PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Classifying urban public spaces according to their soundscape

Kang SUN[1]; Karlo FILIPAN[2]; Francesco ALETTA[3]; Timothy VAN RENTERGHEM[1];
Toon DE PESSEMIER[1]; Wout JOSEPH[1]; Dick BOTTELDOOREN[1]; Bert DE COENSEL[1,2,*]

[1] WAVES Research Group, Department of Information Technology, Ghent University / imec, Belgium

[2] ASAsense CVBA, Bruges, Belgium

[3] University College London, UK

ABSTRACT

Cities are composed of many types of outdoor spaces, each with their distinct soundscape. Some of these soundscapes can be extraordinary, others are often less memorable. However, most locations in a city are not visited with the purpose of experiencing the soundscape. Consequently, the soundscape will not necessarily attract attention. Existing methods based on the circumplex model of affect classify soundscapes according to the pleasure and arousal they evoke, but do not fully take into account the goals and expectations of the listener. Therefore, in earlier work, a top-level hierarchical classification method was developed, which distinguishes between spaces based on the degree to which the soundscape creates awareness of the acoustical environment, matches expectations and arouses the listener. This paper presents the results of an immersive laboratory experiment, designed to validate this classification method. The experiment involved 40 participants and 50 audiovisual recordings drawn from the Urban Soundscapes of the World database. It is shown that the proposed classification method results in clearly distinct classes, and that membership to these classes can be explained well by physical parameters, extracted from the acoustical environment as well as the visual scene.

Keywords: Soundscape, Classification, Urban space

1. INTRODUCTION

The soundscape of urban outdoor public spaces contributes to the perceived quality of those spaces, and eventually to the identity of a city or a community. Over the past decades, awareness of the transience and the heritage value of these soundscapes has risen, and more and more efforts are spent to record and catalog urban soundscapes for posterity. A well-known early example is the World Soundscape Project initiated by R. Murray Schafer (1), which has led to a set of recordings called the World Soundscape Library. Technological progress is providing a helping hand in speeding up this process. More recent examples are the use of mobile applications and citizen science to collect urban soundcape recordings (2, 3), or several projects that aim to collect high-quality immersive audiovisual recordings at urban locations worldwide, such as the Urban Soundscapes of the World project (4). In this light, soundscape data collection methods have become a subject of standardization efforts within the ISO 12913 series of standards (5).

Data collection often goes hand in hand with data classification, and this is also the case for urban soundscapes. Several approaches to classify urban spaces according to their soundscape have been proposed in the past. These classification methods can be based either solely on acoustical properties, such as in (6), or on a combination of acoustical, visual and other perceptual properties, such as in (7). Holistic methods, which assess the soundscape as a whole, are often inspired by the circumplex model of affect (8), and classify soundscapes according to the pleasure and arousal they evoke (9). However, the application of the core affect model assumes that one actually pays attention to the soundscape. While this might be the case by design in most laboratory experiments, most often, city dwellers do not have the explicit purpose of experiencing the soundscape. Moreover, the acoustic environment in many urban locations does not necessarily attract attention if sound is not the purpose of being there.

[*] bert.decoensel@ugent.be

Taking into account the goals and expectations of the listener may lead to a more nuanced classification. In (4), a hierarchical approach is proposed, to categorize urban locations concerning how the soundscape at these locations is typically perceived. In a first stage, locations are classified according to the degree to which the soundscape influences the perception of the total environment. The scale varies from no influence at all (meaning that one does not pay attention to the soundscape) to a strong influence (meaning that the soundscape immediately attracts attention and creates awareness). In a second stage, locations are scaled according to how well the soundscape matches expectations. These expectations can be both place-related (e.g. congruence with the visual surroundings) and person-related (e.g. how much the soundscape interferes with the goals and activities of the listener). The scale ranges from disruptive to supportive, and relates to the pleasantness axis in the core affect model (9). Finally, in a third stage, locations are scaled according to how they arouse the listener. The scale ranges from calming to stimulating, and relates to the arousal axis in the core affect model.

In (10), an operational method (in the form of a small questionnaire) is presented, which allows assessing audiovisual recordings according to the hierarchical classification approach proposed above. In Sections 2 and 3, this paper briefly presents the results of an immersive laboratory experiment, designed to validate this operational method. Results of this experiment have been published in extended form in (11). In Section 4, a discussion of these results is presented.

2. METHODOLOGY

2.1 Audiovisual stimuli

The audiovisual stimuli used in the laboratory experiment are drawn from the Urban Soundscapes of the World database (4). This database consists of immersive audiovisual recordings, i.e. combined, simultaneous spatial audio, and 360-degree video, collected in a range of cities worldwide. In each city, recording locations were selected in a systematic manner, through an online survey among 30 to 50 inhabitants (depending on the city). Participants in this survey were asked to pinpoint those outdoor public spaces within their city that they perceive as either being full of life and exciting, chaotic and restless, calm and tranquil, or lifeless and boring. These adjectives were taken from the four quadrants of the principal component analysis (PCA) performed in (9). A spatial clustering analysis allowed compiling a shortlist of prototypical locations with a variety of soundscapes, more or less uniformly covering each of the four quadrants in the two-dimensional core affect perceptual space. More information about the site selection protocol can be found in (4). For the present experiment, 50 one-minute stimuli (including first-order ambisonics spatial audio track) were extracted from the database, recorded at different locations within the cities of Montreal, Boston, Tianjin, Hong Kong and Berlin. We refer to (11) for a complete list of all stimuli.

From the recordings, a range of acoustical and non-acoustical indicators was calculated. Acoustical indicators included the one-minute L_{Aeq} of the stimuli, which ranged from 53.3 dB(A) (recorded in Tiergarten, Berlin) to 77.0 dB(A) (recorded in Peking Road, Hong Kong), as well as percentile levels, loudness, sharpness and auditory saliency (12). Non-acoustical indicators included the people density (qualitatively labeled on a 5-point scale) and the amount of green elements (by proxy of the percentage of green pixels) in the 360-degree scene. Figure 1 illustrates the method for identifying green elements, on the basis of the image extracted from the opening scene of the 360-degree video.

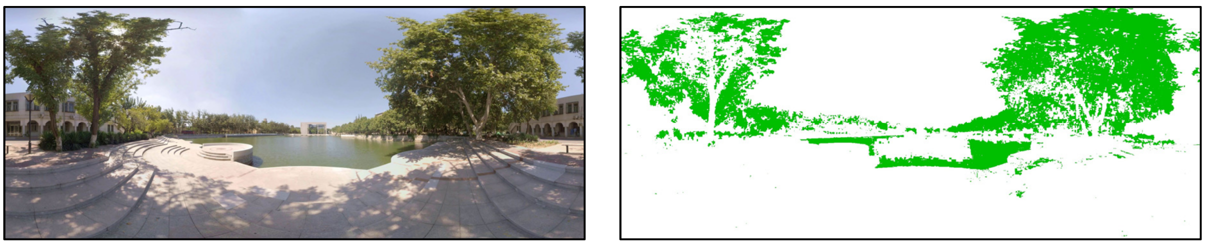

Figure 1 – Illustration of the calculation of green coverage; this example contains 20.6% green elements.

Considering all 50 recordings, 22% had no people at all in them, 30% had a small number of people, 26% an average number, 14% a high number, and 8% a very high number. The percentage of green elements ranged from 0% to about 60%. Overall only a very small negative correlation (r^2 = 0.104) between L_{Aeq} and amount of green elements is observed.

2.2 Experiment setup

During the experiment, participants were seated inside a soundproof booth, with the experimenter monitoring the experiment from outside. Audiovisual stimuli were played back using a PC placed outside the booth. The 360-degree video was presented through an Oculus Rift head-mounted display; participants could freely move their head and look around. The audio was played back through a pair of Sennheiser HD650 headphones, driven by a HEAD acoustics LabP2 headphone amplifier calibrated to ensure that the audio is played back at the original sound level. To have a smooth run of the experiment, questions were projected within the head-mounted display, and participants were asked to answer verbally; the experimenter would then mark down the answers. This way, the participants did not need to take the head-mounted display off all the time during the experiment.

2.3 Participants

In total, 40 participants (29 male, 11 female, mean age 29.5 yr, range 22-46 yr) took part in the experiment, recruited mostly among students at Ghent University. All had normal hearing, assessed via pure tone audiometry, and normal color vision, tested by the Ishihara test for color deficiency (13). Participants were offered a gift voucher as compensation for taking part in the experiment.

2.4 Experiment outline

The participants performed the experiment individually. At the start of the experiment, each participant was briefly informed about the experimental procedure. Subsequently, the 360-degree videos with first-order ambisonics audio track were presented in random order. To keep the total duration of the experiment within bounds, participants were split into two groups of 20 people, each being presented with only half of the stimuli. After each audiovisual stimulus, participants were asked two sets of questions. In the *first set of questions*, participants had to rate the locations they had experienced, on an 11-point scale in terms of the adjectives used for selecting the recording locations, as mentioned in Section 2.1. The *second set of questions* consisted of those proposed in (10) for the hierarchical classification approach, and are listed in Table 1. Note that either question 5a or 5b is asked, depending on the answer on question 1. At the end of the experiment, a small questionnaire was administered, containing questions of demographic nature.

Table 1 – List of questions and possible answers for the hierarchical soundscape classification method.

1. In general, how would you categorize the environment you just experienced? [5-point scale from "calming/tranquil" to "lively/active"]
2. In general, what kind of activities could you imagine doing in this environment? [list of activities]
3. How much did the sound draw your attention? [5-point scale from "not at all" to "extremely"]
4. Would the sound environment prevent you from doing the activities mentioned above? [5-point scale from "not at all" to "extremely"]
5a. How much does the sound environment contribute to the *calmness/tranquility* of this place? [5-point scale from "not at all" to "extremely"]
5b. How much does the sound environment contribute to the *liveliness/activeness* of this place? [5-point scale from "not at all" to "extremely"]

3. RESULTS

3.1 Principal component analysis

The first set of questions leads to average scores for each of the 50 soundscapes considered, on the four core affect model classes "full of life/exciting", "chaotic/restless", "calm/tranquil", and "lifeless/boring". In order to obtain a similar score for the four classes "backgrounded", "disruptive", "calming" and "stimulating" that are considered in our proposed hierarchical classification method, a fuzzy set membership analysis is applied to the answers on the second set of questions. Answering "not at all" or "a little" on Question 3 (Table 1) leads to a score of 1.0 for the "backgrounded" class. The other possible answers on the 5-point scale gradually lead to lower scores; the answer "extremely" leads to 0.0. Answering "highly" or "extremely" on Question 4 leads to a score of 1.0 for the "disruptive" class, and in a similar way scores for "calming" and "stimulating" are obtained from the answers on questions 5a and 5b. This procedure is applied to each soundscape/participant combination, and for each soundscape, the average membership over all participants for each class is calculated.

Subsequently, two principal component analyses (PCA) are performed on the average scores: once based on the scores for the four core affect model classes, and once based on the scores for the four classes in the alternative hierarchical classification method. Figure 2 shows scatterplots of all soundscapes along the two principal components, for each of the two PCA analyses.

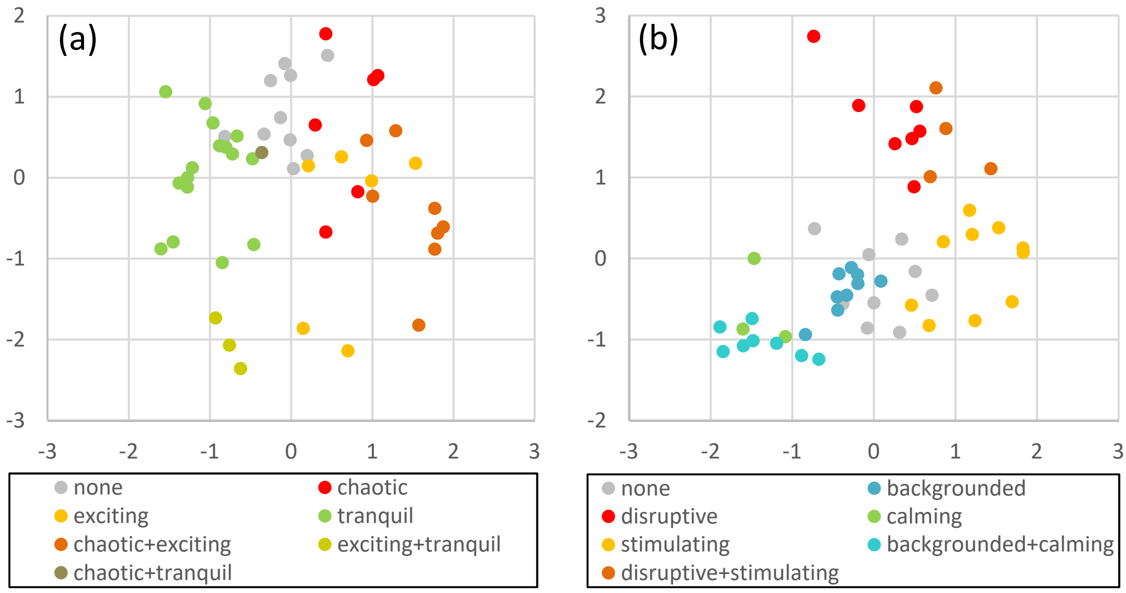

Figure 2 – PCA component plots for (a) the core affect model, and (b) the hierarchical classification method.

For the PCA based on the core affect model scores (Figure 2a), component 1 explains 55.1% of variance, while component 2 explains 30.9%. For the PCA based on the hierarchical classification method scores (Figure 2b), component 1 explains 71.1% of variance, while component 2 explains 22.1%. Thus, a slightly higher total variance is explained by the hierarchical classification method. Individual soundscapes in Figure 2 are colored according to the class with the highest score, or labeled "none" if there is no clear class that has the highest score for the given soundscape.

3.2 Models based on acoustical and visual factors

To analyze the contribution of underlying acoustical and visual factors to the classification of each location, a generalized linear mixed model (GLMM) is constructed for the four proposed classes, using a stepwise procedure, with participant as random factor. Table 2 shows the fittest model for each of the classes of soundscapes, based on the Akaike Information Criterion (AIC). Acoustical factors include sound level (L), loudness (N), sharpness (S) and saliency (SL), together with associated extreme and percentile levels. Visual factors include greenness (G) and person density (P), as a categorical value, dummy coded where P_1 means no people at all, up to P_4, meaning a high amount of people.

Table 2 – Generalized linear mixed models for the classes of the hierarchical classification method.

Soundscape class	AIC	Membership
Backgrounded	319.2	$0.458 - 0.041 \cdot L_{A05} + 0.023 \cdot N_{05} - 0.068 \cdot S_{max} - 0.037 \cdot SL_{50} - 0.116 \cdot G$
Disruptive	511.1	$-1.432 - 0.525 \cdot L_{A95} + 0.547 \cdot L_{A90} - 0.035 \cdot SL_{95} - 0.480 \cdot S_{50} + 0.040 \cdot N_{05} - 0.046 \cdot N + 0.302 \cdot S_{95} + 0.145 \cdot S_{05}$
Calming	591.1	$1.327 - 0.020 \cdot L_{AFmax} + 0.172 \cdot P_1 + 0.024 \cdot P_2 + 0.003 \cdot P_3 - 0.057 \cdot P_4 + 0.106 \cdot S_{50}$
Stimulating	535.7	$0.755 - 0.196 \cdot P_1 - 0.077 \cdot P_2 - 0.064 \cdot P_3 + 0.091 \cdot P_4 + 0.067 \cdot SL_{50}$

3.3 Models solely based on acoustical factors

In order to predict the membership of a given soundscape to each of the four classes within the hierarchical classification approach, a set of linear regression models was constructed, exclusively based on the acoustical parameters extracted from the spatial audio recordings. Table 3 shows the resulting linear regression models.

Table 3 – Linear regression models for the classes of the hierarchical classification method.

Soundscape class	Membership	r^2	sig.
Backgrounded	$-0.018 \cdot L_{A05} + 1.464$	0.521	0.000
Disruptive	$0.027 \cdot L_{A05} - 0.015 \cdot L_{A95} - 0.733$	0.488	0.006
Calming	$-0.020 \cdot L_{AFmax} + 0.079 \cdot S_{50} + 1.440$	0.426	0.098
Stimulating	$0.078 \cdot SL_{95} + 0.643$	0.501	0.000

4. DISCUSSION

The GLMM model results presented in Table 2 show that visible green elements reduce the chance for a soundscape to become labelled as backgrounded. Moreover, the model for the backgrounded class has the lowest AIC value. This suggests that, based on acoustical and visual factors, it would be relatively easy to predict if a soundscape will draw attention or not. The model for predicting if a soundscape will be backgrounded solely based on acoustical factors, as shown in Table 3, only retains L_{A05} as an acoustical indicator. Thus, to be backgrounded, soundscapes should simply not contain any loud sounds, whatever their origin or duration.

A disruptive soundscape prevents the users of a space from performing the activities that they would otherwise engage in. The optimal generalized linear mixed model for the disruptive class combines many non-orthogonal acoustical factors but contains no visual factors, which might indicate that sound dominates perception in a disruptive soundscape. The predictive model for the disruptive class in Table 3 contains both L_{A05} and L_{A05}–L_{A95} with a positive coefficient, indicating that the sound level as well as the temporal variability of the sound are determinant for the soundscape to become disruptive.

Supportive soundscapes, either calming or stimulating, are expected to contribute to the overall experience of a place, matching expectations created by the context and purpose of the place. The results of Table 2 indicate that calming or stimulating support is for a large part evoked by visual information. Although the amount of visual vegetation is not a significant factor for explaining why soundscapes are calming or stimulating, however, the visual presence of people plays a key role: too many people present reduces the calmness of the soundscape. Furthermore, sharpness (S_{50}) and the absence of strong peaks in the sound level (L_{AFmax}) appear in both the GLMM and linear regression models for the calming soundscapes. Sharpness is typically higher for natural sounds (which are commonly perceived as calming) and lower for mechanical sounds (which are commonly perceived as non-calming). As expected, the number of people is found to be positively correlated with the stimulating character of a soundscape. Finally, auditory saliency (SL), a property of the sound that characterizes its ability to draw attention, and that focuses strongly on vocalizations (12), appears in both the explorative GLMM and predictive models for the stimulating soundscapes. This suggests that (bottom-up) auditory saliency might be indicative in explaining the stimulating character of a soundscape, whereas voluntary (top-down) auditory attention might be indicative in explaining the calming or tranquil character of a soundscape (14).

5. CONCLUSIONS

This paper presented the results of a laboratory experiment, designed to validate a hierarchical approach for classifying urban outdoor public spaces according to their soundscape. This approach, for which an operational assessment method was proposed in earlier work, is seen as an alternative to the more conventional way of classifying soundscapes based on the circumplex model of core affect. The experiment involved the perceptual evaluation of immersive audiovisual recordings drawn from the Urban Soundscapes of the World database. The proposed hierarchical classification method was shown to result in distinct classes, and membership to these classes could be explained well by parameters extracted from the acoustical environment (spatial audio) as well as the visual scene.

ACKNOWLEDGEMENTS

This study was supported by the HEAD Genuit Foundation through grant agreement No. P-16/11-W, "Urban Soundscapes of the World", and was partially carried out in the framework of the C3PLACES project, which received funding from the European Union's Horizon 2020 research innovation programme under grant agreement No. 693443.

REFERENCES

1. Schafer RM. *The Soundscape: our Sonic Environment and the Tuning of the World*. Rochester, Vermont: Destiny Books (1994).
2. Radicchi A. The use of mobile applications in soundscape research: open questions in standardization. In *Proceedings of Euronoise '18*, Crete, Greece (2018).
3. Li C, Liu Y, Haklay M. Participatory soundscape sensing. *Landscape and Urban Planning* **173**:64-69 (2018).
4. De Coensel B, Sun K, Botteldooren D. Urban Soundscapes of the World: selection and reproduction of urban acoustic environments with soundscape in mind. In *Proceedings of Internoise '17*, Hong Kong, (2017).
5. ISO/TS 12913-2. *Acoustics – Soundscape – Part 2: Data collection and reporting requirements*. International Organization for Standardization, Geneva, Switzerland (2018).
6. Rychtáriková M, Vermeir G. Soundscape categorization on the basis of objective acoustical parameters. *Applied Acoustics* **74**(2):240-247 (2013).
7. Jeon JY, Hong JY. Classification of urban park soundscapes through perceptions of the acoustical environments. *Landscape and Urban Planning* **141**:100-111 (2015).
8. Russel JA. A circumplex model of affect. *Journal of Personality and Social Psychology* **39**:1161-1178 (1980).
9. Axelsson Ö, Nilsson ME, Berglund B. A principal components model of soundscape perception. *Journal of the Acoustical Society of America* **128**(5):2836-2846 (2010).
10. Sun K, Botteldooren D, De Coensel B. Realism and immersion in the reproduction of audio-visual recordings for urban soundscape evaluation. In *Proceedings of Internoise '18*, Chicago, IL, USA (2018).
11. Sun K, De Coensel B, Filipan K, Aletta F, Van Renterghem T, De Pessemier T, Joseph W, Botteldooren D. Classification of soundscapes of urban public open spaces. *Landscape and Urban Planning* **189**:139-155 (2019).
12. Filipan K, De Coensel B, Aumond P, Can A, Lavandier C, Botteldooren D. Auditory sensory saliency as a better predictor of change than sound amplitude in pleasantness assessment of reproduced urban soundscapes. *Building and Environment* **148**:730-741 (2019).
13. Ishihara S. *Test for Colour Deficiency – 24 Plates Edition*. Tokyo: Kanehara Shuppan (1957).
14. Filipan K, Boes M, De Coensel B, Lavandier C, Delaitre P, Domitrović H, Botteldooren D. The personal viewpoint on the meaning of tranquility affects the appraisal of the urban park soundscape. *Applied Sciences* **7**(1):91(1-16) (2017).