How to measure soundscape quality

Östen Axelsson[1]
Gösta Ekman Laboratory, Department of Psychology, Stockholm University, Stockholm, Sweden.

Summary
The Swedish Soundscape-Quality Protocol has been criticized for being insufficient, because it proposes to assess soundscape quality by a Good–Bad Scale, and alternatively by eight attribute scales assessing the perceived affective quality of a soundscape. Critics argue that further alternative definitions of 'soundscape quality' must be explored. In particular they argue for assessing 'soundscape quality' by asking to what extent a soundscape is appropriate to a place. The Sound Cities project at School of Architecture, University of Sheffield, in the UK, investigated this issue by a listening experiment involving 50 university students and 25 urban and peri-urban areas from the UK. The results indicate that the Good–Bad Scale is correlated with the perceived affective quality of a soundscape. Conversely, the appropriateness of a soundscape to a place is orthogonal to the former two assessments and provides additional information. Thus, a soundscape can be appropriate to a place even though it is poor. This raises the issue of which information should be given priority. Probably the best recommendation is to assess soundscape by perceived affective quality. In addition, it is possible to complement this assessment by assessing the appropriateness of the soundscape to the place. However, the latter assessment should not be used on its own, as this may lead to unfortunate conclusions.

PACS no. 43.50.Qp, 43.50.Rq, 43.66.Lj

1. Introduction

Soundscape research is gaining in importance, particularly in Europe. This is much thanks to the European Environmental Noise Directive [1] that provides that the Member States of the European Union must protect what is known as 'quiet areas.' In 2014 the European Environment Agency published a good practice guide on quiet areas [2]. It recommends four methods for identifying and assessing quiet areas. The soundscape approach is one of the methods. The European Environment Agency also calls for further in-depth research on how to assess human appreciation of quiet areas and perceived acoustic quality. In the same year the International Organization for Standardization (ISO) published an International Standard providing a definition and a conceptual framework of 'soundscape' [3]. This development highlights the need for an instrument for assessing soundscape, and soundscape quality in particular.

The Swedish Soundscape-Quality Protocol defines soundscape quality in two ways: (1) by a Good–Bad Scale, and (2) by a set of eight attribute scales for assessing the perceived affective quality of a soundscape (see e.g. [4, 5]). Researchers with a background in environmental planning has criticized this, and claim that soundscape quality rather is a matter of how appropriate the soundscape is to a place (see e.g. [6, 7]).

The Sound Cities project at the School of Architecture, University of Sheffield in the UK, was initiated to investigate this issue further. The present paper reports preliminary results from a listening experiment in which 50 university students assessed a set of 25 urban and peri-urban areas with regards to their social and recreational activities and with regards to their soundscapes. As in the Swedish Soundscape-Quality Protocol, soundscape quality was assessed with a Good-Bad Scale and by perceived affective quality. In addition, a scale for assessing how appropriate the soundscape is to a place was developed and included in the experiment. The aim was to investigate which method for assessing soundscape quality would receive the most empirical support.

[1] Email address: oan@psychology.su.se

2. Method

2.1. Participants

Participants were 25 female and 25 male students from The University of Sheffield, aged 18–32 years (M_{age} = 21.6 years, SD = 3.2). They were recruited through the university's website for research volunteers. All were born and raised in the UK, and had English as their mother tongue. Persons with hearing thresholds at or below 20 dB in both ears for all the tested frequencies (0.125, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6 and 8 kHz) as assessed with an audiometer (Kamplex Clinical Audiometer KC35, or Labat AudioLAB 1.0.2.716), were allowed to take part in the experiment. An exception was made for the frequency 0.125 kHz if the threshold was at or below 30 dB, because this frequency turned out to be difficult for 23 of the 50 participants. This was particularly associated with the Labat AudioLAB audiometer. An additional exception was made for six participants who had a hearing threshold at or below 30 dB in one of the other tested frequencies in one ear. Participants received a small monetary compensation for volunteering in the experiment.

2.2. Stimuli

The stimulus material consisted of 25 videos (46 s) presenting an urban or peri-urban outdoor area from Sheffield, London or Brighton. Each area was documented with binaural recordings and digital photographs taken during the recording sessions. From every area, six photographs were selected to be included in a video, together with the binaural sound excerpts that correspond to the six photographs. Examples of documented areas are the Sheaf Square in Sheffield, Covent Garden in London, and Royal Pavilion Gardens in Brighton. The aim was to achieve a wide range of urban and peri-urban areas from city center to city boarder and where local residents would spend some of their free time.

Photographs and binaural recordings were edited into slideshows, beginning with 1 s of a black frame and then a 2 s fade-in-from-black to the first photograph. All six photographs were presented for 5 s, followed by a 2 s crossfade between the photographs. All slideshows finished with a 2 s fade-out-to-black, and finally a 1 s black frame. The binaural soundtracks were edited with an initial 2 s fade in, making a soundtrack to start 1 s before the first photograph. The intention was to center the participant's attention to the soundtrack rather than to the photographs. This is also the reason why photographs were used instead of video recordings. Thereafter 2 s crossfades between the binaural excerpts were synchronized with the crossfades for the slideshow. The soundtracks finished with a 2 s fade out, making a soundtrack to finish 1 s after the last photograph in a slideshow, again emphasizing the importance of the soundtrack.

Every area was photographed according to one of three basic principles, or a combination of the three principles. An area could be photographed from one single spot. The photographer then rotated at the spot to take six successive photographs to present a panorama of the area. Alternatively the photographer walked along the perimeter of the area, taking six photographs across, towards the center of the area. Finally, the photographer could walk along a path through the area, taking six successive photographs along the path. As stated above, the acoustic environment was recorded simultaneously.

2.3. Equipment

The binaural recordings were conducted with a stereo pair of omnidirectional miniature microphones (DPA 4060) connected to a digital audio recorder (Sound Devices 722; 24 bits, 48 kHz) via a pair of XLR adapters (DAD6001-BC). The miniature microphones were mounted in the photographer's outer ears. The photographs were taken with a Nikon Coolpix 5700, using the widest possible angle of the built in zoom lens.

The experiment was conducted in a semi-sound-proof listening booth. To present the videos and to collect the participants' responses the www.surveygizmo.com Internet based questionnaire tool was used. The tool was run on a Lenovo ThinkPad Edge E530 laptop, with Windows 7 Professional (64 bit) Operating System, in a Mozilla Firefox web browser using Fast Ethernet. The 25 videos were saved in the MPEG-4 format on the www.surveygizmo.com server. Sounds were played back at the authentic sound-pressure level (01dB sound calibrator), using a pair of closed, diffuse-field headphones (beyerdynamic DT 770 PRO, 250 ohms) connected to an external audio interface (RME Babyface) via a stereo headphone amplifier (Lake People G109-P). Closed, as opposed to open, headphones were used to protect from any background sounds in the listening booth.

2.4. Soundscape assessment instrument

The participants assessed the areas presented in the videos by answering five questions. Question 1 concerned to what extent 15 social and recreational activities were suitable for an area. The participants responded to each of them by the aid of slide bars delimited by "Not at all (0 %)" and "Perfectly (100 %)". Question 2 concerned to what extent the participants could perceive five different kinds of sound sources in a video: (1) traffic, (2) construction, maintenance, industry, loading of goods, etc., (3) sounds of individuals, (4) crowds of people, and (5) natural sounds. They responded by the aid of slide bars delimited by "Do not hear at all" and "Dominates completely". Question 3 concerned to what extent the participants perceived the acoustic environment in a video as good. They responded by the aid of slide bars delimited by "Very bad" and "Very good". Question 4 concerned the perceived affective quality of the acoustic environment in a video. The participants assessed to what extent they agreed that eight affective attributes applied: 'pleasant,' 'vibrant,' 'eventful,' 'chaotic,' 'annoying,' 'monotonous,' 'uneventful,' and 'calm.' They responded by the aid of slide bars delimited by "Strongly disagree" and "Strongly agree". Question 5 concerned to what extent the acoustic environment was appropriate to the place in a video. The participants responded by the aid of a slide bar delimited by "Not at all (0 %)" and "Perfectly (100 %)". The response values of all slide bars were set to 0–100 (increment = 1).

Questions 1 and 2 were presented in the same order to all participants and for all videos. Questions 3–5 where presented in a new random order to all participants and for every video. Also all items in a list (i.e., the 15 social and recreational activities in Question 1, the five kinds of sound sources in Question 2, and the eight affective attributes in Question 4) were always presented in a new random order to all participants and for every video.

2.5. Design

Every participant assessed 10 of the 25 videos in an individual, irregular order. 50 subsets of 10 videos were organized and assigned randomly to the 50 participants. Every video was assessed by 20 participants.

2.6. Procedure

First, the experimenter checked the participant's hearing ability. Persons who passed the hearing test took part in the experiment. After the experiment the participants received a small monetary compensation for volunteering in the experiment. An experimental session lasted for 30–60 min.

3. Results

Arithmetic mean values were calculated for all variables across the 20 participants who had assessed a video. Thus, cases in the data matrix used in the analyses represent the 25 videos.

The set of variables was reduced with the aid of Principal Components Analysis (PCA) using SPSS 22 for Windows. The set of 15 social and recreational activities in Question 1 was reduced to three oblique components with Eigenvalues lager than 1.0 (Direct OBLIMIN rotation). They explained 58, 19 and 10 % of the variance in the data set. Components scores were calculated by the regression method. The first component was positively associate with active street life and shopping, and negatively associated with outdoor exercise, picnic and escaping city stress. It was interpreted as to represent *Urban Environments*. The second component was positively associated with socializing, appreciation time with friends and family, and with shopping. It was interpreted as to represent *Social Environments*. The third component was positively associated with swimming and bathing, boating and fishing, as well as appreciation inland water. It was interpreted as to represent *Aquatic Environments*. The first two were uncorrelated, whereas *Aquatic Environments* had a statistically significant negative Pearson's correlation coefficient with *Urban Environments* ($r = -0.503$, $p = 0.01$).

Also the set of eight affective attributes in Question 4 were reduced, into two orthogonal components with Eigenvalues larger than 1.0 (VARIMAX rotation). They explained 58 and 37 % of the variance in the data set. Components scores were calculated by the regression method. The first component was positively associated with 'eventful,' and negatively associated with 'uneventful.' It was interpreted as to represent *Eventfulness*. The second component was positively associated with 'pleasant,' and negatively associated with 'annoying.' It was interpreted as to represent *Pleasantness* (cf. [5]).

In order to explore the relationships between the reduced set of variables (the five components identified above, the five kinds of sound sources in Question 2, and Questions 3 and 5), it was subjected to a new PCA. It resulted in three principal components with Eigenvalues larger than 1.0. They explained 43, 33 and 11 % of the variance in the data set. The first component was positively associated

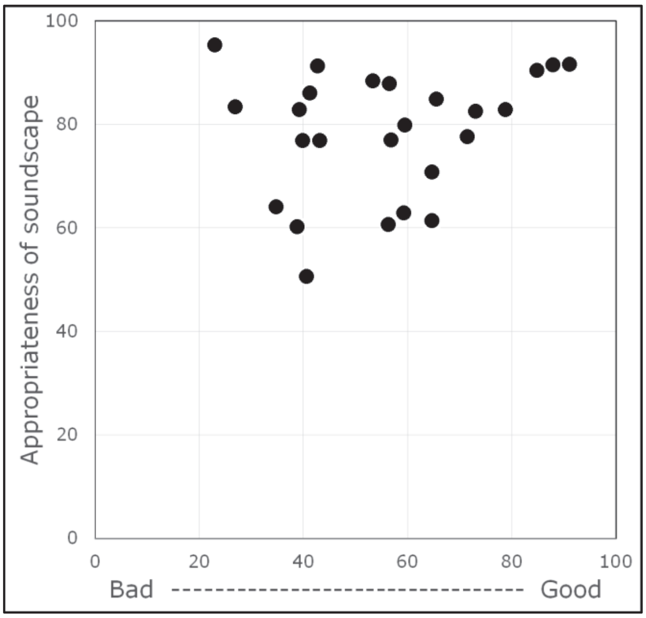

Figure 1. Scatterplot presenting the extent to which a soundscape is good on the X-axis and how appropriate a soundscape is to a place on the Y-axis. Data points represent the 25 videos.

with natural sounds, *Pleasantness*, and to what extent the soundscape was perceived as good (Question 3). It was negatively associated with *Urban Environments*, construction noise etc., and the sound of traffic. This indicates that the first component represents Natural vs Urban Environments, and that natural environments are perceived as pleasant and good. The second component was positively associated with *Eventfulness*, *Social Environments*, the sound of crowds and the sound of individuals. This indicates that the second component represents Social Environments, and that such environments are likely populated with people and perceived as eventful. The third component was mainly associated with the extent to which the soundscape was perceived as appropriate to a place (Question 5). Thus, whereas all the other variables are well explained by the first two components in the model, how appropriate the soundscape is to a place seems to represent a separate kind of quality.

Figure 1 presents a scatter plot of the relationship between the arithmetic mean values for the Good–Bad Scale (Question 3) and for how appropriate the soundscape is to a place (Question 5). It shows that the Good–Bad Scale serves as a lower limit for how appropriate a soundscape is. Among the 25 areas included in the study, no soundscape was perceived as good and inappropriate. However, there are areas with a soundscape that the participants perceived as bad but appropriate to the place. This was in particular at a road bridge over the very busy motorway M1 outside Sheffield, and the Charing Cross roundabout by Trafalgar Square in the center of London.

In order to investigate what variables may predict soundscape quality, the data was subjected to a set of linear multiple regression analyses using the stepwise method (SPSS 22 for Windows). The three oblique components *Urban Environments*, *Social Environments* and *Aquatic Environments*, derived from Question 1, were used as independent variables, together with the five sound-source variables from Question 2.

In the first regression model the Good–Bad Scale (Question 3) was used as dependent variable. It was best predicted by the sound of traffic ($\beta = -0.777$, $t = -9.33$, $p < 0.001$) and by *Urban Environments* ($\beta = -0.289$, $t = -3.47$, $p < 0.01$). Both variables had a negative influence. Together they explained 87 % of the variance (R^2) in the Good–Bad Scale ($F_{2,22} = 75.65$, $p < 0.001$).

In the second model the component *Pleasantness*, derived from Question 4, was used as dependent variable. It was best predicted by natural sounds ($\beta = 0.268$, $t = 2.09$, $p = 0.049$), the sound of traffic ($\beta = -0.547$, $t = -6.28$, $p < 0.001$), and by *Urban Environments* ($\beta = -0.300$, $t = -2.88$, $p < 0.01$). Natural sounds had a positive, and the other two a negative influence. Together they explained 92 % of the variance in *Pleasantness* ($F_{3,21} = 78.71$, $p < 0.001$).

In the third model the component *Eventfulness*, derived from Question 4, was used as dependent variable. It was best predicted by the sound of crowds ($\beta = 0.677$, $t = 6.34$, $p < 0.001$) and by *Social Environments* ($\beta = –0.321$, $t = –3.01$, $p < 0.01$). Both had a positive influence. Together they explained 91 % of the variance in *Eventfulness* ($F_{2,22} = 104.27$, $p < 0.001$).

In the fourth model the appropriateness of the soundscape to a place (Question 5) was used as dependent variable. It was best predicted by *Urban Environments* ($\beta = 0.937$, $t = 5.02$, $p < 0.001$), the sound of traffic ($\beta = –0.796$, $t = –4.35$, $p < 0.001$), and the sound of individuals ($\beta = –0.596$, $t = –3.34$, $p < 0.01$). The first had a positive, and the other two a negative influence. Together they explained 57 % of the variance in how appropriate a soundscape is to a place ($F_{3,21} = 9.45$, $p < 0.001$).

4. Discussion and concluding remarks

The results indicate that, for the present 25 areas, the Good–Bad Scale largely provides the same information as the component *Pleasantness*, which is related to whether the environment is natural or urban. *Eventfulness*, which is orthogonal to *Pleasantness*, provides additional information, related to the presence of people. Also, the appropriateness of the soundscape to the place, which was found to be orthogonal to the former three variables, provides additional information.

The Good–Bad Scale, *Pleasantness* and *Eventfulness* were all well predicted by the two components *Urban Environments* and *Social Environments*, together with the dominance of the five kinds of sound sources assessed through Question 2. For all, approximately 90 % of the variance was explained. On the other hand, only 57 % of the variance in the appropriateness of a soundscape was explained by the independent variables. Nevertheless, the regression model for the appropriateness of a soundscape indicates that an appropriate soundscape would be one in a downtown area void of the sounds of traffic and people. That is, a ‘quiet urban area.’ Conversely, an inappropriate soundscape would be one in a rural area dominated by the sound of traffic and people. Still, 43 % of the variance remains to be explained.

That the appropriateness of a soundscape deviates so markedly from the other soundscape-quality variables raises some concerns as to how useful this information is to soundscape assessment in practice. As Figure 1 shows, it is possible for a soundscape to be appropriate to a place even though it is poor. On the one hand, this means that the appropriateness of a soundscape provides information beyond whether the soundscape is good or bad. On the other hand, if the appropriateness of a soundscape is the only information available, there is a risk that one concludes that the soundscape is fine as it is, although it could be poor. Which information should then have priority, the appropriateness of the soundscape or how good or bad it is?

Based on the present and previous research (e.g., [4, 5]), this author recommend to assess soundscape by perceived affective quality in terms of *Pleasantness* and *Eventfulness*. These two components seem to summarize most of the relevant information. In addition, it is possible to complement this assessment by assessing the appropriateness of the soundscape to the place. However, the latter assessment should not be used on its own, as this may lead to unfortunate conclusions.

Acknowledgements

This research was conducted while the author was a Newton International Fellow at the School of Architecture, University of Sheffield, in the UK. The fellowship was sponsored jointly by The Royal Society and British Academy. Professor Jian Kang was the host of the fellowship in Sheffield. The author also acknowledges the Marianne and Marcus Wallenberg Foundation.

References

[1] EC: 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 of the European Communities 45 (2002) 12–25.

[2] EEA: Good Practice Guide on Quiet Areas. European Environment Agency (EEA), Copenhagen, 2014.

[3] ISO: ISO 12913-1:2014. Acoustics—Soundscape—Part 1: Definition and Conceptual Framework. International Organization for Standardization (ISO), Geneva, 2014.

[4] Ö. Axelsson, M.E. Nilsson, B. Berglund: A Swedish instrument for measuring soundscape quality. In J. Kang (Ed.), Euronoise 2009: Action on noise in Europe. Institute of Acoustics, Edinburgh, 2009, Paper EN09_0179. [Available on CD]

[5] Ö. Axelsson, M.E. Nilsson, B. Berglund: A principal components model of soundscape perception. Journal of the Acoustical Society of America 128 (2010) 2836–2846.

[6] A.L. Brown: A review of progress in soundscapes and an approach to soundscape planning. International Journal of Acoustics and Vibration 17 (2012) 73–81.

[7] A.L. Brown, J. Kang, T. Gjestland: Towards standardization in soundscape preference assessment. Applied Acoustics 72 (2011) 387–392.