CONTINUOUS EVALUATIVE AND PUPIL DILATION RESPONSE TO SOUNDSCAPES

S. Graetzer[1] **A. Landowska**[1] **L. Harris**[1]
T. J. Cox[1] **W. J. Davies**[1]
[1]Acoustics Research Group, School of Science, Engineering and Environment,
University of Salford, United Kingdom
s.n.graetzer@salford.ac.uk, W.Davies@salford.ac.uk

ABSTRACT

We investigated human response to soundscapes using a continuous second-by-second rating of soundscapes and a more conventional overall rating of each sample at the end of each audition. In this work, our primary aim was to explore what continuous ratings tell us about soundscape perception. Our secondary aim was to understand how pupil dilation response (*i.e.*, changes in pupil size) relates to such perception. We used ambisonic soundscapes simulated in a laboratory environment. During playback, listeners were asked to perform a continuous evaluation on pleasantness and eventfulness dimensions.

A comparison of the results of the two rating methods (continuous and overall) indicates that the first can be used reliably for soundscape evaluation. We found a strong effect of soundscape category – human, nature or manmade industrial or domestic – on the pleasantness and eventfulness ratings and on the first two principal components derived from the overall ratings. These components related to pleasantness-eventfulness dimensions. In addition, we observed some effects of category on pupil size, which were broadly consistent with perceptual ratings. The perceived differences between soundscape categories were consistent with findings in the literature of greater pleasantness ratings for natural sounds than mechanical or industrial sounds.

1. INTRODUCTION

Soundscapes can be defined as acoustic environments as perceived or experienced and understood by a person or people, in context [1]. They consist of multiple sound sources and multiple sounds that may occur simultaneously or separately in time. Typically, human emotional responses are modelled as two-dimensional. For example, in Russell's circumplex model of affect [2], emotions evoked by an environment are represented in a two-dimensional space with axes valence (pleasure-displeasure) and arousal (arousal-sleepiness or active-passive). Västfjäll [3] argued that this model can be applied to the context of emotions evoked by particular sounds. Axelsson *et al.* [4] redefined the dimensions as 'pleasantness' and 'eventfulness', the latter on a continuous scale of eventfulness-uneventfulness. Davies *et al.* [5] suggest that the weight of evidence in the literature is now sufficient for two dimensions of pleasantness/calmness and eventfulness/activity to be regarded as the standard model for the perceptual dimensions of soundscapes.

Soundscape category, like other informational properties of soundscapes, significantly contributes to soundscape perception (e.g., [6]). Previous work by Bones *et al.* [7] indicated that isolated sounds were classified by listeners in a free sorting task into three high level categories: *people*, which was subdivided into *voices* and *music*, *nature*, which was subdivided into *animals* and *weather*, and *manmade*, which was subdivided into *industrial* (mechanical and transport) and *household/domestic*.

Pupil size has been found to change spontaneously during visual and audio stimulus presentation to reflect arousal and cognitive processes such as decision making, problem solving and language comprehension (e.g., [8]). Previous studies of pupil dilation response to non-speech and non-music auditory stimuli suggests that the pupil dilates in response to louder sounds, surprising sounds *vs.* unsurprising sounds, sounds with high positive/negative valence *vs.* neutral sounds, and for active *vs.* passive listening (e.g., [8]). A review of pupil dilation response to auditory stimuli is provided by Zekveld *et al.* [9].

2. AIMS AND RESEARCH QUESTIONS

In the study reported in this paper, listeners rated thirty second excerpts of soundscapes of various categories using a two-dimensional pleasantness-eventfulness model. Soundscapes were reproduced by a 16-channel loudspeaker array. Listener pupil dilation response was measured during soundscape exposure. The research questions included the following: firstly, do soundscapes associated with the categories human, nature, manmade domestic and manmade industrial differ from one another according to pleasantness (valence) and eventfulness (arousal) both on a continuous evaluation and on a more conventional soundscape evaluation performed after exposure, *i.e.*, semantic differential scales, to the stimulus? Secondly, how do real-time pleasantness-eventfulness ratings compare to semantic differential scale ratings? Do the scale ratings have underlying dimensions of pleasantness and eventfulness on a principal component analysis? Thirdly, is there an effect of soundscape category on pupil dilation response and, by ex-

tension, on arousal?

The primary aim was to investigate how soundscapes are perceived by human listeners on pleasantness and eventfulness dimensions on a continuous basis, and how this evaluation compares with a conventional semantic differential scale evaluation. A secondary aim was to understand how pupil dilation response relates to soundscape perception. Continuous subjective evaluation constitutes a new method for soundscape research. Furthermore, the use of pupillometry to measure arousal is a new method in the context of spatial audio reproduction of soundscapes.

3. METHOD

In the study reported in this paper, forty subjects listened to thirty second soundscape excerpts assigned to four categories – human, nature, manmade industrial and manmade domestic – while interacting with a custom MATLAB Graphical User Interface (GUI) [10]. Listeners subsequently completed a nine attribute semantic differential scale evaluation. Pupil size was measured during the soundscape playback and for five seconds immediately preceding playback (baseline), during which time listeners looked at a fixation point. During soundscape playback, listeners completed a continuous pleasantness-eventfulness evaluation via the GUI, while directing their gaze towards the interface.

The experiment took approximately 1.5 hours, including set up. The luminance in the room was kept constant at 2.7 EV to ensure that pupil size was not affected by changes in ambient light intensity. A light grey background was used in the GUI.

3.1 Participants

All listeners had normal hearing and normal or corrected-to-normal vision by their own report. Nineteen male and twenty female listeners, and one listener identifying as other, were recruited (mean = 32.6 years, std = 6.8 years). Of these, 33 were right-handed and 7, left-handed; 34 were native English speakers, and 6 were highly fluent. All participants were dominant or native English speakers who reported that they did not use autonomic nervous system blockers, and had not used stimulants, including caffeine, within the 5 hours preceding the session.

3.2 Soundscape recordings

First order ambisonic reproduction was used after Davies *et al.* [5]. Recordings were made using a Soundfield ST450 microphone recording to ambisonic B-format. The microphone was connected to a Soundfield ST450 MKII control unit with preamplifier connected to a Zoom H6 portable recorder in ambisonic B-format with a sampling rate of 44.1 kHz (16 bits). The soundscapes were selected so as to be dominated by three principal categories of sounds: human or people, nature and manmade where manmade was divided into industrial and domestic. In MATLAB, the ambisonic B-format signals were decoded to first-order ambisonics for a 16-channel loudspeaker array. The 16-channel Genelec 8030A loudspeaker array in the audio booth consisted of eight speakers in the horizontal plane (positioned at azimuths +0°, +45°, +90°, +135°, ± 180°, -45°, -90°, -135°), 4 at +39° elevation (positioned at azimuths +45°, +90°, -45°, -135°) and 4 at -39° elevation (positioned at azimuths +45°, +135°, -45°, -135°). The speakers had a radius of 1.26 m at 0° elevation and of 1.54 m at ± 39° elevation.

The recordings were segmented into thirty second excerpts. Ten excerpts were chosen at random from each category, with the constraint that at least one excerpt must be sampled from each recording. Excerpts were equalised in loudness.

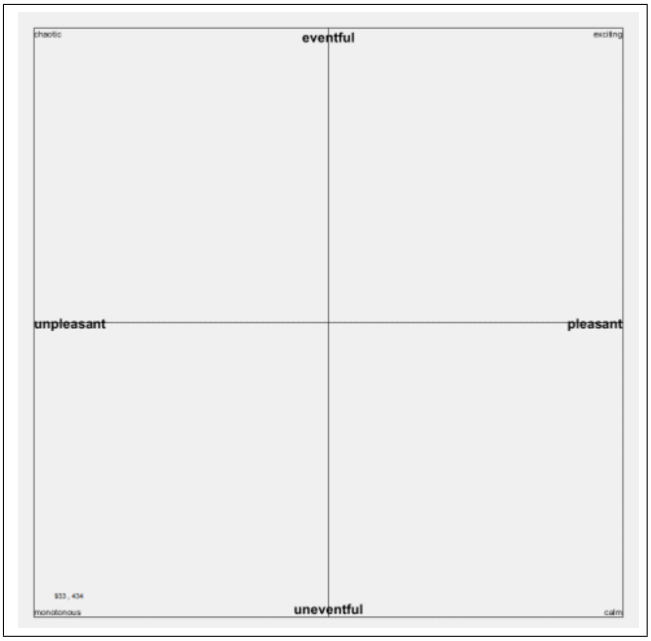

Figure 1. Two-dimensional display for real-time soundscape evaluation.

3.3 Playback

The experiment took place in an audio booth in which the background noise level at the listener position was 26 dB L_{Aeq} and the reverberation time was approximately 0.1 s. The loudspeaker array was first calibrated such that each loudspeaker reproduced a pink noise signal at 76 dB(A) (slow integration time) at the listening position. The loudspeakers were connected via an RME Digital/Analog Interface M-32 DA and an RME Hammerfall DSP soundcard to a PC running the GUI. The GUI was implemented such that the soundscape excerpts were reproduced at a level of 65 dB L_{Aeq} in the position of the listener, as measured with a Brüel & Kjær (B&K) Head and Torso Simulator (HaTS) with Type 4190 microphones, connected to a B&K data acquisition unit Type 3560-B-130, connected via LAN to a PC running Pulse LabShop.

Real-time pleasantness and eventfulness ratings were measured during soundscape playback via mouse movements within a computer representation of a two-

dimensional emotional space (shown in Fig. 1) after the method used previously for music and speech by Coutinho and colleagues [11]. The labelling of the dimensions and the descriptors at the corners of the space – 'exciting', 'calm', 'monotonous' and 'chaotic' – derive from Axelsson and colleagues' pleasantness-eventfulness model (see [4], Figure 4, p. 2844). The data were collected with a 10 Hz sampling rate. The forty soundscape excerpts were presented in a pseudo-random order to each listener.

3.4 Tasks

As part of the training stage, participants were given anchor stimuli for each of the four extremes of the two-dimensional space and given the opportunity to practice. Participants were shown the interface and instructed to "rate the soundscapes according to your perception of pleasantness and eventfulness. You can move the mouse anywhere you like within the four quadrants to reflect your perception. If your perception changes, move the mouse to reflect this change." The initial cursor position was at the centre of the four quadrants. After listening to each soundscape excerpt, the listeners were asked to rate the entire excerpt on nine visual analogue scales from 0, 'not at all', to 100, 'extremely'. The scale attributes were pleasant, chaotic, vibrant, uneventful, calm, annoying, eventful, monotonous, and unpleasant [12]. After listening to the soundscape excerpt, the listeners were asked to rate the entire excerpt.

3.5 Pupillometry

The Pupil Labs [13] head-mounted, video-based eye-tracking device used in this study comprised a high-speed optical camera and two global shutter infra-red (IR) cameras with IR illumination. These had a resolution of 200 x 200 pixels at 124 frames per second. Data were captured and initially processed via Pupil Capture and Pupil Player v. 1.12-17 [13].

3.6 Signal processing

Non-pupillometric signals were processed with MATLAB v2019b [10]. Post-processing and statistical analysis was conducted in MATLAB and R v. 3.5.3 [14].

Pupillometric data were captured and processed using Pupil Labs (Germany) Pupil Capture software v. 1.12-17 [13] and post-processed in MATLAB using the Kret and Sjak-Shie [15] method. After processing, signals were normalised by subtracting the baseline from the PDR on a trial-by-trial basis. A Generalised Additive Mixture Model (GAMM [16]) was used for statistical analysis.

4. RESULTS

4.1 Continuous ratings

Continuous ratings indicated that median pleasantness was highest for nature, lowest for the manmade categories, and intermediate for the human category (Fig. 2). Eventfulness tended to be higher for the human category than the other categories. Nature was primarily associated with the pleasant eventful quadrant, while manmade domestic was primarily associated with the unpleasant eventful quadrant.

A mixed-effects linear regression model with response variable, pleasantness, predictor variables, time (s), category (with reference level, human), and an interaction of time and category, and random effects listener and stimulus, indicated that nature was associated with higher pleasantness than human ($B = 0.22, SE = 0.05, t = 4.58, p < 0.0001$), and that there was a main effect of time, such that pleasantness decreased very slightly as time passed ($B = -0.0005, SE = 0, t = -8.67, p < 0.0001$). There was an interaction of category and time such that the nature category was associated with an increase in pleasantness over time relative to the human category, while the manmade categories were associated with a decrease, with the difference emerging within the first five seconds (nature∼human, $B = 0.012, SE = 0.0, t = 76.85, p < 0.0001$; industrial∼human, $B = -0.01, SE = 0.0, t = -47.46, p < 0.0001$; domestic∼human, $B = -0.01, SE = 0.0, t = -56.45, p < 0.0001$). $R^2 = 0.47$.

A mixed-effects linear regression model with the same fixed and random effects for real-time ratings of eventfulness indicated that there were main effects of time such that eventfulness increased slightly as time passed, as might be expected as events accumulate ($B = 0.01, SE = 0, t = 49.77, p < 0.0001$). There was an interaction of category and time such that the human category was associated with higher eventfulness ratings than the other categories, with the difference emerging within the first five seconds (nature∼human, $B = -0.01, SE = 0.0, t = -30.23, p < 0.0001$; industrial∼human, $B = -0.01, SE = 0.0, t = -43.07, p < 0.0001$; domestic∼human, $B = -0.005, SE = 0.0, t = -7.21, p < 0.0001$). $R^2 = 0.25$. In both models, the manmade categories did not differ from one another.

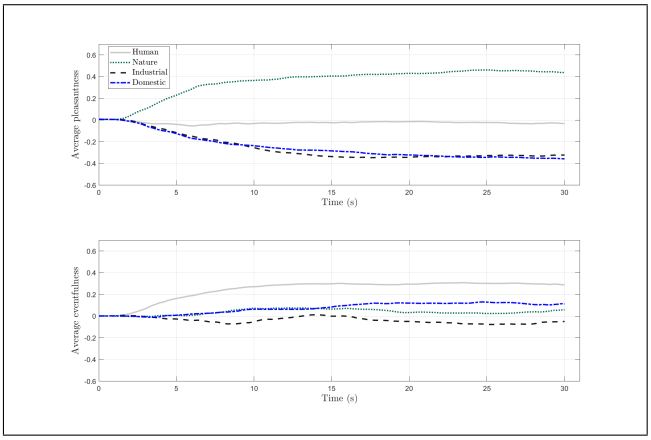

Figure 2. Trajectories of real-time pleasantness (upper panel) and eventfulness (lower panel) ratings by category, averaged over listeners.

4.2 Conventional post-stimulus ratings

Immediately after playback, soundscapes were evaluated on nine attributes: pleasant, chaotic, vibrant, uneventful, calm, annoying, eventful, monotonous, and unpleasant (Fig. 3). The nature category was associated with greater scores on pleasant and calm attributes, and lower scores on chaotic, annoying, monotonous and unpleasant attributes than other categories. While on many attributes the manmade subcategories did not differ markedly, the industrial category was rated as being less eventful than all other categories, and the domestic category was rated as being more vibrant and less monotonous than the industrial category.

For the 40 soundscape excerpts, the first two principal components explained 43 and 29% of the variability (72% in total). PC1 was inversely related to pleasantness and PC2 to eventfulness. The results indicated that attributes unpleasant and annoying are synonymous in the context of the study. PC3 explained only 7.5% of the variance, and PC4, 5.8%.

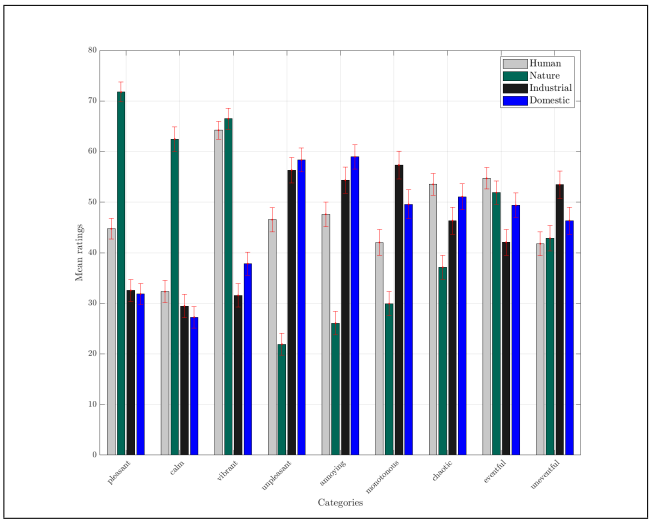

Figure 3. Bar plot of mean ratings on nine attributes by category with standard error bars (95% confidence intervals).

Linear mixed-effects models of PCs 1 and 2 were run with category as a fixed predictor (reference level: human) and ID and stimulus as random effects. There was an effect of category on pleasantness (likelihood ratio test statistic, $lrt = 75.49, p < 0.0001$): the human category was associated with lower pleasantness than the nature category but higher pleasantness than the manmade categories. The nature category was associated with higher pleasantness than the manmade categories, which did not differ from one another. There was an effect of category on PC2, or eventfulness ($lrt = 8.20, p < 0.05$): the human category was associated with higher eventfulness than the manmade industrial category. The nature and domestic categories were similarly eventful.

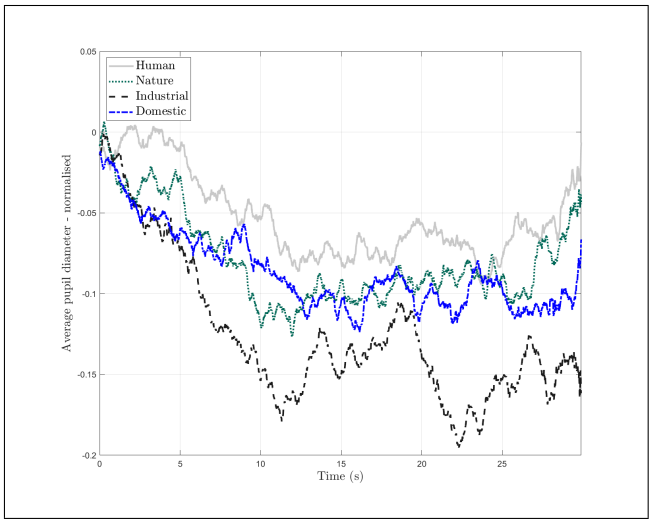

Figure 4. Pupil dilation response over time per soundscape category.

4.3 Pupil dilation response

Pupil dilation response over time is shown per soundscape category in Fig. 4. The human category is associated with a slight constriction relative to baseline but the least average constriction, with early peaks. The nature category is associated with pupil constriction relative to baseline, which suggests a decrease in arousal. The manmade industrial category is associated with the greatest constriction relative to baseline. This is consistent with the finding that the industrial stimuli were rated as less eventful than others. A GAMM [16] indicated that the human category was associated with greater pupil dilation relative to baseline than the industrial category ($B = -0.05, SE = 0.02, t = -2.30, p < 0.05$), and there was a tendency for the human category to be associated with greater pupil dilation relative to baseline than the domestic category ($B = -0.05, SE = 0.02, t = -1.95, p = 0.051$). However, there was no reliable difference between the human and nature categories ($p = 0.2$). The model explained 84.5% of the deviance.

5. DISCUSSION

In this study, it was confirmed that soundscapes of the categories human, nature and manmade differ from one another in perceived pleasantness and eventfulness on a real-time, continuous evaluation, on a conventional post-stimulus semantic differential scale evaluation, and on a pupil dilation response evaluation.

5.1 Continuous ratings

The nature category was associated with higher real-time pleasantness ratings than the human category across the time course, while the manmade categories tended to be associated with lower pleasantness ratings. Pleasantness ratings decreased slightly as time passed, perhaps due to de-sensitisation. Within the first two seconds of playback, the difference in pleasantness between the nature and

other categories emerged. The human category tended to be rated as more eventful than the other categories, perhaps partly due to the number of talkers in the 'crowd' scenes. The lack of differentiation between the manmade categories on both pleasantness and eventfulness supports the claim of Bones *et al.* [7] that these are subordinate categories of a higher order 'manmade' category. The findings that nature sounds are more pleasant than manmade/technological sounds and that human sounds, such as voices, tend to be perceived as more pleasant than manmade sounds, but not as pleasant as nature sounds, are consistent with the literature (e.g., [6]).

A model of real-time eventfulness ratings with predictors soundscape category and time (s) explained a smaller proportion of variance (25%) than a model of real-time pleasantness ratings with the same predictors (47%). This could be primarily because perception of eventfulness is more explained than perception of pleasantness by personal factors that were not modelled, such as personal preferences and sensibility, and previous experiences and expectations. Secondarily, it could be that the concept of 'eventfulness' is more ambiguous than the concept of pleasantness. If so, in future, a less ambiguous term might be used.

5.2 Conventional post-stimulus ratings

Soundscapes were evaluated after playback on nine attributes: pleasant, chaotic, vibrant, uneventful, calm, annoying, eventful, monotonous, and unpleasant. Overall, the results indicated that unpleasant and annoying are (nearly) synonymous in the present context, which is consistent with the recommendations in PD ISO/TS 12913 2:2018 [12]. 72% of the variability was explained by two principal components corresponding inversely to pleasantness and eventfulness. It was therefore confirmed that the scale ratings have underlying dimensions of pleasantness and eventfulness on a principal component analysis. Soundscape category was found to predict both of these principal components.

Ratings on the pleasant attribute were highest for the nature category, lowest for the manmade categories, and intermediate for the human category. The nature category was rated as least chaotic, monotonous and unpleasant, and most calm. On many attributes, the manmade categories were not well distinguished. The finding that the human category tended to be associated with higher eventfulness than the manmade industrial category is likely because the human soundscapes involved crowd scenes, while the industrial soundscapes involved fewer sound sources, such as a small number of people working in an acoustic laboratory with intermittent ventilation.

5.3 Pupil dilation response

There was some evidence of an effect of soundscape category on pupil dilation response. The nature category was associated with pupil constriction relative to baseline, suggesting decreased arousal. This is consistent with ratings on the calm and chaotic attributes. The human category was associated with only a small amount of constriction. There was little evidence of a difference between the manmade categories, as was the case on several semantic differential scale attributes. Overall, the results were consistent with the ratings on the eventful attribute, which supports the argument that eventfulness is closely related to arousal, which is reflected in pupil dilation.

6. CONCLUSIONS

The broad aim of the present study was to investigate human perception of soundscapes, including emotional response, while controlling for the overall soundscape loudness. Evidence has been provided that the sound categories human, nature and manmade, as proposed by Bones *et al.* [7], hold for complex audio in the form of soundscapes in addition to isolated sounds. These categories can be distinguished on the basis of pleasantness and eventfulness ratings. The results of a real-time continuous evaluation of soundscapes on dimensions pleasantness and eventfulness were consistent with the results of a principal components analysis of semantic differential scales derived from PD ISO/TS 12913-2 [12]. The finding of least pupil dilation relative to baseline for the industrial category *vs.* other categories is consistent with the rating of these stimuli as least eventful.

7. REFERENCES

[1] *BS EN ISO 12913-1 Acoustics — Soundscape Part 1: Definition and conceptual framework*. London, United Kingdom: British Standards Institution, 2014.

[2] J. A. Russell, "Circumplex model of affect," *J. Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[3] D. Västfjäll, "The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments," *Cyberpsychology and Behavior*, vol. 6, no. 2, pp. 181–188, 2003.

[4] O. Axelsson, M. E. Nilsson, and B. Berglund, "A principal components model of soundscape perception," *J. Acoust. Soc. Am.*, vol. 128, no. 5, pp. 2836–2846, 2010.

[5] W. J. Davies, N. S. Bruce, and J. E. Murphy, "Soundscape reproduction and synthesis," *Acta Acust united Ac*, vol. 100, no. 2, pp. 285–292, 2014.

[6] C. Guastavino, "The ideal urban soundscape: Investigating the sound quality of french cities," *Acta Acust united Ac*, vol. 92, no. 6, pp. 945–951, 2006.

[7] O. Bones, T. J. Cox, and W. J. Davies, "Sound categories: Category formation and evidence-based taxonomies," *Frontiers in Psychology*, vol. 9, no. July, pp. 1–17, 2018.

[8] T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," *Int. J. Human Computer Studies*, vol. 59, no. 1-2, pp. 185–198, 2003.

[9] A. A. Zekveld, T. Koelewijn, and S. E. Kramer, "The pupil dilation response to auditory stimuli: Current state of knowledge," *Trends in Hearing*, vol. 22, pp. 1–25, 2018.

[10] *MATLAB Version 2019b*. Natick, Massachusetts: The MathWorks Inc., 2019.

[11] E. Coutinho and N. Dibben, "Psychoacoustic cues to emotion in speech prosody and music," *Cognition and Emotion*, vol. 27, no. 4, pp. 658–684, 2013.

[12] *PD ISO/TS 12913-2 12913-1 Acoustics — Soundscape Part 2: Data collection and reporting requirements*. London, United Kingdom: British Standards Institution, 2018.

[13] M. Kassner, W. Patera, and A. Bulling, "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction," in *UbiComp 2014 - Adjunct Proc. of the 2014 ACM International Joint Conf. on Pervasive and Ubiquitous Computing*, (Seattle, Washington), pp. 1151–1160, 2014.

[14] *R: a language and environment for statistical computing [Online]*. R Core Team, 2019.

[15] M. E. Kret and E. E. Sjak-Shie, "Preprocessing pupil size data: Guidelines and code," *Behavior Research Methods*, vol. 51, no. 3, pp. 1136–1342, 2019.

[16] J. van Rij, P. Hendriks, H. van Rijn, R. H. Baayen, and S. N. Wood, "Analyzing the time course of pupillometric data," *Trends in Hearing*, vol. 23, pp. 1–22, 2019.