This is a repository copy of *Assessment of Soundscapes Using Self-Report and Physiological Measures*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/196527/

Version: Published Version

Article:

AUDIO ARTICLE

OPEN ACCESS

Assessment of soundscapes using self-report and physiological measures*

Simon Durbridge* and Damian Thomas Murphy

Audio Lab, School of Physics, Engineering and Technology, University of York, York YO10 5DD, United Kingdom

Abstract – Soundscape studies evaluate the subjective and objective qualities of an environment and attempt to develop a holistic view of the interplay between the acoustic scene and the listener's experience. Descriptors are used to express the perception of the acoustic environment, while further subjective and quantitative measures are used as indicators that represent features of the acoustic environment. The relationships between descriptors and indicators for a particular soundscape study are often identified by developing linear statistical models. This work describes an experiment to assess heart rate measures, including ultra short term heart rate variability, within the context of the predictor descriptor framework of a soundscape study. The aim of this work is to provide evidence in support of the psychophysiological basis of measures of affect in soundscape evaluation. In this study 15 participants evaluated a randomly ordered set of 8 soundscape recordings in a repeated measures directed listening experiment. Subjective evaluation of the soundscapes was performed using the self-assessment manikin and a sound classification survey. Participants' heart rate was measured throughout the experiment with a Polar H10 ECG heart rate monitor. Statistically significant relationships were identified between indicators and descriptors that reflect results present in the literature. However, there were no significant interactions between heart rate measures and self-reported affect or classification scores. Future studies should focus on improving the selection of stimuli and the experiment methodology to boost the sensitivity of the experiment in light of small effect sizes.

Keywords: Soundscape physiological heart rate variability

1 Introduction

Environmental noise has been identified as a critical pollutant. The European Environment Agency estimate that in the European region environmental noise contributes to 12,000 premature deaths and 48,000 new cases of ischemic heart disease each year [1]. Traditional environmental noise management strategies target key sources of environmental noise through a cyclical process of strategic identification and mitigation, measuring and modeling the levels of environmental noise from certain sources within agglomerations [2]. However, critics of these strategies of noise abatement suggest that they are insufficient to address the greater issues of environmental noise, being fundamentally reactionary and restricted to a few specific types of noise. The field of soundscape research focuses on understanding the perceptual quality of the acoustic environment from the perspective of the experience of the listener, within the context of the whole environment. Soundscape research

has been associated with developments in approaches to both urban planning and noise abatement, attempting to change the preconception of noise as a waste product to sound as a resource for urban designers to manage. Soundscape was reputedly conceived in the context of urban design by Michael Southworth, and was later popularized by R. Murray Schafer who founded the World Soundscape Project [3]. More recently soundscape has come to be more broadly associated with interdisciplinary approaches for environment evaluation and sound management in urban planning. A soundscape is defined in BS ISO 12913-1:2014 as the *acoustic environment as perceived or experienced and/or understood by a person or people, in context* [4].

Primary focuses of soundscapes research include identifying relationships between subjective and objective measures of the soundscape, and the the development of experimental methodologies that can be used to assess soundscapes from different perspectives [5–7]. The interdisciplinary nature of soundscape has resulted in a variety of methodologies for soundscape evaluation, suggesting that the field is still in the early stages of its development as a scientific field [8]. This is reflected in the inclusion of three different soundscape evaluation methodologies within the standards for soundscape [4, 9, 10]. The ISO/TS 12913-2:2018 technical specification for soundscape evaluation

---

The display of Audio files embedded in this PDF depends on the software used (PDF reader, video player, installed codec, direct display in the browser, etc.). Please see the Adobe Acrobat page for more explanation. You can also find the audio files at https://github.com/sedurCode/ActaAcustica22Resources.

identifies that descriptors and indicators should be used to reflect the evaluation and assessment of the concerned individuals [9]. A descriptor in this instance is a measure of how an individual experiences a soundscape, often through self-report of emotion or by the use of semantic descriptors of the soundscape [11]. An indicator is a measure of a feature of the soundscape, such as the subjective presence of a class of sound, or a measure of the average sound level, loudness and prominence ratio [12, 13].

The effective ground truth of the quality of the soundscape is implicit in the emotions and attitudes of the people within the environment, and so the primary tools available to evaluate soundscapes involve subjective estimation. Soundscape descriptors often include measures such as the subjective pleasantness and vibrancy of the environment, which some have equated to states of affect [14]. Affect in this context is the psychological definition as referring to the experience of feeling, emotion and mood [15].

In the fields of psychology and engineering there is a growing body of evidence identifying that changes in emotional state can be identified through physiological measures [16], and this has been extended into evaluating the experience of sounds and soundscapes [14]. The use of physiological measures that are influenced by emotional state may provide soundscape researchers with evidence of the psychophysiological basis for changes in affect in response to different soundscapes. Such evidence could be used to establish a ground-up theory of the perception of soundscapes. Several theories in psychophysiology support the relationship between states of affect and the behavior of the autonomic nervous system [17], including changes in heart rate variability and the regulation of emotions. The aim of this study is to understand whether participants exhibit changes in physiological behavior when listening to recordings of soundscapes in the context of a directed listening experiment under laboratory conditions.

1.1 Soundscape study methodologies

Soundscape studies are typically performed on location for the purposes of environment evaluation. However, in circumstances where experimental control is critical or the environment under evaluation does not yet exist, researchers have used virtual approximations of environments [18, 19]. The goal of a soundscape study might be to identify how the sounds of road-traffic, water features and acoustic mitigation might change the subjective quality of the soundscape [19–22]. The soundscape standards recommend three different methodologies for soundscape study [9]:

- Method A: A general questionnaire taken on location.
- Method B: A soundwalk where participants are guided between locations and are surveyed at those locations.
- Method C: A guided interview methodology.

Variations of the soundwalk methodology have been previously used in laboratory conditions as a *virtual soundwalk* [19, 23]. Surveys are often used in soundscape studies,

utilizing Likert or categorical scales that measure a subjective estimate of a quantity in ordinal levels. As the subjective ordinal data is specific to the participants' internal reference and participants often perform several assessments, data analysis often includes non-parametric statistical tests and procedures that allow the researchers to accommodate for the violation of the assumptions of statistical models. Principal component analysis has been used to identify the minimum effective dimensionality of soundscape perception, resulting in three dimensions of pleasantness, eventfulness and familiarity representing 78% of the variance in results of soundscape quality estimation across 116 attribute scales [24]. This three dimensional representation of the variance in the semantic-differential data reflects theories of affect being represented by three primary dimensions. Researchers have previously identified that three dimensions often represent a high proportion of variance in affect self-report and these are labelled valence, arousal and dominance [25–27]. The self-assessment manikin (SAM) is a pictographic form of survey that is used to evaluate these three most common dimensions of affect, providing a simplified form of affect self-report that has also been used in soundscape research [27, 28]. A paradigm often used as a measure of quality in soundscape research is the model of pleasantness and vibrancy [10], which is considered comparable to the affect dimensions of valence and arousal [14]. The SAM was extended from the semantic differential analysis procedure proposed by Mehrabian and Russell [15, 26], and has been used in studies of physiological responses to auditory stimuli [29].

Another aspect that is considered in soundscape evaluation is that of the classification of sound sources and soundscapes. As soundscape is a perceptually driven concept, the ground truth of the types of sounds perceived within an environment is also defined by those experiencing the environment. To compare the effect of sounds in different instances of a given context, researchers might compare the presence or absence of different sound classes to differences in affect in order to find evidence of a causal relation. Identifying classes of sounds requires the definition of a taxonomy of sound sources. Several forms of soundscape taxonomy have been proposed such as those by Bones et al. [30] and Trudeau and Guastavino [31]. The soundscape standards include a recommended taxonomy to be used for soundwalks in Annex C of the data collection technical specification [9]. Despite the development of these taxonomies, researchers often used a simplified taxonomy that is reduced to the three classes of natural, mechanical and human sounds [32].

1.2 Psychophysiological response studies in soundscape research

Several studies have attempted to identify physiological responses to sounds and soundscapes. Erfanian et al. published a systematic review of research studying psychophysiological factors in the context of soundscape [14]. Six relevant studies were identified from the literature search, most of which used a stimulus-locked repeated measures design. In this type of experiment design all participants

are exposed to a set of stimuli following a particular protocol of exposure, response and recovery. All of the studies included an evaluation of perceptual attributes related to experiencing the stimulus, and more specifically emotional or affective attributes. Some studies included the evaluation of participants' emotional state through the survey of affective dimensions [33], though in studies such as [34] affective evaluation is performed separately to physiological evaluation. Some studies evaluate the subjective quality of the stimuli by surveying with measures that describe the soundscape itself, as opposed to the participants' experience, including measures such as eventfulness, pleasantness and vibrancy [24]. These scales are identified by Erfianet al. as analogous to the affective dimensions of valence and arousal.

The studies include a variety of different physiological measures and measurement equipment, including galvanic skin response [35] and functional magnetic resonance imaging [34]. Average heart rate measurement was included in all studies but one, in which high frequency heart rate variability measures were included [36]. The outcome of this review was that results from the studies discussed were generally weak or conflicting when relating physiological responses to one emotional state. These conflicting results may be unsurprising, given the range of psychophysiological studies in the music psychology literature that come to similarly disparate conclusions [37]. However, it could also be that the size of the effect of the manipulation was insufficient for the given sample sizes in the presence of the other confounds that are likely to influence physiological measures. For example, Irwin et al. included 150 stimuli that were only 8 s in length, and used only 16 participants for the physiological experiment [34]. The heart rates of these participants are likely to have been strongly influenced by the experimental procedure, that is being in an fMRI scanner. Irwin et al. subsequently did not find evidence of a statistically significant effect of the stimuli on heart rate.

1.3 Heart rate variability measures in emotion estimation

Heart rate variability (HRV) analysis is becoming increasingly popular in affective computing and psychological research, in part thanks to the improvements in the availability and affordability of electrocardiogram (ECG) heart rate monitors [38, 39]. Heart rate variability measures are interpretations of a series of inter-beat intervals, the time periods between successive heart beats. There are several proposed theories that describe how HRV might reflect the psychophysiological state of an individual, suggesting a causal link between changes in emotional state and the behavior of the systems that regulate the heart. This causal link is facilitated by the balance between the sympathetic and the parasympathetic nervous systems (PNS), and it is theorized that the periodicity of the activation of the Vagus nerve or vagal tone is an indicator of the activity of the parasympathetic nervous system. The parasympathetic nervous system in turn is theorized to be representative of cognitive and emotional function [40]. Two key

theories that support the connection between emotion regulation and the function of the autonomic nervous system are the Polyvagal and Neurovisceral Integration theories. The Polyvagal theory was proposed by Porges [41] as a model of the neural regulation of the autonomic nervous system that also supports the relation between autonomic function and primary emotions [42]. This theory suggests that the adaptive mechanism of heart rate regulation are mediated by neurological mechanisms that are influenced by the environment and associated with behaviors include fight-or-flight and social interactions. The Neurovisceral Integration theory proposed by Thayer and Lane provides a model of a network of neurological systems that is theorized to be important for system regulation. This network is used to control several systems including the regulation of heart rate including heart rate variability [40]. Shaffer et al. suggest that healthy natural heart beat regulation is both periodic and highly complex [17]. Alternatively, under stress heart rate increases and HRV decreases as the vagal breaking mechanisms releases, allowing the heart rate to increase. Decreased HRV has also been associated with reduced PNS activation and has been observed in people suffering from stress, anxiety and panic [17]. There are several HRV metrics that reflect different statistical properties of the inter-beat interval series. Time domain measures are more suited to very-short-term and ultra-short-term HRV analysis (generally considered less than 5 min periods under analysis), as spectral methods require longer measurement periods to accurately reflect low frequency and very low frequency phenomena [43].

2 Methods

The aim of this study is to understand whether participants exhibit changes in physiological behavior when listening to recorded soundscapes, in the context of a directed listening experiment under laboratory conditions. To this end, heart rate measures (such as average heart rate and heart rate variability) were the chosen physiological measures, as they are theorized to reflect changes in affect. It therefore remains to identify if such heart rate measures are appropriate indicators of soundscape experience in the context of a seated lab based experiment, by identifying changes in these measures when experiencing different soundscapes. The experiment included the assumption that soundscapes reported as having a higher proportion of mechanical or natural sound sources would be reputed as having subjectively different qualities, such as the presence or absence of environmental noise. Under this assumption, the expectation would be that stimuli with differing classification scores would elicit different states of affect which would be reflecting in the physiological measures. Based on the reports of natural soundscapes being restorative or eliciting positive valence, a further expectation would be that soundscapes classified as being more natural would elicit higher valence. Another assumption was that ultra-short-term time domain heart rate measures are representative of changes in mood and emotion elicited by

Table 1. Stimuli selection summary.

Location type	Recording number	Clip	NDSI group	NDSI score	Stimuli	Player
Woodland	5	5	High	0.90	1	🔊
Woodland	7	1	Low	−0.86	5	🔊
Train station	4	5	High	−0.27	2	🔊
Train station	6	3	Low	−0.81	6	🔊
Park	3	3	High	0.41	3	🔊
Park	2	2	Low	−0.75	7	🔊
Beach	7	4	High	−0.03	4	🔊
Beach	2	3	Low	−0.85	8	🔊

experiencing soundscapes. The methodology used in this experiment follows on from Stevens et al. [11, 44] by performing a descriptor indicator comparison of changes in heart rate, affective report and classification score.

2.1 Participants

Participants were recruited from groups of audio engineering students via email. All participants had training in subjective testing, and could be considered expert listeners (see Section 5.4.1 of [45] for a discussion on the definition of expert listeners). A total of 15 participants were recruited to take part in this experiment. Of the participants 12 identified as male, 1 as female and 2 preferred not to say. The average age of participants was 26 and the standard deviation of the age was 5. The participants were screened by self-report for the following exclusion criteria:

- Abnormal or damaged hearing.
- Skin damage or known reactions to materials on the heart rate monitor.
- Known heart conditions, ailments or using medications that might directly effect heart rate.

2.2 Experimental stimuli

The experimental stimuli were selected from the EigenScape dataset, a set of high-order Ambisonic B-format recordings made in various locations around the UK [46]. The version of the dataset used in this experiment is comprised of first order ambisonic recordings, and is available from Zenodo [47]. The stimulus lengths used in the literature identified by Erfanian et al. [14] ranged from 8 s to 4 min, however the stimulus length used by Stevens et al. was 30 s [11]. A stimulus length of 40 s was chosen for this experiment in order to maximize the length of stimulus and the proportion of recall and recovery time for each test interval, while keeping the length of the test to within 30 min to avoid listener fatigue.

The EigenScape dataset was sliced into contiguous clips from which the test stimuli were then selected. All of the stimuli were screened for markers that could be used to identify individual people such as clear discernible speech. Stimuli were selected by algorithmically evaluating the proportion of natural and mechanical sounds in each clip, which was calculated using the Normalized Difference Soundscape Index (NDSI). NDSI is an acoustic index that

is intended to identify the proportions of natural and mechanical sound sources in a recording by comparing the proportions of spectral energy in two frequency bands, 1–2 kHz for mechanical sounds and 2–11 kHz for natural sounds [48–51]. NDSI scores range from +1.0 to −1.0, where a positive score indicates the presence of a higher proportion of natural sounds, and a negative score a higher proportion of mechanical sounds. The NDSI score was calculated for the 0th order omnidirectional W-channel of each soundscape sample by using a Python3 implementation of the R Acoustic Indices package [52].

The dataset contains soundscapes from 8 location types, this was reduced down to only include clips from the 4 locations with the greatest variance in NDSI. From each of these four locations, the 2 clips at the limits of the centre quartiles of the NDSI scores were selected as test stimuli. From each location type the clip at the limit of the positive quartile of NDSI score is an example from the set that is maximally natural and minimally mechanical and is therefore considered a high NDSI example. Conversely the clip at the limit of the negative quartile of NDSI scores is an example of a soundscape from that location type that is maximally mechanical and minimally natural, this clip is therefore a low NDSI example of that location type. The two training stimuli were the two samples in the data set with an NDSI score closest to zero. The selected test stimuli are summarized in Table 1. In the dataset recordings are organized by location type i.e., beach, woodland etc. Individual recordings and locations can be identified using a map provided by the author [47]. In Table 1 the location type and recording number are both from the dataset structure. The clip number identifies which contiguous 40 second slice of the recording number is used. The NDSI group is provided to identify whether the clip has a higher proportion of natural sounds (a more positive NDSI score) or mechanical sounds (a more negative NDSI score).

The selected stimuli were converted from B-format first order ambisonic to 2 channel binaural which could be easily reproduced over headphones. The B-format to binaural conversion was performed using the Binaural Decoder VST plugin that is part of the IEM plug-in suite, which is a selection of free tools that can be used for spatial audio processing [53]. The Binaural Decoder plug-in settings were set to SN3D normalization and 1st order Ambisonics, with headphone equalization disabled. The head related transfer functions used in the Binaural Decoder plug-in were recorded using a Neumann KU 100 dummy head, and the

Table 2. Stimuli loudness information.

Stimuli	1	2	3	4	5	6	7	8
Momentary	−43.35	−37.92	−38.04	−44.53	−51.19	−35.14	−36.35	−39.89
(LUFS)	(2.76)	(5.56)	(1.59)	(2.79)	(7.09)	(2.39)	(1.58)	(1.74)
Short-term	−42.61	−37.70	−37.76	−44.24	−49.76	−34.76	−36.11	−39.66
(LUFS)	(1.75)	(5.43)	(0.87)	(1.91)	(8.45)	(0.83)	(1.04)	(0.48)
Integrated	−43.32	−32.11	−38.14	−43.21	−42.44	−35.16	−36.71	−39.52
(LUFS)	(1.18)	(1.81)	(1.20)	(2.14)	(14.15)	(2.40)	(0.69)	(0.23)
Range	3.31	7.72	2.66	3.51	12.85	1.54	1.83	0.97
(LU)	(1.72)	(5.32)	(1.21)	(1.82)	(13.62)	(0.70)	(1.06)	(0.40)
Peak	−36.65	−31.57	−32.00	−38.10	−43.96	−28.36	−30.47	−33.67
(dBTP)	(2.77)	(5.33)	(3.00)	(3.32)	(6.19)	(3.20)	(2.74)	(2.70)

binaural rendering was performed using a magnitude least squares approach [54]. The stimuli were peak normalized to ensure a consistent relative level is maintained. No further processing was performed, and any low frequency rumble or wind noise in the recording was not compensated for.

The loudness of each stimuli is presented in Table 2. The loudness of each stimuli is given in loudness units full scale (LUFS) which is a standardized time-weighted and gated unit of the loudness of a signal relative to digital full-scale representation. The standard suite of loudness meters is calculated for each stimulus according to the algorithm defined in ITU-R BS.1770-4 [55], using the loudness meter that is built into Matlab. Table 2 presents each of the loudness metrics in the form mean (standard deviation).

2.3 Data collection instruments

Two forms of data gathering were used in the experiment, physiological sensing and self-report. For each test interval participants were asked to complete a survey based on their experience of the soundscape. Figure 1 presents the test page of the user interface. The survey featured two components as represented in Figure 1:

- A five-point self-assessment manikin in which participants report their affect.
- A classification task in which participants are tasked to report the proportions of sound sources present in the soundscape.

The self-assessment manikin featured three sets of graphics with 5 elements in each set, representing the three primary dimensions of affect; valence, arousal and dominance. The participants were instructed to select which of the five levels of each dimension affect reflected their state when listening to each soundscape. The dimensions were described to participants as follows:

- *Valence*: positive or negative emotions, analogous to feelings of pleasantness and happiness.
- *Arousal*: analogous to excitement and apathy.
- *Dominance*: analogous to the participants feeling or control or presence within the situation.

These descriptions were based on those from Stevens et al. [28]. The classification task included three sound

source classes from a typically used taxonomy of sound sources [14, 22]:

- *Natural* e.g. animal sounds, bird song and environmental sounds such as the wind or the ocean.
- *Mechanical* e.g. air, rail and road traffic, as well as construction and industrial sounds.
- *Human* e.g. foot fall, masked speech and laughter.

Each class was represented by a continuous linear slider, as opposed to the 5 point scale used by Stevens et al. [11]. The participants were instructed to consider the collection of sounds they could identify when listening to a soundscape, and position the three sliders to indicate the general proportionality of the different classes of sounds. Participants were given the following example scenario for a soundscape. A soundscape that is primarily composed of mechanical sounds, with no natural sounds and few human sounds, participants should move the slider associated with mechanical sounds toward the top position, the natural slider to the bottom position, and the human slider to somewhere between the bottom and middle position. Participants were directed to disregard any numerical value associated with the sliders, and instead use the slider's position to indicate the proportionality of sounds from each class.

2.4 Physiological measurement

The heart rate monitor used in the experiment was a Polar H10 ECG based chest strap monitor [56]. The H10 is popular, robust, relatively low cost, and has been shown to provide data of a quality similar to Holter style monitors [57]. Polar have published a white paper article suggesting that the H10 has an overall 95% accuracy in reporting R–R intervals during sports, which is an improved error rate in comparison to the group of Holter monitors tested in the same sporting activities [56]. Though the exact processing used in the Polar H10 is not disclosed, it could be inferred that some correction processes are used to ensure that the R–R series remains stable and accurate during these sporting activities [56, 58].

Heart rate data was streamed from the Polar H10 via Bluetooth to an Android mobile device (A Google Pixel 4 XL using version 10.0 of the Android operating system), which was running the Polar Sensor Logger App (version 6) [59]. The app records time-stamped heart rate
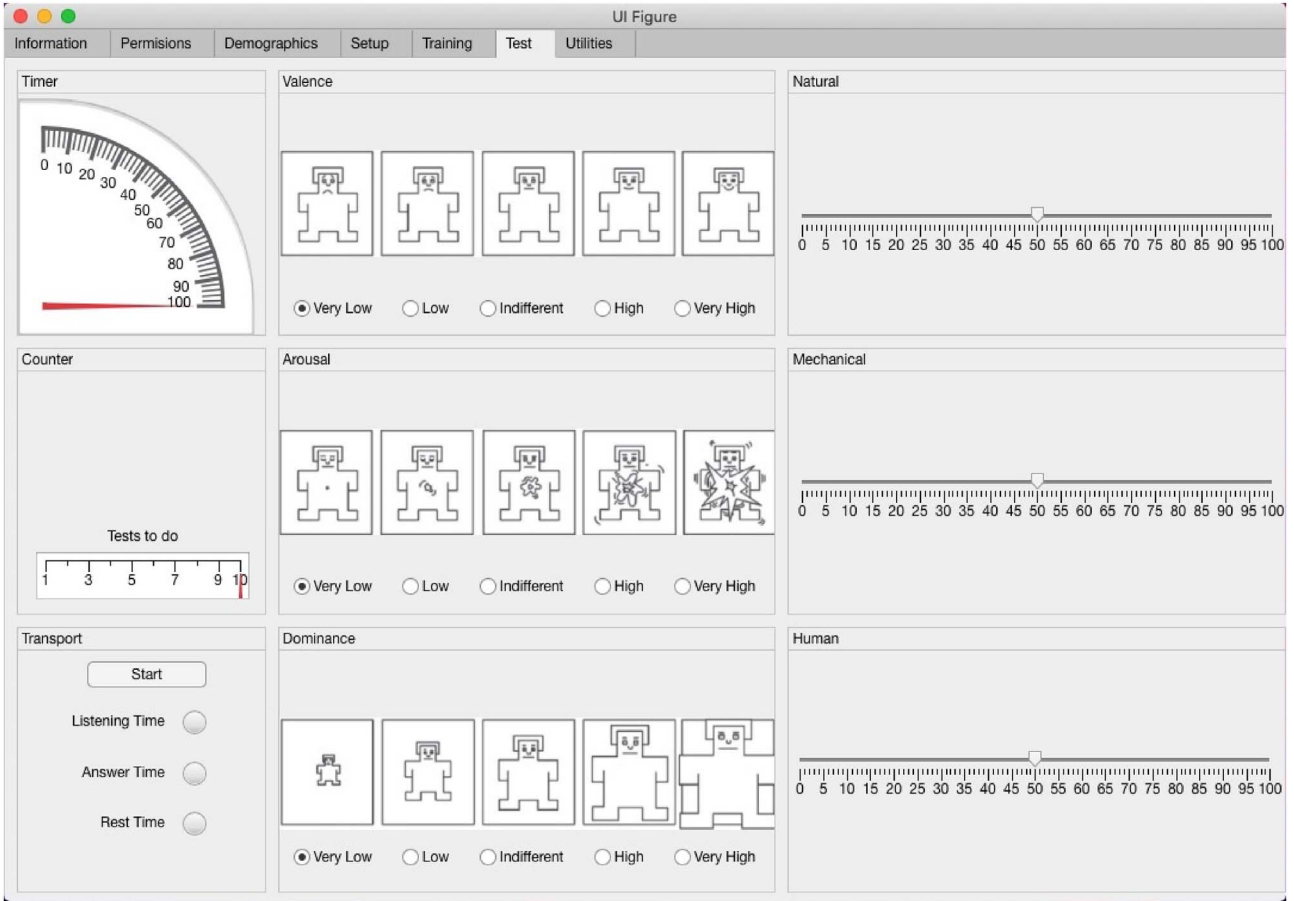
Figure 1. Listening test user interface.

and inter-beat interval estimates into comma-separated value (CSV) files. In post-processing of the heart rate data, sections of recorded heart rate that correspond to each interval of the test were indexed and sliced by the timestamps that were recorded by the listening test user interface. As the sensor sent data to the mobile device at a regular 1 s interval, a variable number of inter-beat intervals were reported with each message step. The inter-beat intervals were subsequently rearranged in post-processing before HRV analysis was performed.

In HRV analysis the inter-beat intervals need to be pre-processed to remove anomalous artifacts and ectopic beats [43]. All processing of inter-beat interval data was performed in Matlab using the HRVTool toolbox [60]. The data was first filtered to remove artifacts using the default thresholds provided by the toolbox. Where several false beats had been replaced by empty values, linear interpolation was performed to recover an estimate of the missing inter-beat intervals. The test data was subsequently sliced into the periods that align with each of the test intervals. Five HRV measures were computed using the HRVTool toolbox [61]:

- Root mean square of successive differences (RMSSD)
- Standard deviation normal to normal inter-beat intervals (SDNN)
- Percentage of successive normal intervals of more than 50 ms (pNN50)

- The median distance to the centre of the RR interval return map (rrHRV)
- The triangular interpolation of the NN interval histogram (TINN)

RMSSD has been indicated as a representative of vagal tone and is reputed to have good correlation with high frequency HRV [38]. SDNN is representative of the median variability in inter-beat intervals [62]. pNN50 is reported to be closely correlated with vagal tone and the the activity of the PNS [43]. rrHRV is a robust geometric measure of HRV that can be applied to short measurements [61]. TINN represents the spread of the histogram of inter-beat intervals by approximating the spread of the data with a triangle, greater variance is represented with a larger triangle [63].

2.5 Experimental design and procedure

After the initial screening and survey stage, participants were invited to fit the heart rate monitor as per the manufacturer's instructions, in privacy. Once fitted and tested, participants were guided to a waiting area to acclimatize to wearing the heart rate monitor. After the acclimatization period of 10 min participants were guided back to the formal test environment. The experiment followed a repeated measures design as used in the literature [11, 14].
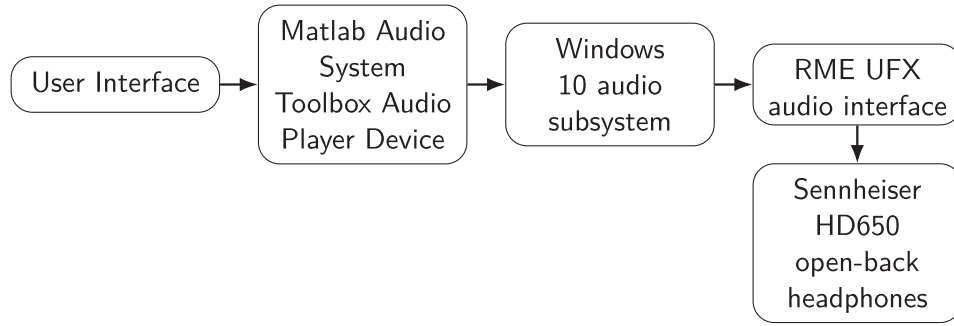
Figure 2. Signal chain of playback system.

Participants blindly listened to the randomly ordered stimuli, after which they reported their affect and performed the classification task. The experiment procedure featured a set of 2 training intervals followed by a set of 10 test intervals. In the first 8 test intervals all test stimuli were presented to participants. Two of the 8 test stimuli were randomly selected and were re-played in the last 2 test intervals. The test process was managed by a participant facing user interface that was presented via computer. The timing of the test was kept uniform across all test intervals, following the format:

- 40 s of listening to a soundscape.
- 30 s of reporting via the user interface.
- 60 s of rest.

As the test progressed a timestamp was recorded at each step in the procedure to allow for later synchronization between the stimuli playback and heart rate measurements. The listening portion of each test took approximately 26 min to complete. A further 15 min was required for the preparation and debrief stages.

Throughout the experiment a safe playback system level was maintained. The experiment was always performed in the same listening room and with the same equipment. The audio playback system signal chain is presented in Figure 2.

The playback system level was set to 90 dB sound pressure level (SPL) at full scale. The system was calibrated using a Behringer ECM8000 measurement microphone, which itself was calibrated using a Tenma 72-7260 sound level meter calibrator.

3 Results

To accommodate for the repeated measures design and the expected individual differences in responses, linear mixed effects models were used in the analysis below. The models included participant and stimuli as random intercepts to satisfy the assumptions of independence between samples. Statistical analysis was performed using Matlab [64], including the statistics and machine learning toolbox, the econometrics toolbox, the curve fitting tooblox, the normalitytest toolbox [65] and the HRV toolbox [60]. The fitting method used in the model was the restricted maximum likelihood estimate, meaning the models are less sensitive to outliers and biased estimates of the random effects terms.

3.1 Subjective estimates

The distribution of the subjective responses of participants with respect to the stimuli are summarized by the series of box plots in Figure 3. Figures 3a–3c summarize the distribution of affect scores from all participants across the stimuli. Figures 3d–3f summarize the distribution of classification scores from all participants across the stimuli. The box plots of Figures 3a–3c indicate considerable similarities in the state of affect elicited by the different stimuli, particularly for the dimensions of arousal and dominance. Figures 3b and 3c give little evidence of a significant relationship between the stimuli and self-reported arousal or dominance, with the median of affect scores generally sitting at 0 or indifference. However, there was a larger range in valence scores across the stimuli leading to an identifiable positive relationship between the stimuli and self-reported valence $F(1,118) = 12.099$, $p \ll 0.001$, $\eta_p^2 = 0.46$. Comparing Figures 3a and 3b it appears that decreased valence scores are associated with increased arousal scores, and a significant interaction was identified $F(1,118) = 9.125$, $p = 0.003$, $\eta_p^2 = 0.08$. Figures 3d and 3e highlight that natural and mechanical scores were generally in opposition, with the notable exception of stimuli 5 which was considered to have a high proportion of both natural and mechanical sounds. Natural scores varied significantly between stimuli $F(7,98) = 56.72$, $p \ll 0.001$, $\eta_p^2 = 0.8$ as did mechanical scores $F(7,98) = 35.18$, $p \ll 0.001$, $\eta_p^2 = 0.72$ and human scores $F(7,98) = 19.91$, $p \ll 0.001$, $\eta_p^2 = 0.59$, suggesting that the selection of stimuli included a reasonable variety of content.

Of the classification scores, two stimuli have elicited results of particular interest, stimuli 8 and 5. Stimulus 8 elicited a wide range of natural and mechanical scores, suggesting that in this context participants disagreed on whether the soundscape was explicitly highly mechanical or highly natural sounding. Stimulus 8 ◀⁾)   was a beach soundscape with the sound of people walking and talking, the jangle of metal and the rolling of a pram. Stimuli 5 ◀⁾)   was a woodland soundscape with the sound of a steam locomotive, eliciting both highly mechanical and highly natural scores. The participants reported positive valence for this stimuli, giving a context in which highly mechanical sounds in a highly natural setting elicit positive valence. This provides a counter point for the expectation
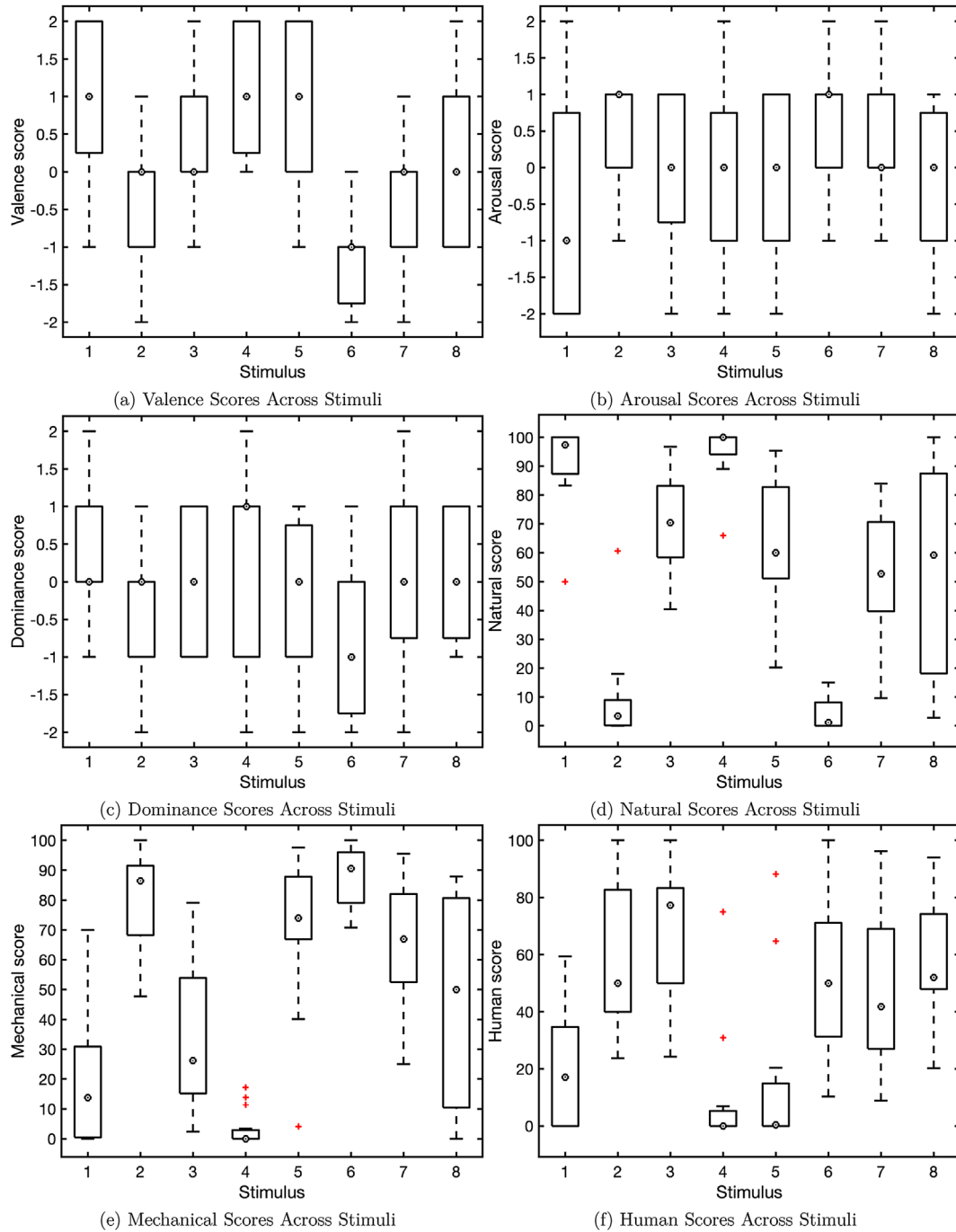
Figure 3. Distribution of subjective estimates for each stimuli across all participant.

that mechanical sounds in the context of highly natural settings would lead to reports of negative valence.

3.2 Physiological responses

Heart rate measurements were averaged across each test interval. The average resting heart rate of each participant was calculated by averaging across the heart rate recorded in the pre-test rest condition. Each participant's average resting heart rate was then subtracted from each respective test interval in order to normalize the data, giving the average heart rate Δ. Figure 4 presents the mean and standard deviation of average heart rate Δ across all participants per stimuli.

It is clear from Figure 4 that the mean normalized average heart rate is similar across the stimuli, and there is an obvious overlap in the distribution of heart rates across the stimuli. It is clear from Figure 4 that there are no instances where a stimulus has elicited a consistent large change in average heart rate from rest.

Figure 5 presents boxplots of the distribution of heart rate metrics per stimuli across all test intervals. It is clear
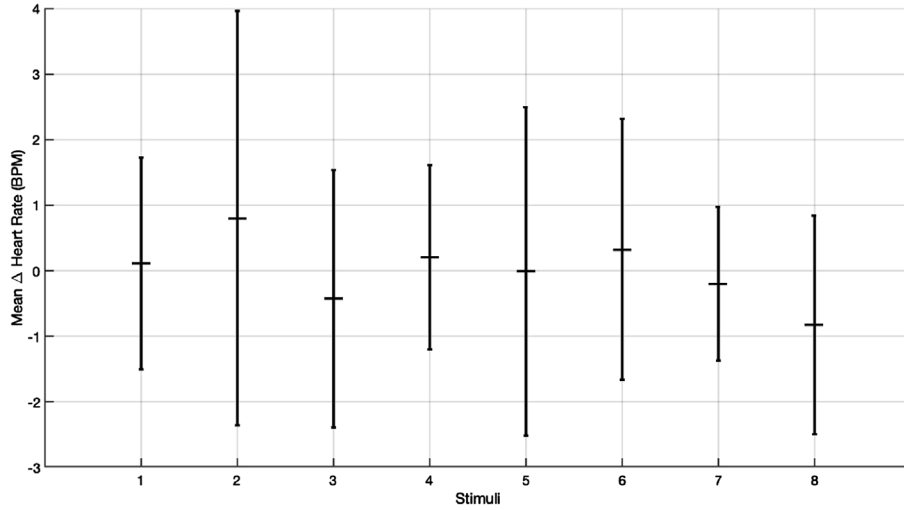
Figure 4. Average heart rate Δ across all participants for each stimulus.

from all of the subplots of Figure 5 that there are no large differences in the effect of any stimuli for any HRV measure. Though there appears to be differences in the medians for several of the measures, the differences appear very small suggesting any effect size would in turn be very small. The analysis of HRV measures are often performed on data recorded over much larger periods of time. To identify if any experimental effect may be present at all, a comparison is presented for an HRV measure between the pre-test and test conditions for all the participants. Figure 6 presents the absolute mean log RMSSD for each participant across the test and rest conditions. The height of the bars presented the absolute mean, and the size of the error bars represents the standard deviation.

The data presented in Figure 6 shows there are no large differences between these HRV values between the pre-test and test condition. This indicates that the test condition is unlikely to have have had a consistent, systematic or distinguishable effect on heart rate variability compared to the pre-test condition of rest. However, the lengths of the pre-test and test periods used in the data from Figure 6 are not the same, the test period was significantly longer than the pre-test period. The data was log transformed to improve the visibility of differences on the chart, given the small scale of the RMSSD data.

Table 3 presents the mean and standard deviation of the difference of average heart rate between test intervals and the pre-test period for each participant.

The data in Table 3 shows a large variation in the change of heart rate from pre-test and test intervals across the participants. However, most participants' heart rates decreased in the test compared to the pre-test condition. Though no statistically significant effect of the stimuli on heart rate was detected, in most cases participants may have relaxed when sitting down to perform the test.

3.3 Correlation

A correlation table comparing measures is presented in Table 4. The correlation coefficients were calculated using Spearman's rank correlation which was chosen due to the nature of the types of data being analyzed. Only correlation coefficients with a p value of less than 0.05 are presented in the table.

The correlation coefficients presented in Table 4 show that natural, mechanical and human scores are correlated with valence, arousal and dominance reports respectively. Natural scores are correlated with mechanical scores and share collinearity, and as such different models are presented below to evaluate these indicators independently. Only two correlations appear between HRV measures and other variables. SDNN is positively correlated with test interval number, suggesting this value may increase over the course of the experiment. rrHRV is positively correlated with valence report, suggesting that the variance of interbeat intervals might increase when experiencing more pleasant soundscapes.

3.4 Linear mixed-effects models

Given the correlations present within the sets of subjective indicators, descriptors and HRV measures, single factor linear mixed-effects models were computed for pairwise descriptors, indicators and HRV measures. The statistics of these models are presented in Table 5, including the model estimates, standard errors for the primary effects and the upper and lower bounds of the 95% confidence intervals. The model residuals were tested for normality using the Shapiro–Wilk test, and plots were evaluated for heteroscedasticity.

Analysis of variance was performed on the fixed effects of the models presented in Table 5 that resulted in statistically significant estimates. From these results we find that valence varied significantly with natural scores $F(1,118) = 44.8$, $p \ll 0.001$ and mechanical scores $F(1,118) = 25.05$, $p \ll 0.001$. Arousal varied significantly with natural scores $F(1,118) = 14.2$, $p < 0.001$, mechanical scores $F(1,118) = 12.1$, $p < 0.001$ and human scores $F(1,118) = 8.37$, $p < 0.01$. These relationships are summarized in the subplots of Figure 7. Figure 7 presents the classification scores with
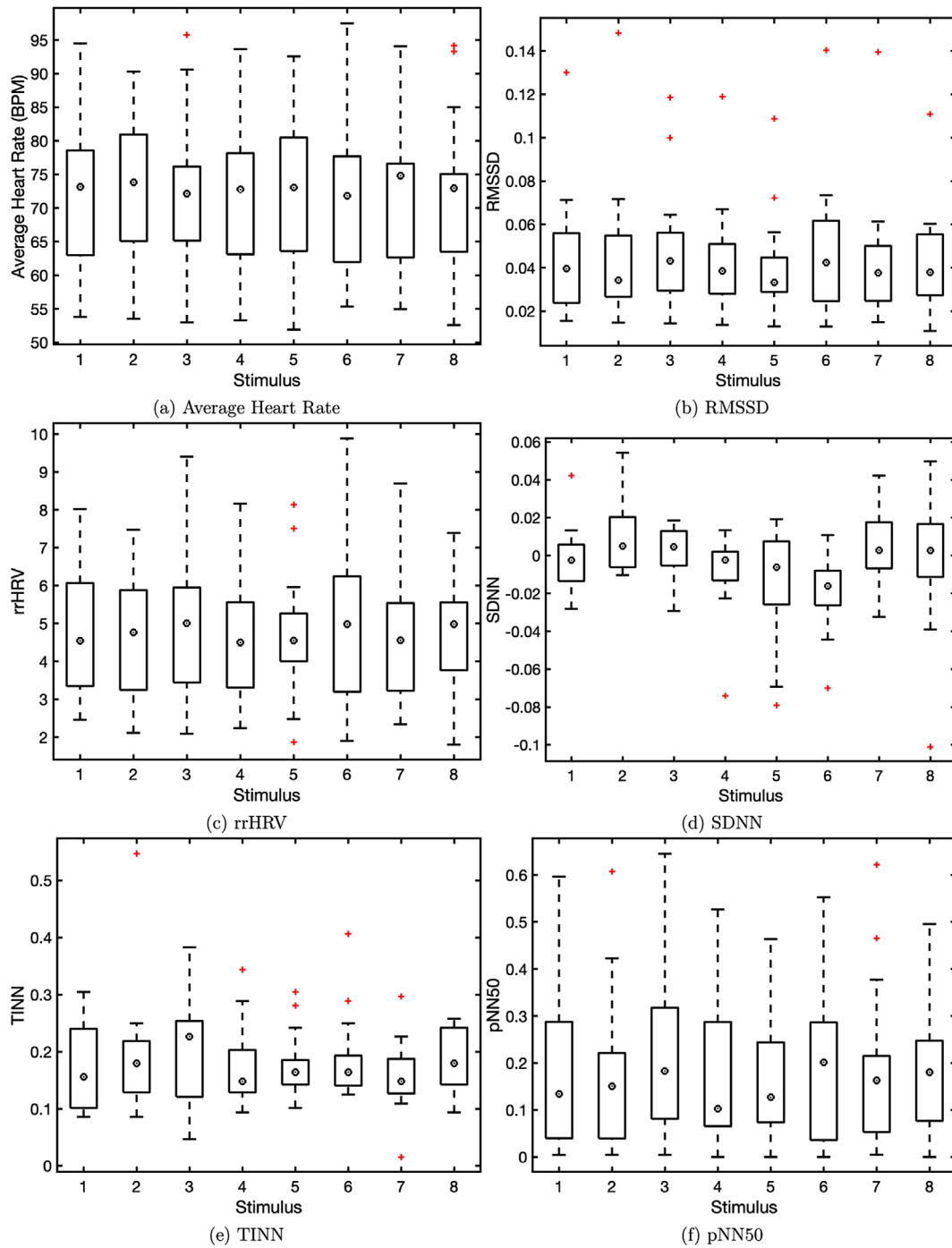
Figure 5. Distribution of heart rate metrics for each stimuli across all participants.

respect to the fitted values of valence and arousal across the respective models. The models are presented with lines of best fit to make the data more clear.

Both Figures 7a and 7b show clear trends between the classification scores and affective report scores for the associated models. The only HRV metric to vary significantly with valence was rrHRV $F(1,118) = 5.04$, $p = 0.026$. No other statistically significant relationships were observed between a

HRV measure, a descriptor or an indicator. Figure 8 presents the rrHRV measures with respect to the fitted values of valence, including a line of best fit to make the data more clear.

There appears to be a trend in the rrHRV data presented in Figure 8 with respect to the fitted values of the model for rrHRV and valence, though given no other relationships were found this effect should be confirmed in a study with a larger sample size.
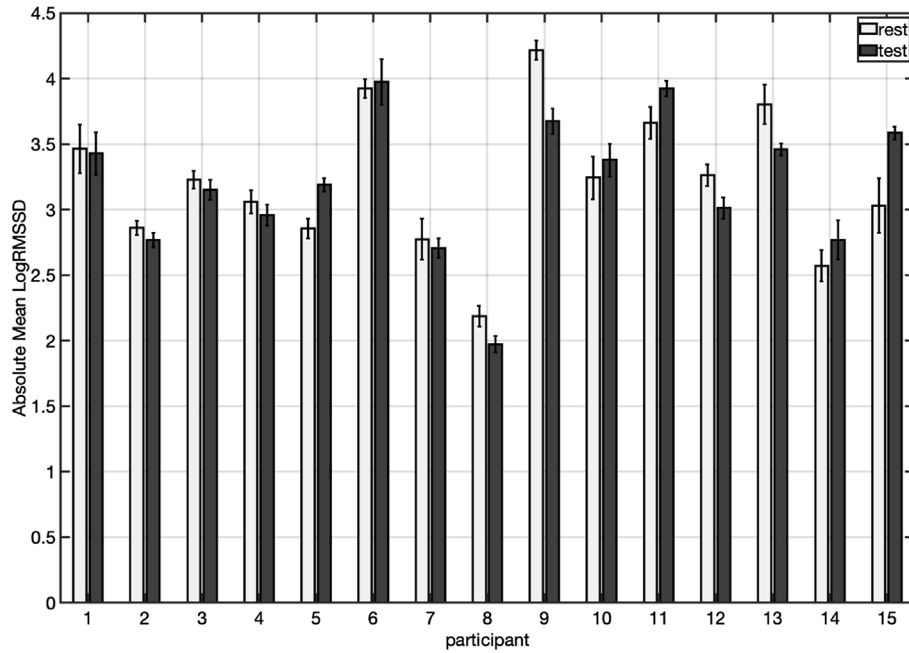
Figure 6. Absolute mean log RMSSD for each participant in the test and pre-test rest conditions.

Table 3. Average difference in mean heart rate between pre-test and test conditions. Below the mean differences, the standard deviations are presented within the parentheses.

Participant	1	2	3	4	5
	−6.57	−3.70	−2.88	0.72	−6.33
	(1.36)	(2.25)	(1.27)	(1.40)	(1.68)
Participant	6	7	8	9	10
	−2.80	−4.44	−9.02	−10.93	−3.40
	(2.69)	(1.91)	(1.09)	(2.33)	(1.19)
Participant	11	12	13	14	15
	−10.23	−15.90	−11.56	−0.62	3.017
	(1.18)	(2.93)	(2.42)	(3.22)	(1.39)

Table 4. Correlation coefficients between subjective estimates and HRV measures.

| Factor	Interval	Stimulus	Valence	Arousal	Dominance	Natural
Arousal			−0.36		−0.34	
Dominance			0.44	−0.34		
Natural	0.19	−0.23	0.63	−0.31	0.35	
Mechanical	−0.24	0.25	−0.57	0.28	−0.38	−0.76
Human			−0.38	0.22		−0.38
Heart rate						
RMSSD						
rrHRV			−0.22			
SDNN	0.23					
TINN						
pNN50						

4 Discussion

In this experiment 15 participants evaluated a randomly ordered set of 8 soundscapes, performing a survey of classification scores and self-reported affect. Heart rate and heart rate variability measures were recorded throughout the experiment. Statistically significant relationships between soundscape classification and affect self-report were identified, with higher natural and lower mechanical scores reflecting higher reported valence and lower reported arousal. The subplots of Figure 7 visualize the relationships between classification scores, valence and arousal respectively. Interestingly the classification of human sounds follows a similar trend to mechanical sound classifi-

Table 5. Linear mixed-effects models.

Factor	Valence				Arousal			
	Estimates	Std. error	95% CI		Estimates	Std. error	95% CI	
Natural	**0.0189**[***]	0.0028	−1.289	−0.471	**−0.0091**[***]	0.0024	−0.0140	−0.0043
Mechanical	**−0.0162**[***]	0.0032	−0.022	−0.0098	**−0.0089**[***]	0.0025	0.0038	0.0140
Human	−0.0062	0.0035	−0.013	0.7e−3	**−0.0094**[**]	0.0032	0.0029	0.0158
Heart rate	0.0059	0.0096	−0.013	0.025	−0.0099	0.0097	−0.0291	0.0092
RMSSD	−3.329	3.75	−10.82	4.034	2.837	3.98	−5.04	10.72
rrHRV	**−0.123**[*]	0.055	−0.232	−0.146	0.0039	0.065	−0.124	0.1327
SDNN	5.313	3.49	−1.613	12.24	−1.305	3.97	−9.23	6.5
TINN	−0.2189	1.145	−2.486	2.048	2.188	1.26	−0.32	4.7
pNN50	−0.708	0.613	−1.924	0.506	0.367	0.674	−0.968	1.7
			Model information					
Observations	120							
Fixed effects coeffs	2							
Random effects coeffs	23							

[***] $p < 0.01$.
[**] $p < 0.05$.
[*] $p < 0.1$.

(a) Fitted Models for Valence

(b) Fitted Models for Arousal
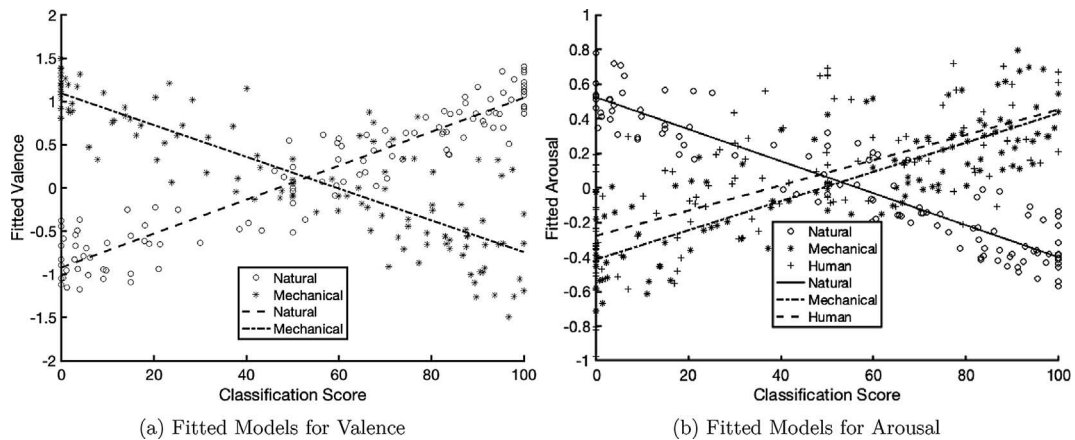
Figure 7. Distribution of subjective estimates for each stimuli across all participants.

cation with respect to arousal responses. These results reflect similar findings to previous research, highlighting the context independent effect of natural sounds on valence and arousal. Stevens et al. identified a position correlation between natural scores and valence scores, and a negative correlation between mechanical and human scores and valence scores across two experiments focussed on soundscape assessment [11, 28].

Only one relationship between a descriptor and an HRV metric was identified in statistical modeling. This relationship between rrHRV and self-reported valence is presented in Figure 8 and reflects a trend of a lower rrHRV with increased valence. This result is contrary to the suggestion that HRV would increase under more relaxed conditions such as listening to a more pleasant soundscape. As rrHRV is intended to be a more suitable and robust HRV measure with respect to outliers and shorter time periods, it could be that the other HRV measures would exhibit similar relationships if the stimuli were longer or the effect size was increased. However, as no other HRV metric presented a similar relationship to any descriptor, this result is likely

to be anomalous and warrants further investigation with a larger sample size. These findings or lack there-of are similar to those reported by Irwin et al. and Medvedev et al., their of whom reported finding statistically significant relationships between environmental sounds or soundscapes and heart rate or HRV measures [14, 34, 66].

4.1 Limitations and further research

This experiment had several limitations that should be addressed when designing further studies of physiological responses to soundscapes. Though the sample size in this experiment was sufficient to identify the effects of soundscape classification on emotional affect, the sample size was insufficient to develop confident conclusions on the presence of an effect of soundscape on HRV measures. Future work should focus on improving the methodology of the experiment in order to maximize the effect size. Ciuk et al. previously identified that physiological measures were significantly weaker than self-reported affect at estimating the influence of attitudes on policy agreement in a study
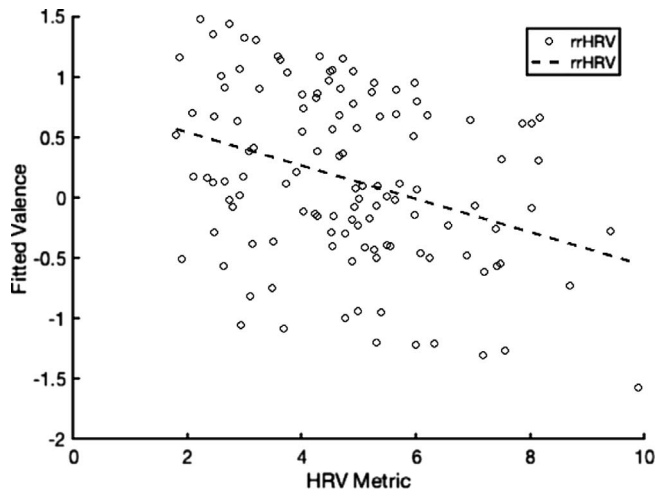
Figure 8. Distribution of rrHRV with respect to valence scores.

of the influence of affect on policy making decisions of federalism [67]. Though the sample size of 106 undergraduate students used in the study was large compared to those typically used in psychophysiological evaluations of soundscape, the researchers concluded that physiological measurements were not appropriate replacements for self-report in studies of political science, and that the stimuli required for an effective study must elicit a very strong emotional reaction. Though large sample sizes are desirable for robust statistical analysis, a greater focus on managing confounds and improving the stimuli may yield more significant results than larger sample sizes alone.

This experiment was designed to minimize external factors that might elicit changes in heart rate. The experiment always took place seated, in a warm insulated environment, with minimal participant engagement. These experimental conditions may have primed participants to be significantly relaxed, indifferent and even bored. Static binaural rendering was utilized to improve the ecological validity of the soundscape reproduction, but there was little impetus for participants to engage in the evaluation and suspend their disbelief. Perhaps the effect size of soundscape experience on HRV can be boosted by significantly improving the experiment design to be more ecologically valid. Recent studies have suggested that the use of VR technologies can improve the immersive nature of such experiments [68]. Future research should utilize advanced environment rendering technologies to improve ecological validity and participant engagement within the experiment.

Another factor in the quality of the study was the demographics of the participants. Future studies should include a greater diversity of the population being sampled in the study design, with the intention of ensuring the sample of participants is representative of the wider population under consideration. The participants in this experiment were not surveyed for factors that could reflect the prior context of their experience such as their nationality and the type of environment they grew up in. Researchers have reported differences in responses to soundscapes that are related to the nationality of the participants [69]. Further, the

participants were not surveyed for their affect and attitudes prior to performing the study. There is evidence that physiological measures might be sensitive to several confounds including a participants' disposition, hydration level and alcohol intake prior to the experiment. Future studies should make appropriate steps to ensure these confounds do not influence the experimental results [70].

The strategy for the selection of stimuli was developed with the intention of avoiding systematic bias. The stimuli were selected through a process of algorithmic evaluation, taking advantage of an established metric for evaluating soundscape ecology [48]. However, researchers using a similar stimuli selection strategy should attempt to compare several metrics that are intended for similar purposes, instead of limiting the range of metrics to one. A further limitation of the stimuli selection procedure was that no further pre-processing was used, and several of the recordings include strong low frequency rumbles that are likely caused by wind noise. This low frequency noise is quite obvious in some playback systems, and a researcher improving on this study should consider using appropriate high pass filtering.

5 Conclusions

In this study fifteen participants took part in a directed listening experiment that was intended to identify if psychophysiological responses would occur when listening to a variety of soundscapes. These responses were measured using a survey and a Polar H10 ECG based heart rate monitor. Results for the objective classification of soundscape composition (indicators) and subjective self-report (descriptors) were similar to previous works such as Stevens et al. [11], indicating that there were appropriate differences in affective report and classification between stimuli. However, no statistically significant changes in heart rate or heart rate variability measures were identified. Heart rate in this case does not appear to be an effective descriptor of differences between the stimuli, given the small sample size used in this repeated measures experiment. Further research should attempt to identify alternative methodological approaches that could elicit and detect such psychophysiological responses.

Conflict of interest

The authors declare that they have no conflict of interests.

Data availability statement

This article includes selected audio files embedded in the article. The complete set of audio files, accompanying data and code are available at https://github.com/sedurCode/ActaAcustica22Resources [71].

Acknowledgments

through the EPSRC doctoral training studentship: reference number EP/R513386/1.

References

1. European Environment Agency: Environmental noise in Europe – 2020, 2020. No. 22/2019
2. European Union: Directive 2002/49/EC relating to the assessment and management of environmental noise (END), no. L189, Official Journal of the European Communities, 2002:12–25Available online: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32002L0049 from=EN.
3. Ö. Axelsson: Soundscape revisited. Journal of Urban Design 25, 5 (2020) 551–555. https://doi.org/10.1080/13574809.2020.1810006.
4. BS ISO 12913-1:2014: Acoustics – Soundscape part 1: Definition and conceptual framework, The British Standards Institution, London, UK, September 2014.
5. W.J. Davies, M.D. Adams, N.S. Bruce, R. Cain, A. Carlyle, D. Cusack, P. Hall, K.I. Hume, A. Irwin, P. Jennings, C.J. Plack, J. Poxon: Perception of soundscapes: An interdisciplinary approach. Health (San Francisco) (2012) 2–7. https://doi.org/10.1016/j.apacoust.2012.05.010.
6. M. Cik, M. Lienhart, P. Lercher: Analysis of psychoacoustic and vibration-related parameters to track the reasons for health complaints after the introduction of new tramways. Applied Sciences 6, 12 (2016) 398. http://www.mdpi.com/2076-3417/6/12/398.
7. M. Lienhart, M. Cik, M. Fellendorf, K. Fallast, E. Marth, W. Freidl, H. Mayerl: Psychoacoustic indicators of road and rail traffic noise, subjective perception and psychological and physiological parameters, in INTER-NOISE and NOISE-CON Congress and Conference Proceedings, Internoise, Crete, 2018, pp. 466–472.
8. Ö. Axelsson, C. Guastavino, S.R. Payne: Editorial: Soundscape assessment. Frontiers in Psychology 10 (2019). https://doi.org/10.3389/fpsyg.2019.02514.
9. Acoustics – Soundscape: Part 2: Data collection and reporting requirements, BSI Standards Publication, 2018.
10. PD ISO/TS 12913-3:2019: Acoustics – Soundscape, BSI Standards Publication, 2019.
11. F. Stevens, S. Smith, D. Forest: Soundscape categorisation and the self-assessment manikin, in DAFx 2017 – Proceedings: 20th International Conference on Digital Audio Effects, Edinburgh, UK, September 2017, pp. 481–488.
12. M.S. Engel, A. Fiebig, C. Pfaffenbach, J. Fels: A review of the use of psychoacoustic indicators on soundscape studies. Current Pollution Reports 7, 3 (2021) 359–378. https://doi.org/10.1007/s40726-021-00197-1.
13. F. Aletta, J. Kang, Ö. Axelsson: Soundscape descriptors and a conceptual framework for developing predictive soundscape models. Landscape and Urban Planning 149 (2016) 65–74.
14. M. Erfanian, A.J. Mitchell, J. Kang, F. Aletta: The psychophysiological implications of soundscape: A systematic review of empirical literature and a research agenda. International Journal of Environmental Research and Public Health 16, 19 (2019) 3533.
15. J.A. Russell: A circumplex model of affect. Journal of Personality and Social Psychology 39, 6 (1980) 1161–1178.
16. A. Dzedzickis, A. Kaklauskas, V. Bucinskas: Human emotion recognition: Review of sensors and methods. Sensors (Switzerland) 20, 3 (2020) 592.
17. F. Shaffer, R. McCraty, C.L. Zerr: A healthy heart is not a metronome: An integrative review of the heart's anatomy

18. and heart rate variability. Frontiers in Psychology 5 (2014) 1–19.
18. W.J. Davies, M.D. Adams, N.S. Bruce, R. Cain, A. Carlyle, P. Cusack, D.A. Hall, K.I. Hume, A. Irwin, P. Jennings, M. Marselle, C.J. Plack, J. Poxon: Perception of soundscapes: An interdisciplinary approach. Applied Acoustics 74, 2 (2013) 224–231. https://doi.org/10.1016/j.apacoust.2012.05.010.
19. S. Harriet, D.T. Murphy: Auralisation of an urban soundscape. Acta Acustica United with Acustica 1014 (2015) 798–810.
20. C. Guastavino, B.F.G. Katz, J.-D. Polack, D.J. Levitin, D. Dubois: Ecological validity of soundscape reproduction. Acta Acustica United with Acustica 91 (2004) 1–19.
21. W. Davies, M. Adams, N. Bruce, M. Marselle, R. Cain, P. Jennings, J. Poxon, A. Carlyle, P. Cusack, D. Hall, A. Irwin, K. Hume, C. Plack: The positive soundscape project: A synthesis of results from many disciplines, in 38th International Congress and Exposition on Noise Control Engineering 2009, INTER-NOISE 2009, vol. 1, December 2013, 2009.
22. F. Stevens: Strategies for environmental sound measurement, modelling and evaluation. PhD thesis, University of York, 2018.
23. T. Oberman, K. Jambrošić, M. Horvat, B.B.O. Šćitaroci: Using virtual soundwalk approach for assessing sound art soundscape interventions in public spaces. Applied Sciences 10, 6 (2020) 2102.
24. Ö. Axelsson, M.E. Nilsson, B. Berglund: A principal components model of soundscape perception. The Journal of the Acoustical Society of America 128, 5 (2010) 2836–2846. https://doi.org/10.1121/1.3493436.
25. C. Osgood, G. Suci, P. Tannenbaum: The measurement of meaning, 1st ed., Urbana, University of Illinois Press, Illinois. 1957. http://yorsearch.york.ac.uk/44YORK:44YORK_ALMA_DS21218486570001381.
26. I. Bakker, T. van der Voordt, P. Vink, J. de Boon: Pleasure, arousal, dominance: Mehrabian and Russell revisited. Current Psychology 33, 3 (2014) 405–421.
27. M.M. Bradley, P.J. Lang: Measuring emotion: the self-assessment manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry 25, 1 (1994) 49–59.
28. F. Stevens, D.T. Murphy, S.L. Smith: Soundscape preference rating using sematic differential pairs and the self-assessment manikin, in Proceedings SMC 2016, Hamburg, Germany, August 2016.
29. M.M. Bradley, P.J. Lang: Affective reactions to acoustic stimuli. Psychophysiology 37, 2 (2000) 204–215.
30. O. Bones, T.J. Cox, W.J. Davies: Sound categories: Category formation and evidence-based taxonomies. Frontiers in Psychology 9 (2018) 1277. https://doi.org/10.3389/fpsyg.2018.01277.
31. C. Trudeau, C. Guastavino: Classifying soundscapes using a multifaceted taxonomy, in Euronoise 2018, Crete, Greece, May 2018, pp. 2487–2492. Available online: http://www.euronoise2018.eu/docs/papers/411_Euronoise2018.pdf.
32. F. Stevens, D.T. Murphy, S.L. Smith: Soundscape auralisation and visualisation: A cross-modal approach to soundscape evaluation, in DAFx 2018 – Proceedings: 21st International Conference on Digital Audio Effects, Aveiro, Portugal, September 2018, pp. 133–140.
33. P. Gomez, B. Danuser: Affective and physiological responses to environmental noises and music. International Journal of Psychophysiology 53, 2 (2004) 91–103.
34. A. Irwin, D.A. Hall, A. Peters, C.J. Plack: Listening to urban soundscapes: Physiological validity of perceptual dimensions. Psychophysiology 48, 2 (2011) 258–268.
35. K. Hume, M. Ahtamad: Physiological responses to and subjective estimates of soundscape elements. Applied Acoustics 74, 2 (2013) 275–281. https://doi.org/10.1016/j.apacoust.2011.10.009.

36. J.J. Alvarsson, S. Wiens, M.E. Nilsson: Stress recovery during exposure to nature sound and environmental noise. International Journal of Environmental Research and Public Health 7, 3 (2010) 1036–1046. Available online: https://www.mdpi.com/1660-4601/7/3/1036.

37. D. Hodges: Music in the human experience: An introduction to music psychology, Taylor & Francis. 2019. Available online: https://books.google.co.uk/books?id=lx0lEAAAQBAJ.

38. S. Laborde, E. Mosley, J.F. Thayer: Heart rate variability and cardiac vagal tone in psychophysiological research – Recommendations for experiment planning, data analysis, and data reporting. Frontiers in Psychology 8 (2017) 1–18.

39. M. Egger, M. Ley, S. Hanke: Emotion recognition from physiological signal analysis: A Review. Electronic Notes in Theoretical Computer Science 343 (2019) 35–55. https://doi.org/10.1016/j.entcs.2019.04.009.

40. J.F. Thayer, R.D. Lane: A model of neurovisceral integration in emotion regulation and dysregulation. Journal of Affective Disorders 61, 3 (2000) 201–216.

41. S. Porges: Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. A polyvagal theory. Psychophysiology 32 (1995) 301–318. https://doi.org/10.1111/j.1469-8986.1995.tb01213.x.

42. S.W. Porges: The polyvagal perspective. Biological Psychology 74, 2 (2007) 116–143.

43. F. Shaffer, J.P. Ginsberg: An overview of heart rate variability metrics and norms. Frontiers in Public Health 5 (2017) 1–17.

44. F. Stevens, D.T. Murphy, S.L. Smith: The self-assessment manikin and heart rate: Responses to auralised soundscapes, in Interactive Audio Systems Symposium, New York, 2016, pp. 1–8. Available online: https://www.york.ac.uk/sadie-project/IASS2016/IASS_Papers/IASS_2016_paper_3.pdf.

45. S. Bech, N. Zacharov: *Perceptual Audio Evaluation - Theory, Method and Application*, Wiley. 2007. Available online: https://books.google.co.uk/books?id=1WGPJai1gX8C.

46. M. Green, D. Murphy: EigenScape: A database of spatial acoustic scene recordings. Applied Sciences 7, 11 (2017) 1204.

47. M. Green, D. Murphy: EigenScape Dataset, Zenodo, October 2017. Available online: https://zenodo.org/record/2628463.

48. E.P. Kasten, S.H. Gage, J. Fox, W. Joo: `The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology. Ecological Informatics 12 (2012) 50–67.

49. J. Sueur, A. Farina, A. Gasc, N. Pieretti, S. Pavoine: Acoustic indices for biodiversity assessment and landscape investigation. Acta Acustica United with Acustica 100, 4 (2014) 772–781.

50. P. Devos: Soundecology indicators applied to urban soundscapes. Inter-noise 2016, 4 (2016) 177–184.

51. M. Green, D. Murphy: Environmental sound monitoring using machine learning on mobile devices. Applied Acoustics 159 (2020) 107041. https://doi.org/10.1016/j.apacoust.2019.107041.

52. P. Guyot: Acoustic indices github repository, 2019. Available online: https://github.com/patriceguyot/Acoustic_Indices.

53. IEM: IEM Plug-in Suite, 2019. Available online: https://plugins.iem.at/

54. C. Schörkhuber, M. Zaunschirm, R. Höldrich: Binaural rendering of ambisonic signals via magnitude least squares, in Fortschritte der Akustik – DAGA 2018, 2018, pp. 339–342.

55. International Telecommunication Union Radiocommunication Bureau: ITU-R BS.1770-4, Tech. Rep., Geneva, 2017. Available online: http://www.itu.int/ITU-R/go/patents/en

56. Polar Research and Technology: Polar h10 heart rate sensor system, 2019. Available online: https://www.polar.com/sites/default/files/static/science/white-papers/polar-h10-heart-rate-sensor-white-paper.pdf

57. R. Gilgen-Ammann, T. Schweizer, T. Wyss: RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. European Journal of Applied Physiology 119, 3 (2019) 1525–1532. https://doi.org/10.1007/s00421-019-04142-5.

58. K. Martinmäki, H. Kinnunen: R-R interval measurement and heart rate variability in polar products, Polar Electro Oy, Kempele, Finland, May 2011, White Paper.

59. J. Happonen: Polar sensor logger, 2019. Available online: https://play.google.com/store/apps/details?id=com.j_ware.polarsensorlogger

60. M. Vollmer: HRVTool – An open-source matlab toolbox for analyzing heart rate variability, in 2019 Computing in Cardiology Conference (CinC). Computing in Cardiology, December 2019. https://doi.org/10.22489%2Fcinc.2019.032.

61. M. Vollmer, A robust, simple and reliable measure of heart rate variability using relative RR intervals, in 2015 Computing in Cardiology Conference (CinC), IEEE, 2015, pp. 609–612.

62. L. Salahuddin, J. Cho, M.G. Jeong, D. Kim: Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, pp. 4656–4659.

63. M. Malik, A.J. Camm, J.T. Bigger, G. Breithardt, S. Cerutti, R.J. Cohen, P. Coumel, E.L. Fallen, H.L. Kennedy, R.E. Kleiger, F. Lombardi, A. Malliani, A.J. Moss, J.N. Rottman, G. Schmidt, P.J. Schwartz, D.H. Singer: Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. European Heart Journal, 17, 3 (1996) 354–381.

64. MATLAB version 9.10.0.1739362 (R2021a) Update 5, The Mathworks, Inc., Natick, Massachusetts. 2021.

65. Ipek: Normality test package, MATLAB Central File Exchange, 2021. Available online: https://www.mathworks.com/matlabcentral/fileexchange/60147-normality-test-package

66. O. Medvedev, D. Shepherd, M.J. Hautus: The restorative potential of soundscapes: A physiological investigation. Applied Acoustics 96 (2015) 20–26. https://doi.org/10.1016/j.apacoust.2015.03.004.

67. D. Ciuk, A.K. Troy, M.C. Jones: Measuring emotion: Self-reports vs. physiological indicators. SSRN Electronic Journal (April 2015) 1–29.

68. J.Y. Hong, B. Lam, Z.T. Ong, K. Ooi, W.S. Gan, J. Kang, J. Feng, S.T. Tan: Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications. Building and Environment 149 (2019) 1–14. https://doi.org/10.1016/j.buildenv.2018.12.004.

69. J.Y. Jeon, J.Y. Hong, C. Lavandier, J. Lafon, Ö. Axelsson, M. Hurtig: A cross-national comparison in assessment of urban park soundscapes in France, Korea, and Sweden through laboratory experiments. Applied Science 133 (2018) 107–177. https://doi.org/10.1016/j.apacoust.2017.12.016.

70. D.S. Quintana, J.A.J. Heathers: Considerations in the assessment of heart rate variability in biobehavioral research. Frontiers in Psychology 5 (2014) 805. https://www.frontiersin.org/article/10.3389/fpsyg.2014.00805.

71. S. Durbridge, D.T. Murphy: Assessment of soundscapes using self-report and physiological measures Dataset. GitHub, January 2023. Available online: https://github.com/sedurCode/ActaAcustica22Resources.