THE SELF-ASSESSMENT MANIKIN AND HEART RATE: RESPONSES TO AURALISED SOUNDSCAPES

FRANCIS STEVENS[1], DAMIAN T MURPHY[2] AND STEPHEN L SMITH[3]

[1]AudioLab, Department of Electronics, University of York, UK
e-mail: fs598@york.ac.uk
[2]AudioLab, Department of Electronics, University of York, UK
e-mail: damian.murphy@york.ac.uk
[3]Intelligent Systems Group, Department of Electronics, University of York, UK
e-mail: stephen.smith@york.ac.uk

September 23rd 2016.

Abstract

Listening to a soundscape is a very complex auditory experience that can invoke a strong emotional response. There are relatively few objective tools that allow for an individuals experience of a soundscape to be measured and for the benefits of soundscape elements to be explored. The measurement of physiological responses to soundscapes has been examined by several previous studies, each reaching different conclusions or making use of different physiological measures. This paper presents the findings of a soundscape preference rating study where each test subject was presented with sixteen soundscape clips (each 30s in duration) over a 16-speaker surround-sound listening setup. Heart rate (HR) and skin conductance (SC) were measured throughout using a Shimmer device and the self-assessment-manikin (SAM) was used to measure subjective valence, arousal, and dominance. This paper presents preliminary findings from this study and examines the relationship between the HR measurements and the SAM results.

1 Introduction

Environmental noise has been increasingly recognised as a form of pollution equally important as more traditional pollutants [1]. In order to understand how the negative effects of noise pollution arise, a method extending beyond noise level measurement is required. One such method is auralisation where a measured or simulated soundscape can be presented in a lab environment, and subjective responses to that soundscape can then be measured [2].

The established method for gathering these subjective responses is the use of SD pairs [3] to rate soundscapes in terms of multiple scales. This method can be time consuming, due to the cumbersomeness of measuring perhaps 18 or more ratings for each stimulus in a given test. It can also be non-intuitive for non-experts where specific terms are used, and relies on an understanding of the terms used which requires translation and validation for use in multiple languages [4]. The SAM consists of only three scales presented in pictorial form and was generated in an attempt to design a preference rating tool free from these problems.

Alongside these subjective responses, biometric measurement tools can be used to concurrently record an individual's objective physiological response. Many previous studies [5–9] have investigated the relationship between various physiological quantities (heart rate, skin conductance, respiratory rate, EMG and EEG measurement) and subjective rating scales. This project involved the use of a Shimmer GSR+ device [10] to record heart rate (HR) and galvanic skin response (GSR) values in participants exposed to a variety of auralised soundscape recordings.

This paper begins with a description of the methodology, including the rationale behind the collection of the soundscape data, the subjective assessment methods used, and the test procedure itself. Results are then presented examining the relationship between the subjective assessment methods used and the HR measurements made. The paper finished with a concluding section which details avenues for further research both in terms of further analysis of the data collected and future research.

2 Methodology

This section presents each stage of the experimental methodology, including the collection of suitable soundscape data, the subjective assessment tools used, and the procedure of the listening test itself including physiological measurement.

2.1 Data Collection

During Summer 2015 data were collected from 8 locations around the north of England covering a wide range of environments from rural to suburban and urban. At each location audio-visual recordings were made using a 4-channel Soundfield surround-sound microphone and 6-GoPro cameras mounted on a cube allowing for the capture of spherical images. The visual data were collected for use in future experimentation. The aim when choosing the recording locations was to cover as wide a range of sound sources, noise levels, and visual features as possible.

2.1.1 Sound Sources

In order to select a set of recording locations covering as wide a range of soundscapes as possible, previously identified categories of soundscapes and their components sound sources had to be considered. In a significant quantity of soundscape research [11–16] three main groups of sounds are identified:

- **Natural sounds:** These include animal sounds (bird song is an oft-cited example), and other naturally occurring environmental sound.

- **Human sounds:** Any sounds that are representative of human presence/activity that do not also represent industrial activity. Such sounds include footsteps, speech, coughing, laughter etc.

- **Industrial sounds:** Mechanical sounds, such as traffic noise, activity on a building site, or aeroplane noise.

The purpose of covering such a wide range of sources was to ultimately generate a set of stimuli that will elicit a wide range of emotional responses. Generally speaking, natural sounds are the most preferred, human sounds are given a neutral rating, and mechanical/industrial sounds are disliked [15].

2.1.2 Recording Duration

Several minutes of material were recorded at each location, from which two 30-second long clips were extracted. Table 1 contains details of the sound sources present in the two 30-second long clips chosen to represent each recording location. Where here the clips are numbered as ‘1’ and ‘2’ for each location, when referred to more generally they have each been given a number between 1 and 16, determined by the order of the locations. For example, the clips numbered 3 and 4 are respectively the 1st and 2nd clips recorded at location 2.

A survey of previous literature showed the duration of recordings used for soundscape reproduction to vary considerably. Whilst Harriet made use of 7-minute long soundscape representations [13] constructed artificially from recorded material, other studies typically use shorter recordings (especially those presenting visual and aural stimuli simultaneously). For example both Anderson et al. [15] and Viollon et al. [12] used 20-second long recordings. Pheasant et al. have used 32-second long recordings [17,18], and both Watts and Pheasant and Gifford and Ng make use of recordings lasting 1-minute [16,19]. Axelsson’s work as part of the Sound Cities project used binaural recording of 46-seconds in length, presented with a set of six still images of the recording site [20]. Rummukainen et al. used even shorter recordings only 15-seconds in length [21]. From the recordings made at each location two 30-second long were selected. This length was deemed long enough for the test participants to get a real sense of the soundscape (i.e. sufficient for a good level of ecological validity [22]) as well as being long enough for physiological changes to take effect.

2.2 Subjective Assessment

The subjective assessment tools used in the listening test including both Semantic Differential (SD) pairs and the Self-Assessment manikin (SAM). This paper will focus on the results from the SAM only.

The Self-Assessment Manikin (SAM) is a method for measuring emotional responses developed by Bradley and Lang in 1994 [4]. It was developed from factor analysis of a set of SD pairs rating both aural [5, 23] and visual stimuli [24] (using both the International Affective Digital Sounds database, or IADS, and the International Affective Picture System, or IAPS) . The three factors developed for rating emotional

Site	Clip 1 Sound Sources	Clip 2 Sound Sources
1. Dalby Forest	Birdsong, owl hoots, wind	Birdsong, goose honking, insects, aeroplane noise
2. Dalby Forest Lake	Wind, birdsong, insects, single car	Wind, birdsong, insects, water splashing
3. Hole of Horcum	Birdsong, traffic, bleating	Birdsong, traffic, conversation
4. Fox & Rabbit Inn	Traffic, car door closing	Traffic, car starting, footsteps
5. Smiddy Hill	Car starting, car door closing, traffic	Traffic, birdsong
6. Albion Street	Busker performance, footsteps, conversation, distant traffic	Workmen, footsteps, conversation
7. Park Row	Traffic, buses, wind, 'flute' playing	See clip 1 details
8. Park Square	Conversation, traffic, birdsong, shouting	Workmen, conversation, birdsong, traffic

Table 1: Details of the sound sources present in the two 30 second long clips used for each location.

response to a given stimuli are:

- **Valence:** How positive or negative the emotion is, ranging from unpleasant feelings to pleasant feelings of happiness.
- **Arousal:** How excited or apathetic the emotion is, ranging from sleepiness or boredom to frantic excitement
- **Dominance:** The extent to which the emotion makes the subject feel they are in control of the situation, ranging from not at all in control to totally in control.

These results were then used by Bradley and Lang to create the SAM itself as a set of pictorial representations of the three identified factors. The version of the SAM used in this experiment is shown in Figure 1.

The SAM has been used a select number of times for soundscape analysis, recently by Watts and Pheasant [16], and combined with concurrent physiological measures by Hume and Ahtamad [7]. However, a direct comparison of SD pair ratings with SAM results has not been conducted. Other studies have investigated the use of Russell's circumplex affect model [25] to study urban environments. Hull and Harvey found a relationship between the physical characteristics of suburban parks and affective states: tree density, and the presence of undergrowth and pathways [26], while Hanyu found that green, open, and well-kept spaces are related to positive valence, but that the presence of 'disorder elements' (vehicles, wires) is related to negative affective response [27]. Also of note is Viollon and Lavandier's identification of valence and arousal as the two main underlying factors in assessment of environmental quality [28], indicating strong possibility of a high correlation between SD pairs and the SAM.

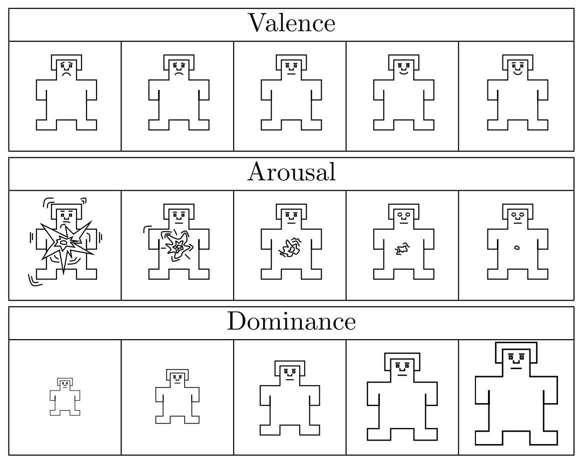

Figure 1: The Self-Assessment Manikin (SAM) as used in this experiment, after [4].

2.3 Physiological Measurement

A Shimmer device is used to measure heart rate (HR) and Galvanic Skin Response (GSR) [10]. The Shimmer device is small measurement unit that is attached to the participant's wrist with a velcro strap that transmits its measurements to a PC via Bluetooth. The HR measurements are made using a pulse oximeter to obtain a Photoplethysmogram (PPG) of the participant's pulse which the Shimmer software then uses to calculate heart rate. Fig. 2 shows the positioning of the Shimmer device to make these measurements. This focus of this paper is on the HR results, and the GSR measurements will be analysed in the future.

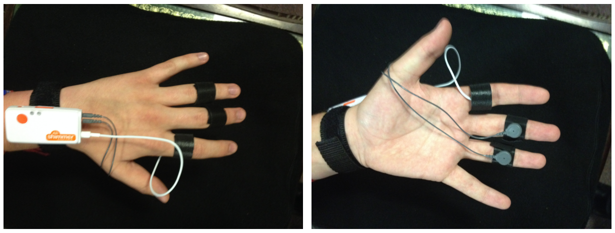

Figure 2: Electrode Positions. The electrodes with the grey leads were used to measure GSR, and the probe on the index finger is the pulse oximeter used to obtain the HR measurement.

2.4 Test Procedure

Each test subject is first presented with the pre-experiment statement and consent form, followed by a demographic questionnaire and a preview of the subjective assessment questionnaire they will use to rate each presented soundscape. This was to give them the opportunity to raise any questions they may have about the questions they will be answering, and to familiarise themselves with the test procedure. Following this they are lead into the listening space.

Subjects are then presented with each of the soundscape recordings in random order over the 16-speaker surround-sound listening setup. After each recording has finished they are given time to fill out a subjective assessment form for each one. The typical duration of the entire procedure was around 40 minutes and for the duration of each test physiological measurements were taken, and the time at which each recording was presented was also noted.

3 Results

This section begins with a brief presentation of some of the SAM results followed by an examination of the HR results. Two different methods are used to analyse the measured HR data, and the results of those analyses are then consider in terms of their relationship to the Valence and Arousal dimensions of the SAM.

3.1 Subjective Data

A full analysis of the subjective data (included results from The SAM and SD pairs) can be found in another study [29] where it was found that the SAM could be justifiably used as a substitute subjective assessment tool to replace SD pairs. The Dominance dimension of the SAM was shown to offer little additional information over that provided by the Valence and Arousal dimensions. This is posited to be due to a general confusion regarding the meaning of the dominance dimension in a soundscape context, given it's basis in individual's reactions to less ambient sounds [4].

Fig. 3 shows the mean Valence and Arousal scores across all test participants for each soundscape recording. As expected these two dimensions of the SAM are shown to not be significantly correlated ($r = -0.4649$, $p = 0.2782$), indicating that they describe different dimensions of the emotional states evoked by the soundscapes presented.

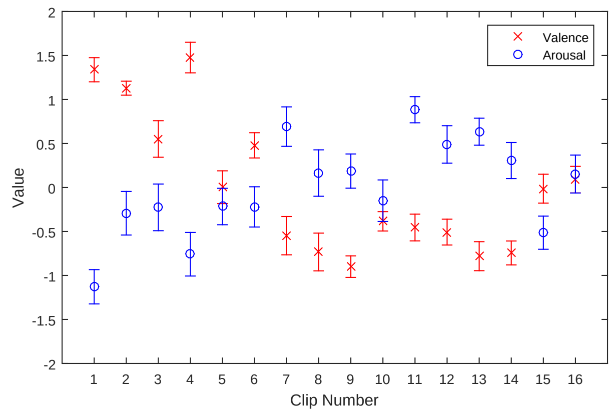

Figure 3: Mean Valence and Arousal scores (and standard error bars) for each clip.

3.2 Heart Rate Data

Similarly the way in which these previous studies have used the collected HR data to reach these results has differed significantly. Making use of 30-second long clips Gomez and Danuser compared the last 15-seconds of each presented stimuli with the last 1seconds of the preceding rest period [6], while Hume and Ahtamad made use of 8-second long soundscape clips and compared the mean HR across the entirety of the clip's duration with the mean HR during the last 8-seconds prior to the clip being played [7]. As such it was decided to examine the HR data recorded here in two ways: firstly comparing mean HR for the entirety of a clip's duration with the preceding 30-seconds; and secondly comparing the mean HR in the last 15-seconds of each clip with the first 15-seconds. These will be referred to as Clip-Rest and Within Clip comparison respectively from here on.

Fig. 4 shows an example HR measurement from the experiment. It shows 1 minute of recorded HR data, where the last 30-seconds correspond to the presentation of a soundscape recording, and the first 30-seconds were part of the preceding rest period. The sample rate for all of the HR measurement made was $f_s = 51.2$Hz.

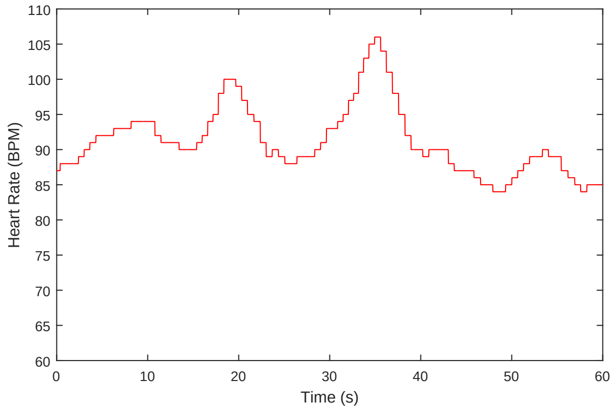

Figure 4: Example HR measurement taken during a particpant's exposure to a soundscape recording.

In order to account for the variation in HR measurement between subjects, rather than using the raw beats-per-minute (BPM) values the change in BPM will be expressed as a percentage. The percentage change in BPM ($BPM_{\%}$) is calculated by using:

$$BPM_{\%} = 100 \cdot \frac{BPM_2 - BPM_1}{BPM_1} \quad (1)$$

where BPM_1 is the mean BPM of baseline measurement currently being used, and BPM_2 is the mean BPM of the are currently being compared the baseline [9]. In the case of the Clip-Rest comparison BPM_1 is mean BPM of the last 30-seconds before the clip is presented, and BPM_2 is the mean BPM for the whole 30-second duration of the soundscape recording. For the Within Clip comparison BPM_1 is the mean BPM in the first 15-seconds of the recording, and BPM_2 is the mean BPM in the final 15-seconds.

3.2.1 Clip-Rest Comparison

Fig. 5 shows the same HR measurement as Fig. 4 with added lines to show the mean BPM during the rest period prior the presentation of the soundscape recording, and during the presentation of the clip itself. The calculate percentage change in BPM is also indicated.

3.2.2 Within Clip Comparison

Fig. 6 shows only the section of HR measurement from Fig. 4 during which the participant was being exposed to a soundscape recording. Again there are added lines to show the mean BPM during the first and final 15s of the clip, with the calculated percentage change in BPM also indicated.

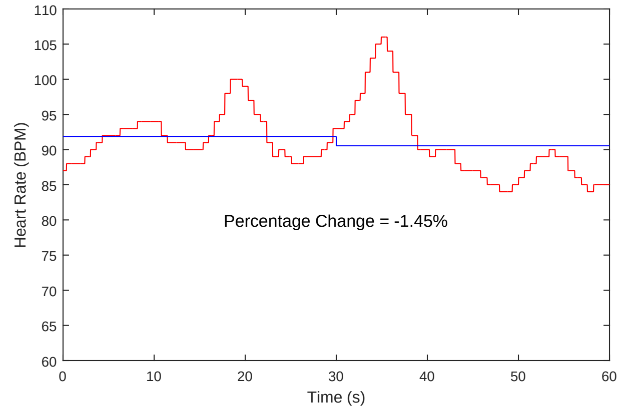

Figure 5: Plot showing the same BPM reading as presented in Fig. 4. Also indicated are the mean BPM values in the first 30-seconds (before the participant was exposed to the clip) and the last 30-seconds (during which the participant was exposed to the clip).

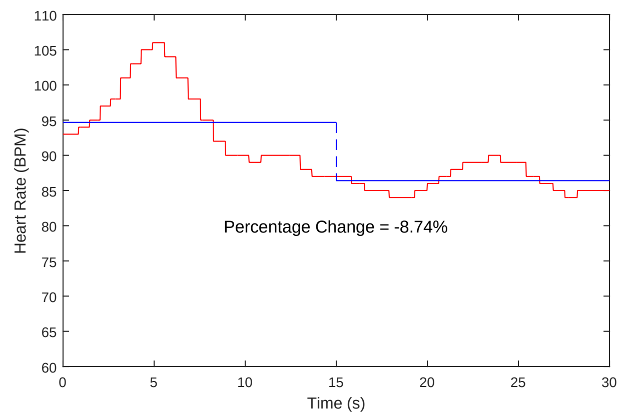

Figure 6: This is the same example BPM reading as shown in Fig. 4. Added here are horizontal lines showing the mean BPM value for the first and last 15-seconds of the clip.

3.2.3 Comparison Results

Fig. 7 shows the mean BPM percentage change for both comparison methods for each clip. The Within Clip results consistently show an average drop in HR, where the Clip-Rest results include some measurements indicating a rise in HR. The Within Clip results are therefore in line with previous research [5,7] that found HR to drop with the presentation of any stimuli. Where these results differ is in the level of change, where here the percentage change is between -8 and -10% - much bigger differences than those found in previous studies. This could be in part due to the use of the Shimmer device, as previous studies have made use of different HR measurement techniques.

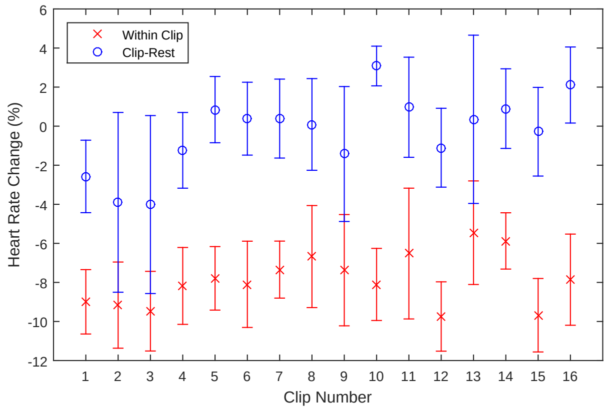

Figure 7: Mean BPM percentage change (and standard error) across all participants for each clip. Shown are the results for both the Clip-Rest and Within Clip comparisons.

3.3 Analysis Of Subjective And Heart Rate Data

Fig. 8 is a plot of the mean percentage change in BPM for both comparison conditions against the mean valence rating for each clip.

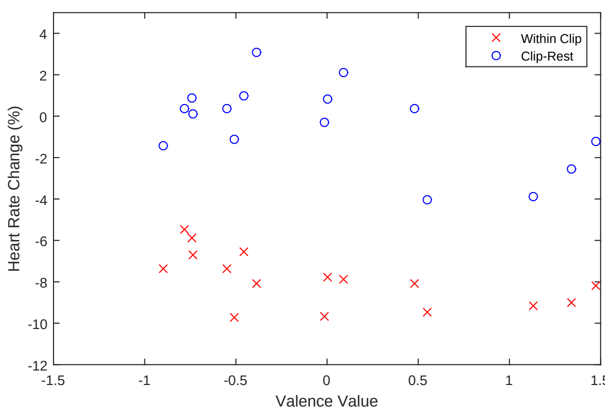

Figure 8: Plots of the percentage change in BPM (for both Clip-Rest and Within Clip comparisons) against mean valence ratings.

For both the Within Clip and Clip-Rest conditions one can see a slight negative correlation between the Valence and HR change. These correlations are statistically insignificant however ($r = -0.1673$, $p = 0.4608$ for the Clip-Rest condition, and $r = -0.0855$, $p = 0.6237$ for the Within Clip results), indicating no significant relationship between HR and Valence. This is somewhat contrary to the findings of Hume and Bradley [5, 7], although one must bear in mind that their results, whilst significant, only indicate a slight change in HR.

Fig. 8 is a plot of the mean percentage change in BPM for both comparison conditions against the mean arousal rating for each clip.

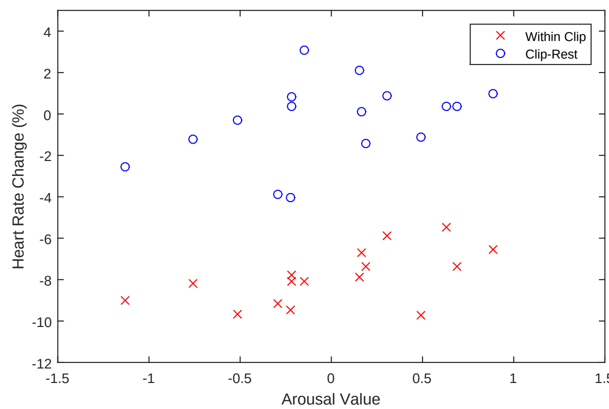

Figure 9: Plots of the percentage change in BPM (for both Clip-Rest and Within Clip comparisons) against mean arousal ratings.

Here there is a very slight positive correlation where increased arousal rating results in greater increase of smaller decrease in HR depending on the comparison method being considered. Again however these correlations are not statistically significant: $r = 0.0638$ and $p = 0.6128$ in the Clip-Rest case; and $rcorrcoefs = 0.1368$ ($p = 0.5735$). All of the above correlation values have been calculated in MATLAB according to Pearson's R [30].

4 Conclusion

This paper has presented primary analysis of HR data and Subjective assessment ratings obtained as part of a listening test presenting test participants with 60 30s-long soundscape recordings. The results show that the presentation of these soundscapes to the participants does have an effect on their heart rate, but correlational analysis has not shown this effect to relate to subjective scores of arousal and valence in a statistically significant way.

This indicates a need for further analysis of the collected data, (for example by the calculation of heart rate variability [31]) or the need to study relationship of heart rate to subjective assessment scales other than Valence and Arousal. Future work will consider both of the avenues, as well as the analysis of the GSR data recorded as part of this test, and the effect of the presentation of visuals on the perception of the soundscape recordings used here.

5 Acknowledgments

This project has been supported by an EPSRC doctoral training studentship.

References

[1] E. U., "Directive 2002/49/ec of the European Parliament and the council of 25 june 2002 relating to the assessment and management of environmental noise," *Official Journal of the European Communities*, vol. 189, no. 12, 2002.

[2] S. Harriet and D. Murphy, "Auralisation of an urban soundscape," *Acta Acustica united with Acustica*, vol. 101, no. 4, pp. 798–810, 2015.

[3] J. Kang and M. Zhang, "Semantic differential analysis of the soundscape in urban open public spaces," *Building and environment*, vol. 45, no. 1, pp. 150–157, 2010.

[4] M. Bradley and P. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[5] ——, "Affective reactions to acoustic stimuli," *Psychophysiology*, vol. 37, no. 02, pp. 204–215, 2000.

[6] P. Gomez and B. Danuser, "Affective and physiological responses to environmental noises and music," *International Journal of Psychophysiology*, vol. 53, no. 2, pp. 91–103, 2004.

[7] K. Hume and M. Ahtamad, "Physiological responses to and subjective estimates of soundscape elements," *Applied Acoustics*, vol. 74, no. 2, pp. 275–281, 2013.

[8] K. Hume, H. Barrett, A. Ip, T. McDonagh, W. Davies, M. Adams, N. Bruce, R. Cain, P. Jennings, G. Czanner *et al.*, "Physiological responses and subjective estimates of sounds: initial results of pilot study," *Proc. I. o. A.*, vol. 30, no. 2, 2008.

[9] B. K. Wiederhold, D. Jang, M. Kaneda, I. Cabral, Y. Lurie, T. May, I. Kim, M. D. Wiederhold, and S. Kim, "An investigation into physiological responses in virtual environments: an objective measurement of presence," *Towards cyberpsychology: Mind, cognitions and society in the internet age*, 2001.

[10] "Shimmer gsr+ module," 2015. [Online]. Available: http://www.shimmersensing.com/images/uploads/docs/GSR_Revision_1.2.pdf

[11] A. Léobon, "La qualification des ambiances sonores urbaines," *Natures-Sciences-Sociétés*, vol. 3, no. 1, pp. 26–41, 1995.

[12] S. Viollon, L. C., and C. Drake, "Influence of visual setting on sound ratings in an urban environment," *Applied Acoustics*, vol. 63, no. 5, pp. 493 – 511, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003682X01000536

[13] S. Harriet, "Application of auralisation and soundscape methodologies to environmental noise," Ph.D. dissertation, University of York, 2013.

[14] W. Yang and J. Kang, "Acoustic comfort and psychological adaptation as a guide for soundscape design in urban open public spaces," in *Proceedings of the 17th International Congress on Acoustics (ICA)*, 2001.

[15] L. Anderson, B. Mulligan, L. Goodman, and H. Regen, "Effects of sounds on preferences for outdoor settings," *Environment and Bevior*, vol. 15, no. 5, pp. 539–566, 1983.

[16] G. Watts and R. Pheasant, "Tranquillity in the scottish highlands and dartmoor national park–the importance of soundscapes and emotional factors," *Applied Acoustics*, vol. 89, pp. 297–305, 2015.

[17] R. Pheasant, K. Horoshenkov, G. Watts, and B. Barrett, "The acoustic and visual factors influencing the construction of tranquil space in urban and rural environments tranquil spaces-quiet places?" *The Journal of the Acoustical Society of America*, vol. 123, no. 3, pp. 1446–1457, 2008.

[18] R. Pheasant, M. Fisher, G. Watts, D. Whitaker, and K. Horoshenkov, "The importance of auditory-visual interaction in the construction of 'tranquil space'," *Journal of Environmental Psychology*, vol. 30, no. 4, pp. 501–509, 2010.

[19] R. Gifford and C. Ng, "The relative contribution of visual and auditory cues to environmental perception," *Journal of Environmental Psychology*, vol. 2, no. 4, pp. 275–284, 1982.

[20] Ö. Axelsson, "How to measure soundscape quality," in *Euronoise2015, Maastricht*, 2015.

[21] O. Rummukainen, J. Radun, T. Virtanen, and V. Pulkki, "Categorization of natural dynamic audiovisual scenes," *PloS one*, vol. 9, no. 5, p. e95848, 2014.

[22] C. Guastavino, B. Katz, J. Polack, D. Levitin, and D. Dubois, "Ecological validity of soundscape reproduction," *Acta Acustica united with Acustica*, vol. 91, no. 2, pp. 333–341, 2005.

[23] M. Bradley and P. J. Lang, *The International affective digitized sounds (IADS): stimuli, instruction manual and affective ratings.* NIMH Center for the Study of Emotion and Attention, 1999.

[24] M. Bradley, B. Cuthbert, and P. Lang, "Picture media and emotion: Effects of a sustained affective context," *Psychophysiology*, vol. 33, no. 6, pp. 662–670, 1996.

[25] J. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[26] R. Hull and A. Harvey, "Explaining the emotion people experience in suburban parks," *Environment and behavior*, vol. 21, no. 3, pp. 323–345, 1989.

[27] K. Hanyu, "Visual properties and affective appraisals in residential areas in daylight," *Journal of Environmental Psychology*, vol. 20, no. 3, pp. 273–284, 2000.

[28] S. Viollon and C. Lavandier, "Multidimensional assessment of the acoustic quality of urban environments," in *Conf. proceedings "Internoise", Nice, France, 27-30 Aug*, vol. 4, 2000, pp. 2279–2284.

[29] F. Stevens, D. Murphy, and S. Smith, "Soundscape preference rating using semantic differential pairs and the self-assessment manikin," in *SMC 2016*, 2016.

[30] J. G. Peatman, *Introduction to applied statistics.* Harper & Row, 1963.

[31] C. R. André, J.-J. Embrechts, J. G. Verly, M. Rébillat, and B. F. Katz, "Sound for 3d cinema and the sense of presence," 2012.