Landscape and Urban Planning 144 (2015) 34–48

Contents lists available at ScienceDirect

Landscape and Urban Planning

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

Research Paper

The acoustic summary as a tool for representing urban sound environments

Damiano Oldoni[a], Bert De Coensel[a,*], Annelies Bockstael[a], Michiel Boes[a], Bernard De Baets[b], Dick Botteldooren[a]

[a] Department of Information Technology, Ghent University, Belgium
[b] Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Belgium

HIGHLIGHTS

- The acoustic summary of a place is a collection of representative sounds.
- Acoustic summaries of several urban and quiet area locations are constructed using an automated procedure.
- A validation test with local residents assesses the quality of the acoustic summaries.
- Local residents can easily identify the acoustic summary extracted at the location of their own dwelling.
- A group of sounds describes the uniqueness of a place, rather than single sounds by themselves.

ARTICLE INFO

Article history:
Received 14 July 2014
Received in revised form 31 March 2015
Accepted 23 August 2015

Keywords:
Sound
Soundscape
Acoustic design
Auditory perception
Self-organizing map
Urban planning

ABSTRACT

Detecting and selecting sound events is emerging as an interesting technique for characterizing and representing the sound environment of a specific location. In this article we propose a computational model for automatically constructing a so-called acoustic summary, i.e. a comprehensive collection of sounds aiming to represent the specific sound environment at a given location. Such an acoustic summary could be used by architects, soundscape designers, and urban planners to explore – by listening – the sonic environment at a certain location as it is perceived by a human listener. The model is based on a self-organizing map, a type of neural network. It starts by extracting several psychoacoustic features from the sound. A specific, extensive and unsupervised training allows this map to be tuned to the typical sounds that are likely to be heard at the microphone location. The learning algorithm takes into account some basic aspects of human perception. For example, salient events tend to be better remembered than the ones that do not stand out, even if they occur less frequently. After the training, the self-organizing map is used to form an exhaustive acoustic summary by means of automatically recording specific sound events for the microphone location. In addition to describing the proposed tool, this paper also presents a validation test with local residents in order to show the ability of the model to pick up sounds which bring out the distinctiveness and the specificity of the soundscape as a local resident would do.

1. Introduction

Livability of the urban environment has always been a compelling issue for urban planners. Citizen well-being is related to the quality of the urban environment in different ways. Person-environment mismatch at the dwelling may lead to stress and related health impacts (Lazarus, 1991) but also the quality of the public space is of utmost importance. High quality public spaces stimulate social cohesion, recreation, and physical activity (Bedimo-Rung, Mowen, & Cohen, 2005). The role of urban green areas in particular has been investigated extensively in this respect. Several studies from the last decades indicate that people's psychological restoration and well-being is enhanced by direct access to nature and restorative areas (Hartig, Böök, Garvill, Olsson, & Gärling, 1996; Kaplan, 1983, 1985; Ulrich et al., 1991; Ulrich, 1981), by visual access to such areas from the dwelling (Kaplan, 1993, 2001; Ulrich, 1984) and by their perceived availability

* Corresponding author at: Ghent University, Department of Information Technology, Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium.
E-mail addresses: damiano.oldoni@ugent.be (D. Oldoni), bert.decoensel@intec.ugent.be (B. De Coensel), annelies.bockstael@intec.ugent.be (A. Bockstael), michiel.boes@intec.ugent.be (M. Boes), bernard.debaets@ugent.be (B. De Baets), dick.botteldooren@intec.ugent.be (D. Botteldooren).

http://dx.doi.org/10.1016/j.landurbplan.2015.08.013
0169-2046/

(Gidlöf-Gunnarsson & Öhrström, 2007). The positive role played by such areas has mainly been studied from the perspective of visual diversity, naturalness and esthetics. However, the role of the soundscape and in particular quietness and tranquility is increasingly being stressed (Gidlöf-Gunnarsson & Öhrström, 2007). Therefore, there is an increasing awareness of the fact that the sonic environment forms an essential component of the urban environment that requires as careful planning as the landscape (Carles, Barrio, & de Lucio, 1999; Liu, Kang, Luo, Behm, & Coppack, 2013; Liu, Kang, Behm, & Luo, 2014; Zhang & Kang, 2007). However, it is also shown that landscape and soundscape planning should not be tackled independently, as landscape indicators have a non-negligible impact on the soundscape (Liu et al., 2013, 2014).

Classically, urban sound has been treated as a waste product to be tackled with suitable noise control policies, for which the most popular and visible tool has been extensive noise mapping. However, the final goal of planning and designing urban environments is not only noise abatement, but the creation of spaces with matching positive acoustic qualities (Botteldooren, De Coensel, Van Renterghem, Dekoninck, & Gillis, 2008). This approach, typically referred to as the *soundscape approach*, is getting increasing multidisciplinary attention and is the subject of several projects and studies (Adams et al., 2006; Brown, Kang, & Gjestland, 2011; Pijanowski, Farina, Gage, Dumyahn, & Krause, 2011a; Pijanowski, Villanueva-Rivera, et al., 2011; Zhang & Kang, 2007). As the soundscape concept extends beyond the sonic or acoustic environment and includes the way it is perceived and understood by a typical user of the space and within a particular context, the tools at the disposal of the urban sound planner and soundscape designer should account for human auditory perception (Oldoni et al., 2013).

Today, physical registration of relevant acoustical parameters is commonly accepted as a first soundscape analysis step (Schulte-Fortkamp, Brooks, & Bray, 2008), followed by an evaluation of the perceptual effects by techniques such as targeted interviews and questionnaires, preferably involving community members who live at the location under study (Axelsson, Nilsson, Hellström, & Lundén, 2014; Brooks, 2006). The combination of these two approaches is called *combined soundscape analysis* (Adams et al., 2006; Schulte-Fortkamp et al., 2008) and it is often deployed by means of soundwalks, in which sound measurements and perception interviews are conducted simultaneously. In a research perspective, the results are combined in order to find quantitative relationships between physical sound indicators and perceptual attributes (Berglund & Nilsson, 2006; Liu et al., 2014). Soundwalks are a popular methodology for understanding outdoor soundscapes (Adams et al., 2008), but they are inherently short-term and typically include only daytime. For this reason, several long-term strategies have been developed, mainly based on mobile sound measurements and community involvement, e.g. with public workers such as local police officers (Schulte-Fortkamp et al., 2008). This approach is surely more detailed and complete, but requires a considerable organizational effort and regular and constant participation, resulting in feasibility and reproducibility issues. In both short and long term approaches, a methodology for systematically selecting and recording a comprehensive collection of sounds that is representative for the sonic environment in the way that it is perceived and understood by so-called "local experts" – inhabitants and visitors – would mean a significant step forward in soundscape methodology.

In this paper a neural-network-based model is proposed that automatically constructs an *acoustic summary*, i.e. a collection of sounds that are likely to be noticed at a particular location and together represent the sonic environment at that location. The acoustic summary can provide a quick overview of the sounds present at a specific location, thus being a useful tool for the urban planner and the soundscape designer. In contrast to most of the computational auditory scene analysis (CASA) models (see Wang and Brown, 2006 for an overview), the major interest here does not lie in extracting as clean as possible sound samples for all components of the auditory scene. On the contrary, the intention is to summarize the sonic environment using only those sounds that a human observer, not particularly focusing its attention to environmental sound, would notice. Note this explicit limitation of the acoustic summary to holistic listening only. Listening is a process that can develop at different cognitive levels, and it could be attentive and analytic rather than holistic. However, within attentive and analytic listening, top-down information is taken into account, which is much harder to implement in a computational model.

The proposed model partly takes inspiration from specific CASA techniques for extracting salient fragments of the auditory scene but it is also inspired by mechanisms underlying human bottom-up attention (Duangudom & Anderson, 2007; Kalinli & Narayanan, 2007; Kayser, Petkov, Lippert, & Logothetis, 2005). Moreover, most CASA techniques are not context dependent. Distinguishing between frequently occurring sounds and out-of-context or rarely occurring sounds is a crucial aspect in constructing an acoustic summary. For this reason, besides a biologically inspired auditory processing model, learning is a very important aspect in the presented model. It is implemented by means of a neural network called *Self-Organizing Map* (SOM) or Kohonen Map (Kohonen, 2001) and a specifically tailored learning technique. Furthermore, the model attempts to create a compromise between biological accuracy and computational efficiency as the model is to be integrated in equipment for long-term outdoor measurement and the data processing underlying the decision whether or not to record particular sound events has to be performed in real-time.

The structure of this paper is as follows: Section 2 describes the neural-network based model to construct the acoustic summary. Section 3 is dedicated to the results of a validation test performed by local residents in order to assess how accurately the acoustic summary is representing the sound environment in their neighborhood. Section 4 discusses the results and future developments. Finally, in Section 5 conclusions are presented.

2. Methods

2.1. Overview

Constructing the acoustic summary requires a computational analysis of the auditory scene that mimics how a human observer would split this auditory scene in its relevant components. Considering the application of the model in long-term outdoor measurement stations, computational efficiency has to be considered. For this reason, existing detailed auditory processing models for loudness (Glasberg & Moore, 2002), masking (Glasberg & Moore, 2005) and auditory saliency (Kayser et al., 2005) are replaced by simplified versions. The proposed model is comprised of two main stages, illustrated in Fig. 1: (I) during the learning phase, a self-organizing map (SOM) is tuned to the typical sounds at the given location based on the sound level and its spectrum, and (II) during the acoustic summary formation phase, for each class of sounds thus obtained, prototypes are recorded to compile the acoustic summary. Real-time operation is required in the second stage due to the limited sound buffer of typical outdoor measurement stations. In both stages, the sound signal recorded by the microphone is first treated in a similar way as in the human peripheral auditory system (I.a and II.a), whereby both a set of acoustical features is extracted and a measure of auditory saliency is calculated. The learning stage classifies the acoustical features based on co-occurrence (I.b) using the incremental SOM algorithm and a training technique called *Continuous Selective Learning* (CSL) that