Using body language indicators for assessing the effects of soundscape quality on individuals

Harry J. Witchel[1], Lisa Lavia[2], Carina E. I. Westling[3], Aoife Healy[4], Rob Needham[4], Nachiappan Chockalingam[4]

[1] *Brighton and Sussex Medical School, Brighton BN1 9PX, UK, Email: h.witchel@bsms.ac.uk*
[2] *The Noise Abatement Society, Hove BN3 1HJ, UK, Email: lisa.lavia@noise-abatement.org*
[3] *School of Media Film and Music, University of Sussex, Falmer BN1 9RG, UK, Email: c.e.i.westling@sussex.ac.uk*
[4] *CSHER, Staffordshire University, Stoke-on-Trent University, UK, Email: n.chockalingam@staffs.ac.uk*

Abstract

"Sounding Brighton" is a collaborative project exploring practical approaches towards better soundscapes focusing on soundscape issues related to health, quality of life and restorative functions of the environment. The project is part of a citywide engagement process working to provide opportunities to demonstrate how an applied soundscape approach might: tackle conventional noise problems, contribute to local planning and improve the environment in areas including urban green spaces, the built environment and traffic noise. So far, a soundscape map of the city has been developed, and a public outreach exhibition and conferences have taken place. One preliminary, experimental soundscape intervention in night noise has been analysed.

This paper reports on further work to develop a better understanding of the effects of soundscapes on individual and community responses to soundscape through the use of body language indicators. Two-minute excerpts of aversive and preferred music were presented to 11 healthy volunteers in a motion-capture laboratory setting. Their responses were quantified computationally using motion-capture-derived parameters for position, absolute movement speed, and stillness. The prevalence of stillness of the head height (based on a 2 cm cut-off during 2-second sectors) was significantly lower when volunteers were exposed to unpleasant music compared to preferred music. This experiment provides proof in principle that changes in soundscape can be associated with subsequent, objective and statistically significant changes in body language that can be detected computationally.

Introduction

Sounding Brighton is a multidisciplinary approach to using a soundscape approach to approach noise and other problems on a large scale in the City of Brighton and Hove [1]. It involves a successful collaboration between the Noise Abatement Society (NAS, a UK-wide registered charity campaigning for improvements in the soundscape), the city council, and a range of scientists, acousticians and sound experts. The NAS along with Brighton & Hove City Council, Brighton & Hove Arts Commission and the EU COST Action TD0804 on the "Soundscapes of European Cities and Landscapes" co-commissioned an artwork, which was in effect a night noise intervention pilot based on a soundscape approach. On 29th October 2011, the project, known as 'West Street Story', created a 3-dimensional soundscape with ambient sounds mixed with relaxing music that was played live onto the main street (West Street) of the city's busiest clubbing and entertainment district, to see if the change in soundscape could improve crowd behavior and decrease anti-social behavior and violence. The preliminary results, based on comparing body language surrogate measures from a control night to the soundscape intervention night, were promising. The soundscape intervention was assocated with a statisically significant decrease in the percentage of territorial (quasi-aggressive) gestures and behaviours compared to non-territorial (open and friendly) behaviours [1]. The interpretation of the data is not without issue because body language metrics, while well-established when performed with investigator-interpreted scoring, is open to criticisms of potential bias; the use of computationally-based body language metrics would answer these criticisms, but such metrics are not currently available, especially when applied to outdoor crowds.

Research "in the wild" is the attempt to situate research in contexts where the conclusions are meant to be applied, and it involves cooperation of different stakeholder groups (e.g. campaigning NGOs) and aligning social identities [2]. However, "in the wild" contexts often interfere with experimenter control of complicating variables, making validation of the methods difficult. Soundscape is fundamentally situated in context, and one key issue with music is that it is central to social identity and territory [3].

The use of postural changes to assay for engagement of seated, healthy volunteers with computer-based stimuli has been validated previously [4-6]; however, these studies did not involve music (without video) as a stimulus, nor did they derive postural measurements that could be used with filmed data.

Methods

Volunteers

Eleven healthy, English-speaking volunteers were recruited for a study on psychobiology and non-verbal behaviour. The study was approved by the local ethics committee. The volunteers were 9 males and two females. The age range was 19 to 62.

Protocol

After being briefed as to the nature of the study, participants were seated in a standard armless "reception room" chair at a desk with a 21 inch (diagonal) monitor. The monitor was raised such that the centre of the screen was at the eye level of the volunteer. Volunteers were allowed to adjust the seat position for comfort. After completing initial background questionnaires, participants experienced audiovisual stimuli, each lasting 180 seconds, and then rated the experience via a set of 10-cm visual analogue scales (VAS). All experimental stimuli were presented in a counterbalanced order. All members of the scientific team left the room before each stimulus, such that the volunteer was alone in the room as they experienced the stimulus. Before the two experimental stimuli, each volunteer was habituated to the protocol with two stimuli that were never part of the analysis (the training stimuli); participants were not informed that the training stimuli were not part of the analysis, so to the participants there was no difference in protocol between the training stimuli and the experimental stimuli. At the beginning of the experiment, each participant was allowed to adjust the volume control of the sound system to a level they found comfortable, and they were encouraged to pick a level that was slightly quieter just for safety; participants were told that they could adjust the volume at any time if they found the sounds too loud.

Stimuli

The experimental stimuli were 180 seconds of two musical excerpts. The aversive excerpt was a piece of solo violin music played incompetently (VIO). The preferred excerpt was user-selected; we asked each volunteer to select a piece of music that they liked (FAV, e.g. their favourite piece of music), preferably a piece that was up-tempo and that they might want to dance to. The training stimuli were structured as follows: 45 seconds of white noise and "television snow" to establish a baseline, 5 second timing signal (black screen with a green flash and clicking sound), and then the main stimulus (lasting 130 seconds). The positive training stimulus was a web-classic video-lecture from the Royal Society of Art in which a hand moving in fast motion draws a cartoon illustrating a lecture by neuroscientist Dan Pink, who is explaining how financial rewards paradoxically diminish performance of healthy volunteers on any task requiring rudimentary mental skill (RSA) [7]. The aversive stimulus was that after the timing signal, the screen went blank and there was no sound for 130 seconds (i.e. without explanation, so volunteers were left alone staring at a blank screen -- BSc).

Measurements

Subjective Responses

Before all the VAS measurements, after each stimulus the volunteer was asked for a few words to describe how they felt. The adjectives for the VAS were: "I felt interested", "I felt bored", "I wanted to see/hear it more", "I wanted it to end earlier", "I was engrossed by the experience", "I put up with it". The anchors for the VAS were 0 = not at all, and 100 = extremely. The University of Florida's Self Assessment Manikin (SAM) was also used.

Motion Capture

Motion capture was performed by video analysis (Kinovea) of video from a lateral aspect (BSMS) or by a Vicon opto-electronic 8 camera-mocap system (Staffordshire). We have previously shown that these two technologies produce comparable results for head attitude and for small translational movements in the sagittal plane [8]. Passive reflective markers were positioned on the head, badge of the deltoid, and middle of the outer thigh. Head markers were placed on the outer canthus of the eye and on the ear behind the tragus (Kinovea) or on a head band as a set of four (left front head, right front head, left back head, right back head); the Vicon movements were corrected for position and angle based on a frame at the beginning of the experiment for each volunteer. The outcome parameters were head pitch (relative to floor), front head marker from screen, front head marker from floor, deltoid marker from screen, deltoid marker from floor, thigh marker from screen, thigh marker from floor. The videos were made by a Canon 850 miniDV recorder and captured by Kinovea at 25 Hz. Vicon captured data at 50 Hz, which was down-sampled by Matlab to 25 Hz.

Statistics and analysis

All statistics reported here are paired T tests calculated in Matlab. For motion capture 80 seconds of each stimulus was used: from the 75th second to the 155th second. This period was chosen to allow participants to settle in to each stimulus and to habituate to being alone; it also avoided any potential artefacts arising from the re-entry of the experimenters into the room. Positions were calculated as the mean of each uni-dimensional parameter.

Motion and Stillness Parameters

For all motion parameters, the time series data was low-pass filtered through a mean filter with a width of 7 time points (i.e. 28 milliseconds). A "speed" parameter was derived by adding up the absolute value of the differences between each successive time point and normalising by total time. In addition a set of "stillness" parameters were calculated to estimate large movements. The rationale for two different calculations is that there is fundamental difference between a person making tiny rocking movements throughout the stimulus compared to a person sitting absolutely still throughout the stimulus except for one second when the person stands up and sits down again – although these two behaviours could potentially result in identical speed measurements. Furthermore, we have previously shown that Kinovea measurements are subject to small "jittery" movements that add to the speed calculation (even with the low-pass filter), while these artefacts are screened out in the stillness calculations.

The stillness calculation units are in the percentage of time that the volunteer's total positional change exceeded an arbitrary cut-off point. These calculations were made in a method analogous to the successful calculation of human motion energy analysis (MEA) used to estimate movement on video-analysis of humans without markers. In brief, the analysis region was divided into two-second sectors, and the absolute value of the range (i.e. the maximum minus the minimum position) for each sector was compared to the arbitrary cut-off (e.g. 2 cm). The stillness (i.e. lack of large-motion) calculation was the percentage of sectors where the volunteer's movement exceeded the cut-off value. A number of cut-off values were selected (in cm 0.5, 1, 2, 5, 10, 15; in angle degrees: 0.5, 5, 10, 15, 20, 25).

Results

Subjective responses

Training Stimuli

Two training stimuli were presented at the beginning of the protocol, both to habituate the volunteer to the process, and also to give the volunteer a sense of the range of how interested or how bored they might feel when presented with this kind of audio-visual stimulus. The VAS "interested" response to the interesting lecture from the Royal Society of Art's animate series (RSA) was 58.8 ± 6.6 (mean ± standard error of the mean) and to 2 minutes of watching a black screen (BSc) was 3.5 ± 2.4; the difference was statistically significant (paired T test, $P < 0.001$). The VAS "bored" response to RSA was 37.1 ± 7.6, and to BSc it was 93.8 ± 2.9 ($P < 0.001$).

Experimental Stimuli

A comparison between the participants' subjective responses to their favourite music (FAV) compared to the incompetently played violin music (VIO) is shown in Figure 1.

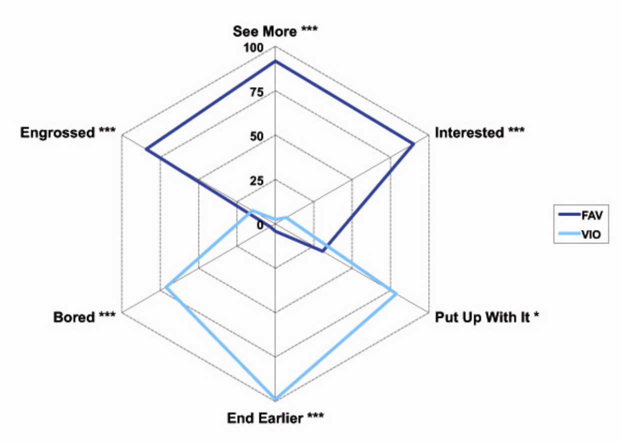

Figure 1: Mean results of subjective Visual Analogue Scale (VAS) in response to the participants' favourite music (FAV – dark) compared to the aversive violin music (VIO - light). Centre is 0 = "not at all", and outermost hexagon is 100 = "extremely". *** $P < 0.001$; * $P < 0.05$

The differences in VAS ratings between the two stimuli were all statistically significant; the differences (as expected) were extreme – the P value for the paired T test for "I wanted to see/hear more" was 1.58×10^{-10}.

Motion Capture

Although nonverbal cues given off during aversive vs. pleasurable stimuli may seem obvious to casual observers (especially when viewing facial expressions), objective measurements based on computational analysis of postural cues is nontrivial because some individuals make many spontaneous movements with no obvious trigger while other individuals make almost no movements at all while being filmed. Although the mean of many parameters were obviously different, the one parameter that reached a statistically significant difference was front of head height from the floor. A representative pair of time series for one volunteer comparing VIO vs. FAV is shown in figure 2.

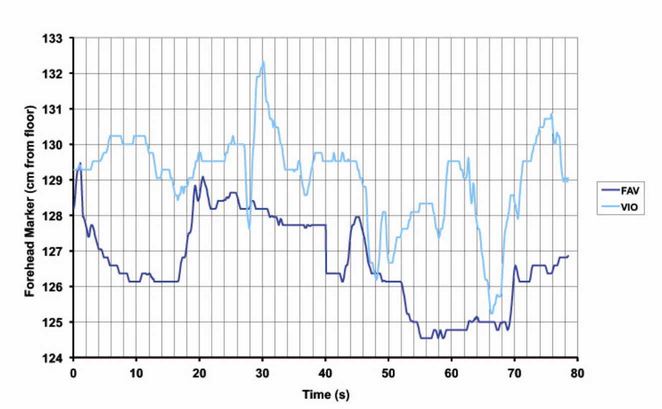

Figure 2: Two representative time series from the same volunteer (Y019) showing the height of the front of head marker from the floor during FAV (dark) and VIO (light). Vertical gridlines show sectors (lasting 2 seconds) for analysis of stillness. Sampling artefacts occur at 40 seconds.

Representative summary parameters derived from these two time series in Figure 2 are shown in Table 1.

Stimulus	Parameter (Front of Head Height)		
	Mean	Speed	Stillness 2
FAV	127.4	96.1	2.4
VIO	130.5	214.8	11.9

Table 1: Parameters describing the representative data in Figure 2. The mean height is the distance of the marker at the front of the head from the floor in cm. The average speed is the sum of the absolute values of all movements (in micrometers) per second. The stillness (measured in percent) is the number of sectors where the total range of positions is ≥ 2 cm, divided by the number of sectors.

In this example there is a difference in mean height (which is idiosyncratic, i.e. not generally true for all volunteers), a large difference in absolute speed (which is true for most volunteers, but does not reach statistical significance), and a difference in stillness with a sector cut-off of 2 cm. The large movements that the stillness parameter detects (e.g. the large upward movement of the head marker seen in Figure 2 in the VIO time series at $28 \leq \text{time} \leq 30$) differ significantly in their prevalence for all volunteers ($P < 0.05$). A summary of the difference in this stillness parameter is shown in Figure 3; note that for a number of volunteers the percent of time-sectors above the cut-off is zero for both music excerpts (i.e. the volunteers were always sitting quite still).

Discussion

This data provides proof in principle that changes in soundscape can be associated with objective changes in body language parameters. This observation in the laboratory is a

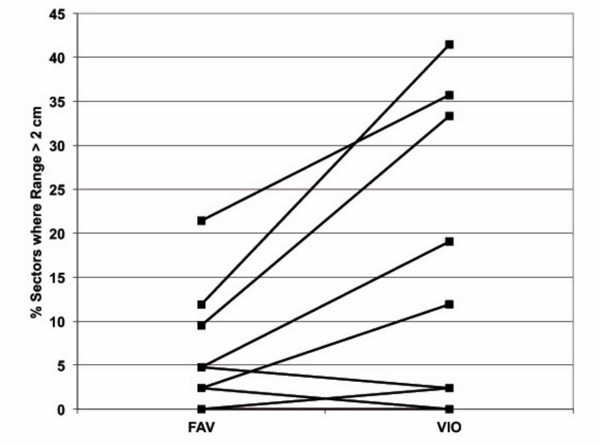

Figure 3: Comparison of mean prevalence of stillness (cut-off 2 cm per 2-second sector) for each volunteer during their favourite music (FAV – left) vs. the aversive violin music (VIO – right), shown by volunteer.

useful first step in validating the use of body language surrogates for assessing soundscape interventions "in the wild".

Limitations on generalising these results

There are many limitations to applying the results of this laboratory experiment to the target situation of making soundscape interventions "in the wild". The experimental volunteers in this study were recruited and paid to take part in a laboratory experiment. The musical excerpts were discrete 180-second stimuli punctuated by interventions with the experimental team (as opposed to a continuous music stimulus such as a playlist). The volunteers were seated. The volunteers willingly subjected themselves to all the stimuli, despite some of the stimuli being incredibly boring or even aversive; none of the volunteers ever left their chair, despite being alone and in a position to get up. The volunteers were facing a blank computer screen while listening to music. The volunteers were alone.

None of the above features of the experiment would be true (or desirable) for people walking through a space with a soundscape intervention. However, this highly controlled experiment shows that even people who know they are being filmed make small (possibly subconscious) postural movements in response to a musical intervention, and that these changes can be detected at a level of statistical significance when testing only a small sample of people.

Limitations of this experiment

This experiment has a range of limitations. Only two musical interventions were tested, and these were at the extreme end of valence**.** Both stimuli were relatively arousing (VIO being irritating), and the self-assessment manikin (SAM) ratings of arousal were not different ($P > 0.05$). However, SAM was sensitve, as both the differences for the mean results for SAM were statistically significant for the indepedence ratings ($P < 0.001$) and for valence ratings ($P < 0.001$).

Stimulus	Mean Rating SAM (1-9)		
	Dependence	Valence	Arousal
FAV	1.6 ± 0.3	8.3 ± 0.3	6.3 ± 0.7
VIO	4.4 ± 1.0	3.0 ± 0.7	4.1 ± 0.8

Table 2: Mean results of the subjective responses using the self-assessment manikin (SAM). Dependence represents the independence-dependence continuum (independent is low). Valence represents the sad-content axis (sadness is low). Arousal represents the quiet-active continuum.

Unlike this experiment, the soundscape interventions for calming crowds are not meant to elicit high states of arousal; they would be geared toward eliciting relaxed or curious states.

Conclusion: This experiment provides proof in principle that changes in soundscape can be associated with subsequent, objective and statistically significant changes in body language that can be detected computationally. Still, much laboratory work needs to be done to validate methodologically that changes in body language surrogates can be used to assess the effects of "in the wild" soundscape interventions.

References

[1] Lavia, L, Easteal, M, Close, D, Witchel, HJ, Axelsson, O, Ware, M, Dixon, M. Sounding Brighton: practical approaches towards better Soundscapes. Internoise 2012 New York, 2012.

[2] Callon, M, Rabeharisoa, V. Research "in the wild" and the shaping of new social identities. Technology in Society. **25**, (2003), 193-204.

[3] Witchel, HJ. You are what you hear: how music and territory make us who we are. 2011. New York, Algora Publishing.

[4] Balaban, C.D. et al.,. Postural control as a probe for cognitive state: exploiting human information processing to enhance performance. International Journal of Human-Computer Interaction, **17** (2004), 275-286.

[5] Mota, S, Picard, RW. Automated Posture Analysis for Detecting Learner's Interest Level. 2003 Conference on Computer Vision and Pattern Recognition Workshop, **5** (2003), 49-49.

[6] Kleinsmith, A., Bianchi-Berthouze, N., & Steed, A.. Automatic recognition of non-acted affective postures. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, **41** (2011), 1027-1038.

[7] The Royal Society for the Arts. Animate. 2010. URL: `http://www.thersa.org/events/rsaanimate/animate/rsa-animate-drive`

[8] Witchel, HJ, Healy, A, Needham, R, Westling, C, Chockalingam, N. Comparing four technologies for measuring postural micromovements during monitor engagement. European Conference on Cognitive Ergonomics, 2012.