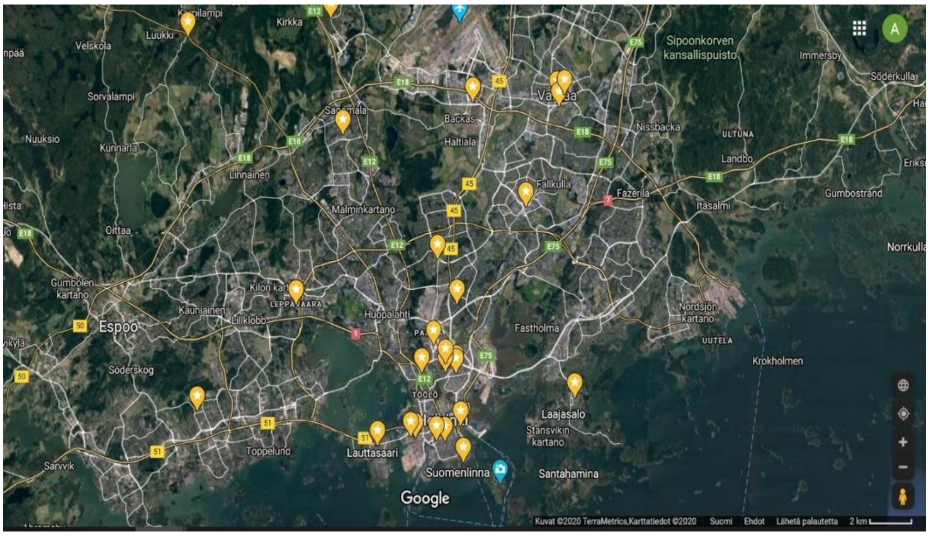

Figure 3A

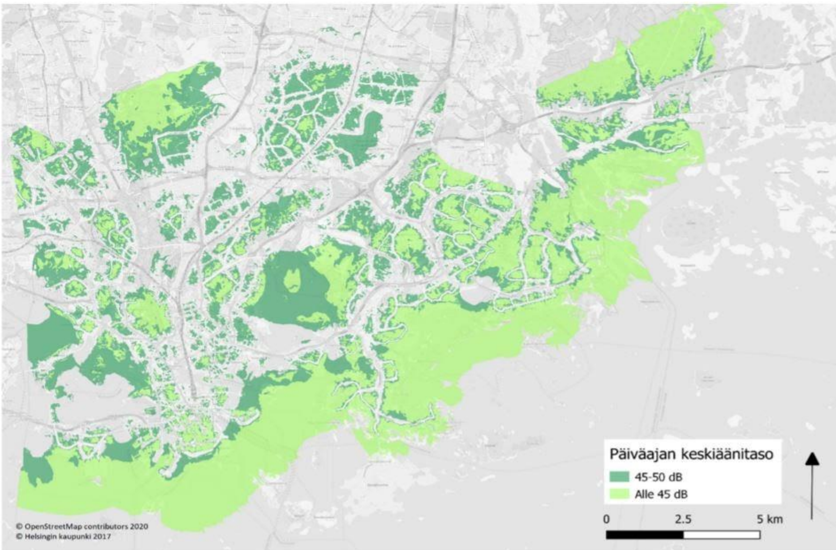

Figure 3B

Figures 3A & 3B. Map of recording locations in Helsinki area (Fig. A) compared to map of locations in Helsinki where the measured decibel level during the day was less than 45–50 dB (Fig. B) (Leppänen & Kuja-Aro, 2020).

Written information. According to the written descriptions, the locations could be divided into two main categories (Table 1):

Semi-public places	Public city locations during quiet hours
– Connected to leisure activities and social events – Presence of people – Known places and local soundmarks – Cozy atmosphere – Fun and enjoyable – Multi-sensory experience	– Low overall volume – Mild amount of sound events – Low traffic noise – Distant sounds, comparable to lo-fi soundscape – Calming and lonely

Table 1. Classification of the recording locations

The semi-public places were shopping centers, cafe terraces, amusement parks, and other locations connected with leisure activities and social events. The presence of people (probably friends and family) was a distinctive feature of these places in addition to having lively, fun, cozy, and active atmospheres. From a soundscape perspective, this manifested as a larger number of sound sources and slightly higher dB levels. None of the locations were noisy or loud, meaning the overall dB levels did not increase to harmful levels, such as at concerts, construction sites, or places with heavy traffic. Moreover, semi-public places can be connected with multi-sensory experiences such as good smells and tastes, lights, and other forms of enjoyable sensations.

In the written descriptions of the soundscapes, the most common words used to describe them were cozy, calm, relaxed, and comforting. The sounds of children, animals, and “living” were those that created the atmosphere. In more than 60% of the answers, there were references to calming down, relaxation, or taking a pause. Cars, traffic, construction sites, and the background hum of the city made the soundscape busy and restless. Although these types of sound deteriorated the soundscape quality, they were acceptable to a certain degree.

In semi-public places, the cozy and relaxed atmosphere emanated from the sounds of living, leisure, and communality. Feelings of belonging somewhere, being part of the community, and not being alone were the emotions invoked by the soundscape. Further, many sounds raised memories from home and childhood. The public city locations felt pleasant during quiet hours, because there were fewer sound sources and less happening, which was in contrast to the normal hectic environment. Water sounds, such as rain and the sea, were mentioned in six descriptions as a soothing element.

It was not possible to determine any values for sound pressure levels, frequencies, or signal-to-noise ratios from the written descriptions. Similar (or even the same) sound sources were described using various words and were mixed with sound events, emotions, and atmospheres. The content and results of the writings were very predictable, following the same main features as other studies conducted by gathering data from written descriptions. (Leppänen & Kuja-Aro, 2020)

The sensory evaluation of these sound files provided an impression of social situations and multiple simultaneous sound sources. Some locations were recognizable to locals because they contained strong sound marks from Helsinki. The presence of human sounds was the most noticeable feature.

During quiet hours, public city locations constitute street areas with no particular sound events, such as empty or quiet metro stations, busses, parks, or other outdoor spaces. The overall noise levels are low, there are few simultaneous sound events, and participants enjoyed "moments alone" simply by sitting still. The soundscapes in this category could be considered lo-fi (Schafer, 1994), even though there was the presence of a low frequency "sound of the city". Even though the soundscapes were not quiet and did not appear to be natural environments, they were described as calm and restful.

It was difficult to determine where sound files had been recorded or what was happening simply by listening. The recognizable sounds included traffic, distant or passing people, dogs, birds, engines, and machines. Even though the impression gained by participants during recording was one of being apart from any social interactions, the soundscapes themselves were not particularly attractive or pleasant.

Sound spectrum analysis represents the amount of vibration (amplitude) at each individual frequency. Usually, spectrum analyses are presented in the form of colorful graphs indicating sound levels and frequencies. Since the soundscape is a complicated mixture of different sound sources that create a combination of vibrations, spectrum analyses can present an overall graphical view of their tonalities, frequencies, and sound levels.

Reference spectrum analyses of generic city and nature recordings help to identify any expected differences between hi-fi and lo-fi soundscape recordings (Fig. 4A & 4B). A sample of a hi-fi natural soundscape with a limited number of sound sources and low overall dB levels looks more balanced and has fewer differences in the frequency bands compared to lo-fi city soundscapes (Fig. 4B). Here, low frequency areas are crowded and individual sound events cannot be observed as clearly (Fig 4B). The spectrum analysis from the city soundscape clearly reveals the noise floor (hum) that masks individual sounds (Fig. 4A).

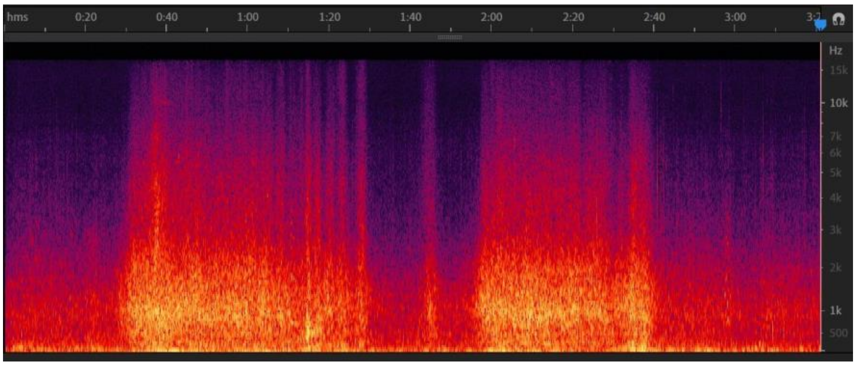

Figure 4A

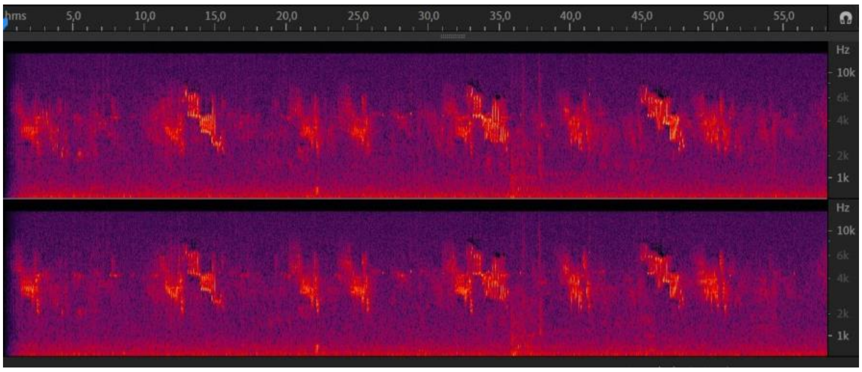

Figure 4B

Figures 4 A & B. Example spectrum analysis of a generic recording from noisy city soundscape (Fig. A) and a nature soundscape (Fig B).

The user recordings were analyzed using Adobe Audition Spectral Frequency Display. The overall visual impression was that no recordings were loud or noisy. Since all recordings were from city areas, a low background noise was present. However, only in 25% of the recordings the low frequency noise reached higher volumes at any point of the sample (Fig. 5).

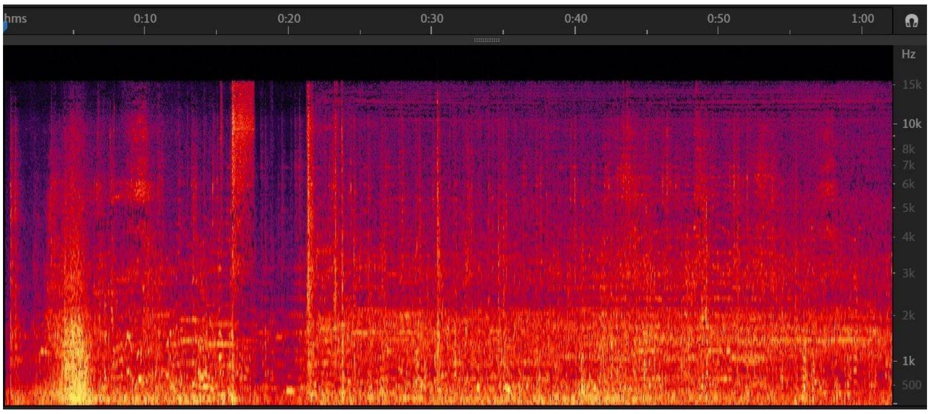

Fig. 5. Example of spectrum analysis of a student's recording from a semi-public place. The soundscape is noisier, as the image indicates high sound levels (bright yellow).

In 55% of the recordings, the analyses revealed a constant sound stream in all appearing frequencies, with no quiet moments or strong variations in dynamics (Fig. 6).

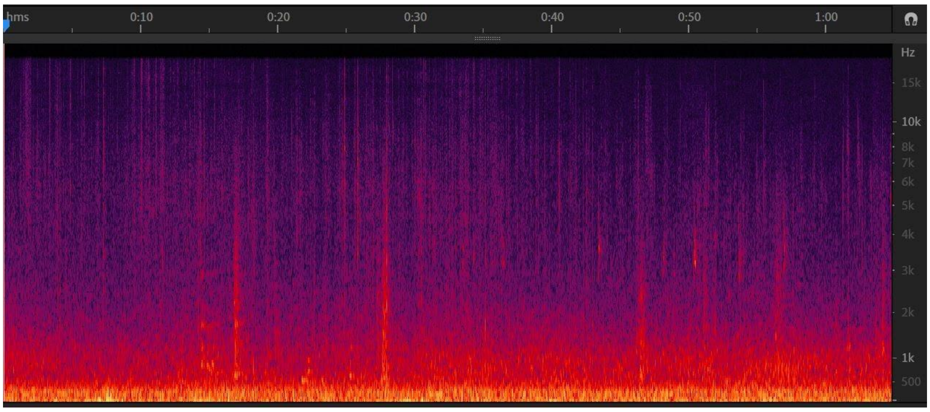

Fig. 6. Example of a spectrum analysis of a student's recording from a public city location during quiet hours. The soundscape has limited dynamics and constant background noise.

It is not possible to classify sound samples according to location or sound sources using a visual analysis. When comparing spectrum analyses from semi-public places to public city locations during quiet hours, there was only a small difference—slightly higher overall dB levels in the semi-public places. This was probably due to the presence of people creating sounds from a closer range compared to the distant sound sources in the quieter (more private) moments in other recordings.

Compared to the sample from a busy street, only 20% of the recordings reached the same loudness level. The difference was most visible in the mid and hi-mid frequency ranges, where the volume was lower. A prominent feature in most of the recordings was that traffic noise was fairly distant, reducing both noise levels and the background humming effect.

When the spectrum analyses from the students' recordings (for example, Figs. 5 and 6) were compared to the sample of a natural soundscape spectrum analysis (Fig. 4B), the recordings lacked more delicate higher frequency sound events. The soundscapes' sound events were more mechanical, and quieter sounds were either more distant or disappeared into the background noise.

3.3 Discussion

According to the test, it would appear that young adults are able to record and share audio information online without any specific technical instructions or special applications. Most of the participants (>80%) chose and uploaded audio files successfully. The few participants who failed to upload the audio file directly via the e-form were still able to share the audio via email. None of the participants asked for instructions or failed to complete the task.

The actual recordings were mostly accomplished using mobile phones according to the file formats and audio quality. All the audio files had a sufficient recording level, meaning the analyzer was able to render graphical representations. None of the recordings were so distorted or poor in quality that they could not be listened to or analyzed. There were some problems with wind and device handling noises, which is inevitable given the quality of microphones in mobile devices and the lack of professional recording skills and equipment. Despite these limitations, audio quality was generally very good.

Since the microphones in mobile devices are usually omnidirectional, they only create an approximate version of the soundscape compared to the actual listening experience. Microphones do not work like human hearing, and with omnidirectional microphones it is impossible to get close enough or to delimit (or focus) the recording on any of the surrounding sound sources. Further, it is quite difficult to avoid traffic noise or background humming in city areas, which inherently became the dominant features in most of the recordings. All the audio files were recorded in the city, with none of the participants choosing a forest (for example), which in Helsinki would have been both easy and possible.

The choice of location was one of the most interesting findings of the experiment. Although the participants were free to choose any location, they chose urban locations where city soundscape features (such as traffic noise) were prevalent. Only a few of the participants chose to record in a park as an example of a pleasant soundscape in the city, which was the closest to a nature-like environment. It is possible that the students could not be bothered to travel or to put more effort into finding pleasant places. However, this meant they revealed pleasant and easily accessible places they enjoyed in their surroundings. It is important to recognize the pleasant soundscapes in everyday lives of the citizens as opposed to the quiet nature areas.

The benefit of applying spectrum analyses is that they visually indicate any differences or regularities in soundscapes. Without sensory evaluations of written information, spectrum analyses can leave some room for interpretation; however, this is also true of recordings and written descriptions. Any physical measurement of a sound lacks any information pertaining to the experience, while any subjective verbal description lacks measurable and comparable data. The spectrum analyses provided clues about what constitutes a pleasant amount of city noise. This information, combined with the location data, suggests there are urban soundscapes within certain noise limits that are both pleasant and homely.

4 Evaluation

The recording and sharing of soundscape samples has become easy and accessible in the mobile device era. From the crowdsourcing and data collection perspective, this is promising. However, the challenge of creating training data for machine learning from soundscapes becomes difficult with insufficient, incomparable, and fragmented data. Mobile device recordings are not sufficiently accurate for measuring actual sound pressure levels in urban areas; hence, they are not suitable for evaluating noise levels for example. However, in this particular case, it is not necessary. If the aim is to understand what citizens hear and experience in a certain location, a recording provides much more accurate, shareable, and analyzable material for research compared to written descriptions or questionnaires alone. The test group appeared to find the task of finding and recording a sample from a place with a pleasant soundscape technically easy. The fact that some students have an interest in media and technologies could have benefitted them in completing the task. Recording a soundscape to produce a facsimile of the original sonic experience requires audio technical skills and equipment so the small advantage the test group had does not invalidate the result.

Sound recordings and spectrum analyses need to be analyzed automatically to create sufficient data. The recordings provide hints of tolerable noise levels and pleasant locations in addition to clues about the density and liveliness of the favorable sound environment.

With a crowdsourced mobile method, it would be possible to create training data for deep learning. Further, there are some options pertaining to how to interpret the spectrum analyses and create useful data from them:

1) Detect individual sound events from the recordings

2) Find common patterns within the recordings

As a starting point, we need to define the sound events of interest. To achieve this, we could analyze existing recordings manually and create the first version of a sound event list. Subsequently, a dataset of audio files would be required to describe those events, which is typically a time-consuming task. Luckily, there are already some limited datasets available that could be employed. For example, the DCASE datasets (Mesaros, et al., 2018), Urbansound8K (Salamon, et al., 2014), and ESC-50 (Piczak, 2015) provide a baseline, which can be completed with missing event audio files.

There are ongoing discussions about suitable approaches for audio preprocessing and deep learning methods (Mesaros, et al., 2018) (Guzhov, et al., 2020). For example, in DCASE 2016 in task 3 (sound event detection in real-life audio), teams mainly used mel-frequency cepstral coefficients for presenting the audio signals and employed deep neural networks, recurrent neural networks, or fusion for deep learning methods. In our case, we need to conduct further studies to determine which approach is most suitable. However, regardless of the chosen approach, we will obtain individual sound events from the recordings. By combining the recordings' metadata, individual sound events and their timings will improve our capabilities for analyses.

Accordingly, we will need more recordings from the urban environment to create reliable datasets for our purposes. As an initial test, we could create two datasets: one for pleasant soundscapes and another for disturbing/annoying soundscapes. Crowdsourcing recordings into these two datasets would be easier compared to our current methodology, as there is no need to collect textual data (metadata) related to recorded soundscapes. For example, by using auDeep (Freitag, et al., 2017), we could reveal features extracted from the recordings and then conduct a further analysis on the common patterns of recordings that each dataset contained. Using this approach, we could find new soundscape categories and then categorize (classify) soundscapes based on these new categories.

The aim of the methodology presented in this paper is to recognize methods for gathering sonic experiences and techniques to transform subjective experiences into objective measures. The second target is to define the parameters for a pleasant urban everyday soundscape. The first objective was reached with a very simple gathering method, demonstrating that smart technology is accessible on smartphones without any specific application or technology. Although smartphone microphones are not yet similar to calibrated measurement microphones, they provide an analyzable and audible sample of the soundscape. While a recording alone is not self-explanatory, an effective combination of data could be created with additional metadata.

5 Conclusions

It is estimated that almost 70% of the world's population will live in urban areas by 2050, and many countries will face challenges in successfully managing urban growth (United Nations, 2018). While sustainability, infrastructure, and housing are already on the agenda of urbanization management, the designing and understanding of changes in urban sonic environments are not a priority. Despite the relevance of soundscapes, there remains a need for a common framework, comparable data, and experiential knowledge.

By recognizing pleasant and unpleasant patterns in soundscapes, they could be either enhanced or reduced with planning, building, land use, or other solutions. The approach presented in this paper opens opportunities for data collection and the storing and sharing of opinions. It is clear that soundscape experiences can be collected easily with mobile devices. Files can be shared, geotagged, and tagged with metadata by using open source platforms, meaning it might not be necessary to create specific applications or systems. Instead, we need to be able to analyze the sound files automatically to use the data when planning processes and urban development. The next steps would be to implement recordings in augmented and virtual spaces. This requires methods for exchanging information with other IT systems, such as map- and 3D model-based city planning systems.

The main question for further development involves determining the things we actually seek. Noise pollution and protection have been the main topics in urban soundscape planning. However, according to this research, the concept of a pleasant soundscape should be re-defined or at least broadened. With the strict limitation of sound pressure levels under 45 dB, we might reach the conclusion that urban soundscapes are not pleasant. Moreover, as the populations in urban areas increase, citizens would need to travel increasingly further to reach locations fitting this criterium. However, it would appear there are small, pleasant soundscapes, and sometimes surprising havens in urban areas that might not resemble obvious natural environments with their low dB levels and singing birds. Sensory pleasure and relaxation are important elements when creating sustainable and equal living environments. Further, this option should be available for all citizens, especially those who cannot travel to quiet, natural locations easily.

In this paper, we have suggested that by creating accessible and simple data gathering methods, and by recognizing the pleasant environments in our everyday urban lives, it would be possible to utilize machine learning techniques to create frameworks for urban soundscape planning. On a larger scale, with this kind of data, it would be possible to predict problematic sound components from urban soundscapes. This data could also be used to create technologies and innovations to secure soundscapes for vulnerable areas such as playgrounds, schoolyards, parks, and living areas. The importance of good, healthy urban soundscapes is clear. With smart data collection, we can start to design and build these places instead of simply enhancing and repairing the existing uncontrolled and unwanted urban soundscapes.

References

Adams, M., Davies, B. & Neil, B., 2009. *Soundscapes: an urban planning process map.* Institute of Noise Control Engineering, pp. 4401–4410.

Aletta, F. & Kang, J., 2018. Towards an Urban Vibrancy Model: A Soundscape Approach. *International Journal of Environmental Research and Public Health,* 15(8).

Brabham, D. C., 2013. *Crowdsourcing.* Pittsburgh: The MIT Press.

EU, 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 *Official Journal L, 189*, 18.7.2002

European Commission, 2002. Directive 2002/49/EC of the European Parliament and of the Council. *Official Journal L, 187*, 189, pp. 0012–0026.

European Environment Agency, 2016. *Quiet areas in Europe—The environment unaffected by noise pollution*. EEA Report, Issue 14.

European Environment Agency, 2018. *Environmental indicator report 2018 – In support to the monitoring of the 7th Environment Action Programme*. EEA report, Issue 19.

European Union (EU), 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. *Official Journal L, 189*, 18 July, pp. 0012–0026.

Freitag, M. et al., 2017. AuDeep: unsupervised learning of representations from audio with deep recurrent neural networks. J. Mach. Learn. Res. 18, 1 (January 2017), 6340–6344. *Journal of Machine Learning Research (JMLR),* 18(1), pp. 6340–6344.

Gontier, F. et al., 2018. *Towards perceptual soundscape characterization using event detection algorithms.*

Guastavino, C., 2007. Categorization of Environmental Sounds. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale,* 61(1), pp. 54–63.

Guzhov, A., Raue, F., Hees, J. & Dengel, A., 2020. *ESResNet: Environmental Sound Classification Based on Visual Domain Models.* [Online]
Available at: https://arxiv.org/pdf/2004.07301.pdf
[Accessed 11 2020].

Irwin, A., 1995. *Citizen Science: A Study of People, Expertise and Sustainable Development.* London and New York: Routledge.

ISO, 2014. *ISO 12913 Acoustics—Soundscape.*

Jennings, P. A. & Cain, R., 2013. A framework for improving urban soundscapes. *Applied Acoustics*, *74*(2), pp. 293–299.

Kang, J., 2010. From understanding to designing soundscapes. *Frontiers of Architecture and Civil Engineering in China,* 4(4), pp. 403–417.

Kang, J. B., Aletta, F., Margaritis, E. C. & Yang, M. L., 2018. A model for implementing soundscape maps in smart cities. *Noise Mapping,* 5, pp. 46–59.

Leppänen, A. & Kuja-Aro, J., 2020. Hiljaisiksi ja rauhallisiksi koetut alueet Helsingissä. *Kaupunkiympäristön julkaisuja,* 16,

Lercher, P. & Schulte-Fortkamp, B., 2003. *The relevance of soundscape research to the assessment of noise annoyance at the community level.* Berlin, Foundation ICBEN, pp. 225–231.

Machler, L. & Milz, D., 2015. The evolution of communicative planning theory. *AESOP Young Academics Booklet Series B—Conversations in Planning Theory*, October.

Mesaros, A. et al., 2018. Detection and Classification of Acoustic Scenes and Events: Outcome of the DCASE 2016 Challenge. *26*(2), pp. 379–393.

Naukkarinen, O., 2011. Arjen estetiikka. *Aalto-yliopiston julkaisusarja TAIDE+MUOTOILU+ARKKITEHTUURI*, Issue 1.

Pallasmaa, J., 2005. The Eyes of the Skin: Architecture and the Senses. Chichester: Wiley.

Piczak, K. J. Ed., 2015. *Esc: Dataset for environmental sound classification.* New York, NY, Association for Computing Machinery, pp. 1015–1018.

Purwins, H. et al., 2019. Deep Learning for Audio Signal Processing. *IEEE Journal of Selected Topics in Signal Processing, 13*(2), pp. 206–219.

Raimbault, M. & Dubois, D., 2005. Urban soundscapes: Experiences and knowledge. *Cities, 22*(5), pp. 339–350.

Salamon, J., Jacoby, C. & Bello, J. P., 2014. *A dataset and taxonomy for urban sound research.* New York, NY, Association for Computing Machinery, pp. 1041–

1044.

Schafer, R. M., 1994. *The Soundscape—Our Sonic Environment and the Tuning of the World.* Vermont: Destiny Books.

Schulte-Fortkamp, B. & Jordan, P., 2016. When soundscape meets architecture. *Noise Mapping,* 3(1).

United Nations, 2018. *News: 68% of the world population projected to live in urban areas by 2050, says UN.* [Online]
Available at: https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html
[Accessed 5 12 2020].

World Health Organisation, 2018. *Environmental Noise Guidelines for the European Region,* s.l.: World Health Organization.

Xiao, J., Lavia, L. & Kang, J., 2018. Towards an agile participatory urban soundscape planning framework. *Journal of Environmental Planning and Management, 61*(4), pp. 677–698.

Zappatore, M., Longo, A. & Bochicchio, M., 2017. Crowd-sensing our Smart Cities: A Platform for Noise Monitoring and Acoustic Urban Planning. *Journal of Communications Software and Systems, 13*(2), pp. 53–67.