Correlations Between Acoustic Features, Personality Traits and Perception of Soundscapes.

PerMagnus Lindborg,[*1]

[*] *Nanyang Technological University, Singapore*
[#]*KTH Institute of Technology, Stockholm*

[1]`permagnus@ntu.edu.sg`

ABSTRACT

The present study reports results from an experiment that is part of Soundscape Emotion Responses (SSER) study. We investigated the interaction between psychological and acoustic features in the perception of soundscapes. Participant features were estimated with the Ten-Item Personality Index (Gosling et al. 2003) and the Profile of Mood State for Adults (Terry et al. 1999, 2005), and acoustic features with computational tools such as MIRtoolbox (Lartillot 2011).

We made ambisonic recordings of Singaporean everyday sonic environments and selected 12 excerpts of 90 seconds duration each, in 4 categories: city parks, rural parks, eateries and shops/markets. 43 participants rated soundscapes according to the Swedish Soundscape Quality Protocol (Axelsson et al. 2011) which uses 8 dimensions related to quality perception. Participants also grouped 'blobs' representing the stimuli according to a spatial metaphor and associated a colour to each.

A principal component analysis determined a set of acoustic features that span a 2-dimensional plane related to latent higher-level features that are relevant to soundscape perception. We tentatively named these dimensions Mass and Variability Focus; the first depends on loudness and spectral shape, the second on amplitude variability across temporal domains.

A series of repeated-measures ANOVA showed that there is are patterns of significant correlations between perception ratings and the derived acoustic features in interaction with personality measures. Several of the interactions were linked to the personality trait Openness, and to aural-visual orientation. Implications for future research are discussed.

I.   BACKGROUND

The present study is inspired by research in the fields of soundscape studies, music emotion, and acoustics. Describing everyday sonic environments, understanding how they affect people, and designing soundscapes for pleasure, improved health and efficient communication are weighty tasks that no one can take on alone. The Positive Soundscape Project (Davies et al. 2007, 2009, Adams et al. 2008), with its broad approach to cross-disciplinary research, involved acoustics, sound artists, social scientists, psychoacousticians, and experts in physiology and neuroscience. The authors hold forth that the multiple methods are complementary: for example, the ecological validity of soundwalk annotations complement the accuracy of measurements in a laboratory setting. A project with a community science and perhaps activist ambition is NoiseTube (Maisonneuve et al. 2009, 2012), aiming to focus resources that can support decision-making about the public sonic environment. The importance of soundscape research for architects and urbanists has been underlined by many authors e.g. (Cain 2007, Andringa 2010, and Jennings 2009).

Nilsson studied perception of the quality of sonic environments in city parks and suburban areas in Stockholm and found that "sound source identification… within the soundscape was… a better predictor of soundscape quality than measured sound levels" (Nilsson 2007). To provide a tool for the measurement of the perception of urban soundscapes, in particular the influence of traffic on quality, he and collaborators developed the Swedish Soundscape Quality Protocol (SSQP; Axelsson, Nilsson & Berglund, March 2011). It consists of 5 items and was designed to enable on-site surveys where minimal amount of time and attention can be expected from ad-hoc raters passing by.

Axelsson, Nilsson and Berglund (2010) investigated how people perceived recordings of soundscapes that had been categorised as predominantly 'technological', 'natural' or 'human'. The authors collected ratings on 116 unidirectional scales, using adjectives (such as lively, brutal, warm…) on 50 soundscapes. A principal component analysis led to a solution with 3 meaningful dimensions: pleasantness (50%) and eventfulness (16%) and familiarity. Hence SSQP includes an item where the quality of soundscapes is rated on 8 adjectival unidirectional dimensions, vector added together in a 2-dimensional circumplex model. They summarised the results as: "soundscape excerpts dominated by technological sounds were mainly perceived as unpleasant and uneventful, and soundscape excerpts dominated by human sounds were mainly perceived as eventful and pleasant". Axelsson developed analysis methods further in his thesis (2011), which presents a model where a latent variable, Information Load, is a key factor underlying aesthetic appreciation.

A pioneering work on perceptual-emotional qualities in music was made by (Wedin 1972), leading over the past decade to a wealth of research and computational methods. Examples include CUEX, focussing on onset detection in music (Friberg et al. 2007, 2012); MIRtoolbox, with a broad range of spectral and other features (Lartillot et al. 2008, Lartillot 2011); and CataRT, running in real-time as an integrated system for analysis and performance (Schwartz, . For an overview, see the MIREX webpages. Much of the knowledge gained in music can be applied to soundscapes.

It seems obvious that people perceive sounds in different ways, but what kinds of patterns are involved? Vouskoski and Eerola (2011, 2012) investigated individual differences in emotional processing, specifically the role on personality and mood in music perception, and preference ratings. They hypothesised that both personality and mood would contribute to the perception of emotions in trait- and mood- congruent manners, and that mood and personality would also interact in producing affect-congruent biases. The authors investigated how mood may moderate the influence of personality traits on

emotion perception in excerpts of film music which had been evaluated in a pilot experiment according to perceived basic emotion in five categories (anger, fear, happiness, sadness and tenderness). They concluded that "the degree of mood-congruence in the emotion ratings is at least to some extent moderated by personality traits". The idea with the authors' analysis method was to parcel out the variability of short-term mood swings from those of the more persistent personality traits. Among other things, they found significant correlations between ratings of perceived happiness in the music with vigorous mood state, interacting with extrovert personality. They also found a correlation between vigour and happiness ratings that increased with increased extraversion.

To gauge their subjects' mood, Vouskoski used The Profile of Mood States, in a version adapted for use with adults, (POMS; Terry et al. 1999, 2003). It is a questionnaire with a single instruction: "Mark the answer which best describes how you feel right now", followed by 24 adjectives. The subject answers by marking a point on a 5-point Likert scale anchored by "Not at all" and "Extremely well". The adjectives include "muddled", "alert", "nervous" and so forth. A score is calculated for each of 6 mood dimensions: Anger, Confusion, Depression, Fatigue, Tension and Vigour. POMS has been reported to have good concordance with other measurement instruments (Morfeld et al. 2006). Mood states are considered relevant for the current or same-day emotional state.

Personality traits are considered to be stable over longer duration, even over life spans. The Big Five (John & Srinathavan 1999, McCrae & Costa 1997) is a well-established model. For our study, we used the Ten-Item Personality Index (TIPI), which is a "light-weight version" developed by Gosling et al. (2001, 2005). TIPI is headed by a single statement, the self-reflecting "I see myself as…" and then lists 10 pairs of adjectives, such as "anxious, easily upset" or "extroverted, enthusiastic". The subject answers by marking a point on a 7-point Likert scale anchored by "Disagree strongly" and "Agree strongly". TIPI produces a score in 5 personality dimensions, named Openness, Conscientiousness, Extrovertedness, Aggressivity, and Emotional Stability. The last one is simply the reverse of the more common Neuroticism, part of the OCEAN models. Gosling has shown that TIPI has a good construct validity when compared with both the 42-item BFI (used by Vouskoski) and larger instruments. It is less specific but has an advantage when experiment designs allot little time.

The Soundscape Emotion Response (SSER) study is part of a research aiming to chart people's responses to everyday soundscapes in different modalities: perceptual, physiological, movement, colour association and verbal commentary. Our research is localised to Singapore, a fast-developing city where people's general attention to the quality of sonic environments has low priority. The long-term aim is to bring knowledge as a resource for architects, urban planners, and decision-makers. One area of application is to contribute to improved learning efficiency in local schools. The singular study in a Singaporean context related to this matter (Nyuk & Wy, 2003) calls for further research into soundscape perception in an urban, tropical environment.

In short, the aims of our present study have been to investigate the interaction between psychological and acoustic features in the perception of soundscapes, using established analytical tools.

II.   METHOD

A. Soundscape Stimuli

We made ambisonic recordings of Singaporean public spaces using a TetraMic (#2144) onto a SoundDevices 788t mobile recording device. For inclusion in the SSER study, soundscapes would have to be sufficiently long for the psycho-physiological study (not part of the present text) and have high intra-stimulus homogeneity. The collection as a whole had to be large and varied enough to represent a meaningful sample of Singaporean environments, in some sense "everyday". The choice was also influenced by the way a soundscape can be understandable as representing a physical environment in the absence of all other sensorial information. The size of the set was limited so that the experiment could be completed in one hour. The final selection consisted of 12 excerpts of 90 seconds duration, in four "a priori" categories roughly dividing the set into rural and urban parks, eateries, and places to shop.

In parallel with the audio recordings, we took SPL measurements at several of the locations with an Extech 407790 using the LeqP(30s) un-weighted setting. The B-format files were transformed into binaural (HRTF KEMAR1) with Harpex-B for the perceptual ratings using studio quality headphones. Acoustic feature computation was performed on the W (mono omni) channel at +3 dB. LeqA and other loudness measures were calculated from on-site LeqP when available, or by comparing the dBFS level to that of a recording with known LeqP. See Table 1 below for the main features of the set of soundscapes used as stimuli in the present study.

Table 1. Overview of the SSER soundscapes.

name	description	LeqA
market	Little India Market, large & old-style, covered roof, medium crowded weekday afternoon, butchers chopping meat	76
hawker	Queen Street Block 270 Foodcourt, large & worn-down, covered roof, very crowded, fans, scraping plates	73
construct	Queen Street, small concrete & grass parl, benches with 5 pax, 1 child playing, diesel generator nearby	77
café	Café "Food for Thought" Waterloo street, aircon, chairs scratching floor, churchbells in distance	69
bolly	Bollywood Veggie, tropical park, no people weekday afternoon, airplanes & diesel pump in distance	52
night	Little India small park near Bagdad Street, benches with 2 pax, some passers-by, bar music in distance	55
resto	Bussorah Road street-side restaurants, very crowded Friday night, dense slow-moving traffic	69
shop	Golden Mile Mall, mainly Thai shops, steady stream of people near escalators, child bouncing a ball, dense traffic in distance	73
oriol	Sungei Buloh nature reserve, mangrove, no people in early morning, 2 oriols singing	56
water	Sungei Buloh nature reserve, water sounds, small waves, no people, facing Johor Bahru (city) in distance	58
crowd	Vivo City Mall outside Golden Village Cinemas, near escalators, extremely crowded Saturday night	84
people	Vivo City Mall rooftop open area towards Sentosa, crowded, groups of young people laughing, chatting	74

B. Participants

Participants (N=43) were university students in Singapore (N=32) and Norway (N=11) currently enrolled in music or sound-related courses, and took part in the study as part of course requirements.

For the present study, we developed a screen-based interface developed in MaxMSP, running on individual computers in a school lab with up to 18 participants at each time. Neither lab (Singapore or Norway) was acoustically isolated, but generally suitable for sound work. There was no loud intermittent sound disturbance at any time (e.g. rain or construction noises). The computer sound output had been adjusted to a predetermined level (three clicks below maximum), same for all participants. Identical studio quality circum-aural headphones were used, but no measure of the actual SPL in the ear canal could be taken. Neither were we able to conduct a test to gauge if any hearing impairment was present.

Participants were given verbal information about the experiment before starting, and the same text was displayed on the computer screen. Participants completed three tasks: provide information about themselves, rate each soundscape, and group the 12 soundscapes by similarity.

C. Procedure

In the first part, participants anonymously filled out forms for a) general participant data (GPD): age, gender, handedness, language and race (5 items); b) activities, i.e. the typical number of hours per day & night spent resting, sleeping, working or studying, making music, doing visual arts, watching TV or playing video games, doing sports or socialising (7 items); c) the relative importance they accorded the five senses; d) the Ten-Item Personality Index (TIPI) as above, but with adjectives in randomised order, and e) the Profile of Mood State for Adults (POMS), as above, but with adjectives in randomised order. The participants were free to complete questions at their own pace, taking typically 10-12 minutes for this part.

In the second part, the participants donned studio quality headphones (and thus became raters). There were two tasks: 'grouping' the 12 soundscapes, and 'rating' each soundscape. The interface for the grouping task consisted of a 'white-square' of apparent screen size very close to 100mm x 100mm, containing 12 anonymous, at first grayish 'blobs'. By clicking on one, the associated soundscape started playing at a random position in the soundfile (smoothly looping at the end-of-file). The blob could be dragged to a position anywhere in the white square; see Figure 2. The written instructions were: "1) click on a blob to listen to its soundscape; 2) doubleclick to open, and rate how you perceive the soundscape; 3) drag blobs around to group similar soundscapes next to each other." There was no further explanation given how to interpret the word "similar", even though some participants asked.

By double-clicking a blob, the user could open a 'rating' interface for that soundscape, as seen in Figure 1. The participants entered ratings by adjusting horizontal continuous sliders with apparent screen length very close to 100mm and marked by labels as in SSQP. 3 items were employed:

"To what extent do you presently hear the following 5 types of sounds?" [Traffic noise, Fan noise, Other noise, Sounds from human beings, Natural sounds] The order of the 5 types/categories was vertically randomised for each rater and soundscape, and the sliders had equidistant labels [Do not hear at all, A little, Moderately, A lot, Dominates completely]. In the analysis, we refer to the results as a 5-dimensional variable 'content'.

"Overall, how would you describe the present surrounding sound environment?" The slider had equidistant label [Very good, Good, Neither good, nor bad, Bad, Very Bad]. In the analysis, we refer to the results as a variable 'overall'.

"To what extent do you agree with the 8 statements below on how you experience the present surrounding sound environment?" [pleasant, exciting, eventful, chaotic, annoying, monotonous, uneventful, calm] The order of the 8 adjectives was vertically randomised for each rater and soundscape, and the sliders had equidistant labels [Agree completely, Agree largely, Neither agree, nor disagree, Disagree largely, Disagree completely].
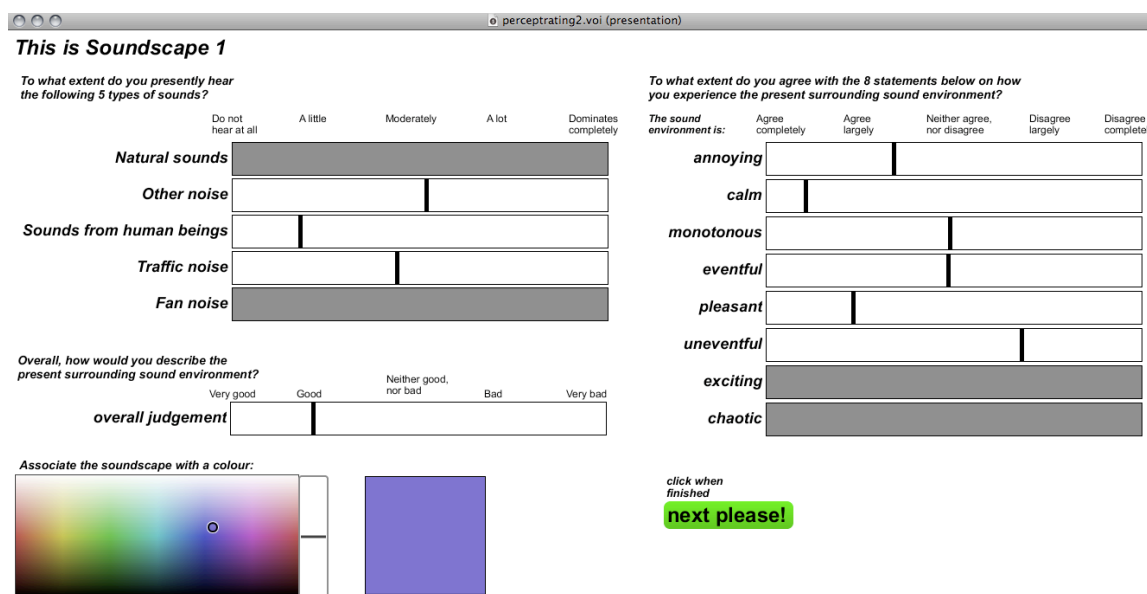

Figure 1. The 'rating' task interface.

SSQP has 2 more items, but as they aim to capture the relationship between soundscape and visual elements of the landscape, they were not employed in this study.

Instead, the rater was asked to "Associate the soundscape with a colour:" using a Hue-Saturation-Luminosity (HSL) colour picker. We were inspired to include colour by (Bresin 2005). Whereas he used a set of predetermined colour patches, our interface had a swatch interface allowing quasi-continuous choice of colours. When all perceptual features had been rated, this colour became visible in the white-square 'grouping' interface (see Figure 2), and when all soundscapes had been rated and grouped (it typically took 50 minutes), the second part was completed, and the experiment as a whole.

Finally, the software made some behind-the-scenes statistics of rater behaviour. The position in the soundfile when a slider was set was registered, because it was suspected that a marked event in a soundscape could cause raters to focus on that, and if significant, this would lead us to question intra-stimulus homogeneity. It also calculated the total amount of time in seconds that a rater spent listening to each soundscape. We speculated that it could be as a measure of attention.
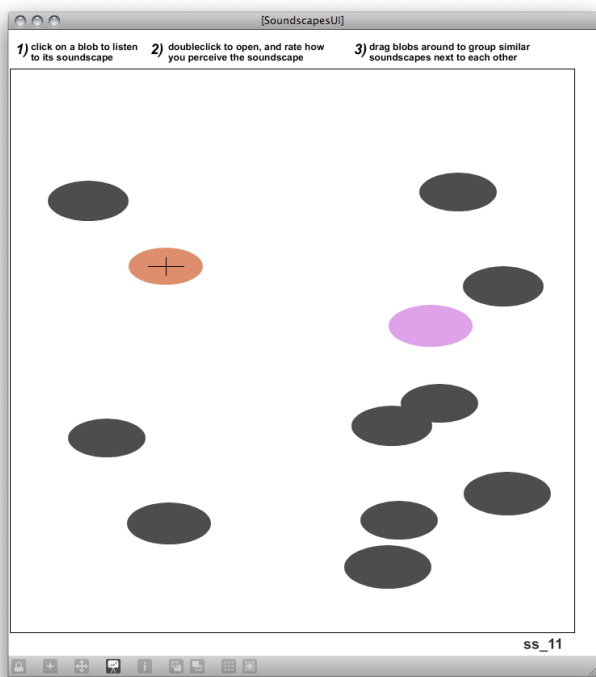

Figure 2 The 'grouping' task interface. Ratings for two soundscapes have been fully completed.

III. RESULTS

A. Participants

The mean age of the 43 participants was 22 years, with most Norwegians being 19 and Singaporeans being between 19 and 26 years old. 33 participants were women (8/11 Norwegians, 25/32 Singaporeans) but the gender imbalance was not significant ($X^2(42) = 6.23$, p=1). Nevertheless results that depend upon gender should be interpreted carefully. We will discuss four aspects of the participants' profiles:

activities, senses, personality traits and mood, and argue that the sample is useable for the analyses undertaken.

A one-way ANOVA with group (Singaporean or Norwegian) as dependent variable against all other participant data showed that the two groups differed significantly at the two-tailed alpha=0.05 level in three regards: age ($F(1,26)=38.2$, p=0.000); the amount of music-making ($F(1, 26)=11.9$, p=0.003); and the amount of time spent watching TV or playing video games ($F(1, 26)=6.80$, p=0.019). The large difference in music is explained by the fact that all the Norwegian participants were students at a conservatory. The difference in watching (Singaporeans twice as many hours as the Norwegians) might be explained by the fact that a large portion of the Singaporean participants were students at a school predominantly for the visual arts. On a typical day, the participants spent 7.9 hours resting, doing work or study 5.5 hours, and socialising 3.8 hours. Sports occupied the participants for 1.3 hours. Doing visualarts, i.e. drawing or painting, clocked in at less than one hour, though with a few Singaporeans spending up to 6 hours daily; however, the group mean difference was not significant ($F(1, 26)=4.07$, p=0.061). As expected, participants self-reported sight as their foremost sense, followed by hearing and touch, then taste and smell. The latter pair correlated strongly, at r=0.66, in line with e.g. (Lindborg 2010a).

More importantly, the ANOVA revealed that there was no significant group difference in regards to any of the personality (TIPI) or mood (POMS) measures; the closest miss was for Emotional Stability ($F(1, 26)=2.15$, p=0.16). This lends support to the assumption that TIPI and POMS are useable in the following analysis. Figure 3 shows the TIPI means with confidence intervals, between groups as well as with normative data from (Gosling 2003).
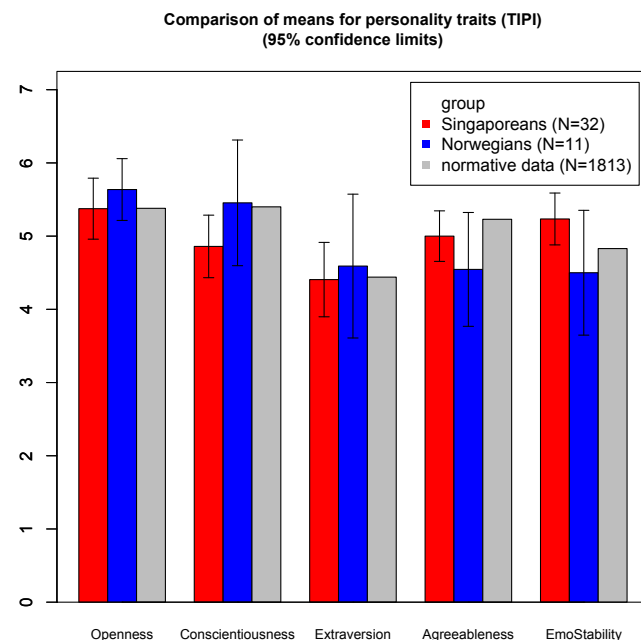

Figure 3. Barplots of TIPI with normative data.

Within TIPI, no correlation was significant and the mean of pairwise correlations was 0.071. The low value indicates that the 5 dimensions are reasonably orthogonal in our

sample. Within POMS, 5 out of the 15 pairwise correlations within the submatrix were significant: anger/depression, anger/tension, confusion/tension, confusion/depression and anger/confusion (r values ranging from 0.62 to 0.41). The mean of pairwise correlations was reasonably low, mean r=0.30. However, the prevalence of covariation indicates that current mood state may not be as well measured by the instrument as one could wish for. Results depending on mood state would have to be carefully considered. Looking at covariation between TIPI and POMS measures, Emotional Stability (the inverse of Neuroticism) correlated negatively and significantly with depression (r=-0.58), confusion (r=-0.50) and anger (r=-0.48).

The mean pairwise correlation of all ratings across soundscapes was 0.645 (Cronbach's alpha was 0.978). The level of agreement can be considered reasonably high.

B. Ratings

1) Grouping task: blobs and colours

There are two ways of understanding the way participants placed 'blobs' inside the 'white square' in the interface: as absolute or relative (i.e. dissimilarities). We will first discuss absolute positions. Since no instructions were given as to how participants ought to interpret the 'grouping' task or the 'white square', we had no hypothesis. As a measure of spread (pos.d2m), we took the Euclidian distance to the middle of the distribution.

Visual inspection of histograms of the spread in different soundscapes indicates that the distributions are reasonably normal and that the mean is representative. The most platykurtic distribution is for soundscape 11 ('crowd'), perhaps indicating a wider range of listening attitudes to this recording than the others. We speculate that some raters might be paying more attention to individual voices of people in the crowd, while others to the crowd as a whole, i.e. their listening strategy focussed either on foreground or background elements.

The colour association for each rating was transformed from RGB to CIE L*ax*b (henceforth Lab) space using default values for white point, standard illuminant D65, "sRGB" source space for the Norwegian group using PC monitors, and "Apple RGB" source space for the Singaporeans. The three dimensions of the Lab space are considered perceptually linear and orthogonal, which means that the difference between colour shades can be estimated computationally. Specifically, the dissimilarity between two given colours can be expressed as the Euclidian distance in Lab space. As a measure of colour spread (col.d2m), we took the Euclidian distance to the middle of the distribution.

2) Ratings task - soundscape quality

The biplot in Figure 4 shows the principal component solution of quality ratings, revealing a neat structure and clearly supporting the results in Axelsson (2011). All adjectives appear in opposite pairs, as predicted by theory. PC1 explains 47% of the variance in the ratings, and PC2 26%. The 2-dimensional model places the 8 dimensions evenly in a circumplex. In our present results, the distribution appears somewhat tilted counter-clockwise, but more important is to analyse how the dimensions relate to each other.

We observe that the 8 unipolar adjectival dimensions appear as bipoles, i.e. as opposing pairs on a "arrows on a line". Theory predicts that 'chaotic' should be half-way
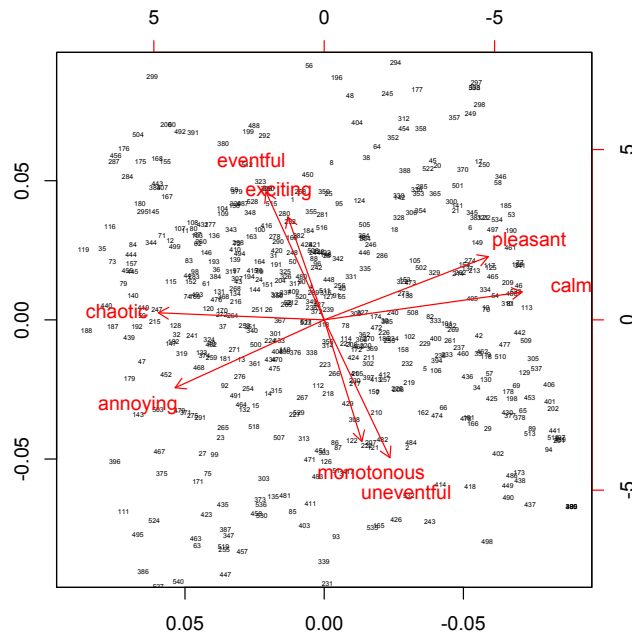

Figure 4. Biplot of the PCA of SSQP quality ratings.

between 'unpleasant' and 'eventful', and this is almost true in our results, albeit slightly nearer to 'unpleasant'. Its opposite,'calm', is resolved to a lesser extent, between 'pleasant' and 'uneventful'. The bipolar dimension 'monotonous-exciting' is not resolved from 'uneventful-eventful', and we note that their loadings are smaller too. Bringing in more dimensions, we see that in the PC1-PC3 plane, 'monotonous' appears orthogonal to 'uneventful', and projections with PC4 resolve 'exciting' nearly orthogonally from 'eventful'.

Based on the model, we directly calculate positions for each rating as:

Pleasantness = $\sum \text{rating}_A * \cos(2\pi * N_A/8)$

Eventfulness = $\sum \text{rating}_A * \sin(2\pi * N_A/8)$

The mean ratings of soundscapes across participants are shown in Figure 5.
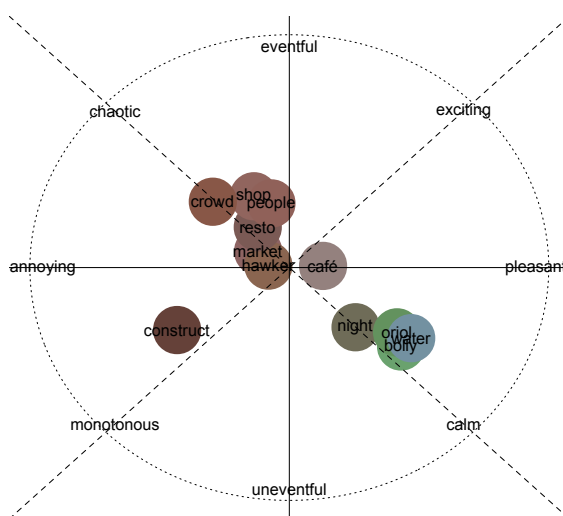

Figure 5. Plot of SSER soundscapes in SSQP Pleasantness-Eventfulness, with mean Lab colour.
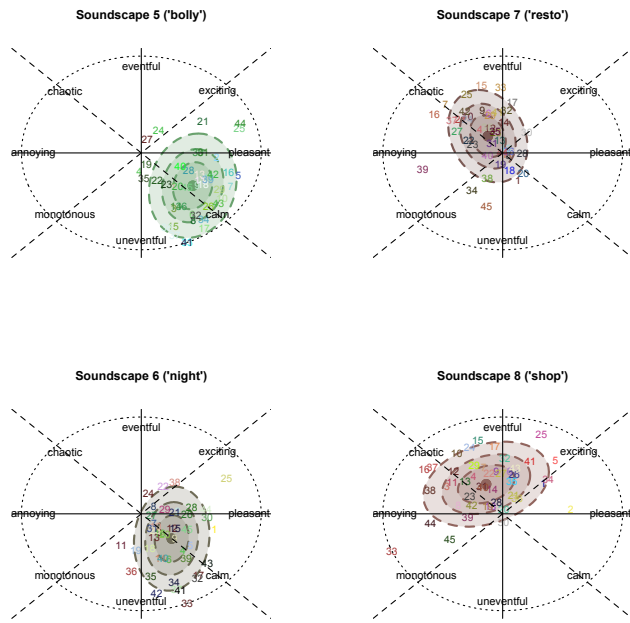
Figure 6. The distributions of N=43 quality ratings for Soundscapes 5-8 in SSQP Pleasantness - Eventfulness.

C. Acoustic features

To find candidates for computing acoustic features, we investigated MIRtoolbox (Lartillot 2011) and Genesis (2009), and a script for ITU loudness from (Nygren 2009). We identified 24 functions that could[3] be meaningful for measuring soundscapes in terms of their loudness, spectral shape, but also "rhythm", which should be thought of not so much in musical terms, but rather as the rate and distribution of events in soundscapes. With only 12 target soundscapes, the rule of thumb says that we should not use more than 2 dimensions to map them, or else risk over-fitting in the succeeding analysis. In order to reduce the number, we tried different sets of candidate features and tested them with principal component analysis. A final selection is listed in Table 2.

Table 2. Overview of the SSER soundscapes.

feature	unit	description
N10	sone	Fastl & Zwicker's model for loudness of "foreground events" i.e. exceeded 10% of the time. Used by Axelsson.
CminusA	dBSPL	"measure of the relative proportion of low-frequency sound" (Nilsson 2007). Used by Axelsson.
zeroXrate	rate	mirzerocross(), rate of zero crossings, "a simple indicator for noisiness" (Lartillot p. 103).
rolloff	Hz	mirrollof(), the frequency below which 85% of the signal energy is contained (Tzanetakis and Cook 2002, Lartillot p. 78).
spectspread	Hz2	mirspread(mirspectrum()), i.e. variance of the Long-Term Average Spectrum.
N10m90	sone	"indicator of the soundscape variability", the relative amount of louder events, calculated as N10-N90 (see above). Used by Axelsson.
eventdens	Hz-1	mireventdensity(), calculated from mironsets(), with default settings.
tempo	BPM	mirtempo(), in the range {12… 20} BPM (= {0.2…20} Hz).

Inspection of the histograms revealed that spectspread and tempo are rather skewed. It makes sense to take the logarithmic transform: spectspread is a measure in the frequency domain, which perceptually is logarithmic; and tempo (in our case, very slow and regular pulsation), being in the time domain, can also arguably be perceived as logarithmic. See Table 3 for cross-correlations.

Table 3. Correlations between the selected acoustic features.

	N10	spectspread	zeroXrate	CminusA	rolloff	N10m90	eventdens
spectspread	-0.95						
zeroXrate	0.34	-0.30					
CminusA	-0.62	0.67	-0.70				
rolloff	-0.53	0.62	0.45	0.09			
N10m90	0.53	-0.67	0.45	-0.75	-0.22		
eventdens	-0.10	0.14	-0.58	0.56	-0.19	-0.50	
tempo	0.19	-0.19	-0.35	0.03	-0.50	-0.18	0.13

The first two dimensions of a principal component decomposition explain 72.7% of the variability in the data. Because PCA is invariant under rotation and mirroring, we can manipulate the solution and make it easier to interpret. By choosing N10 as a reference for rotation, we identify two meaningful components in Figure 7.
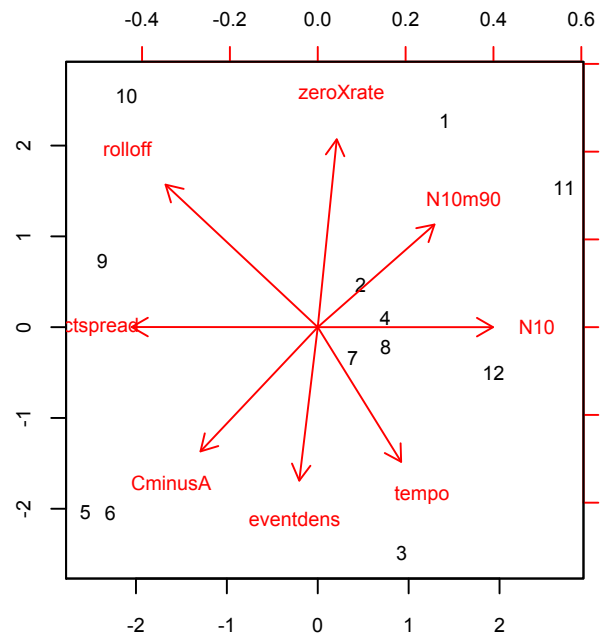

Figure 7. Biplot of the rotated solution.

The first dimension in the rotated solution, PC1', reveals the diametrical opposition between spectspread and N10 for the soundscapes in our study. spectspread is clearly a spectral measure. Could it be that Zwicker's model, based on human perception, effectively has a frequency response curve whose "curvature" increases with higher SPL? Consider also the shape of Fletcher-Mundson equal loudness curves at low and high loudness levels. We tentatively label this dimension Mass. We may think of sounds with large mass as being present and earthy, and sounds with low mass as evanescent and whispery.

The second dimension, PC2', opposes the zero-cross measure, which increases with high-frequency noisiness,

with event density, which detects relatively slow-rate amplitude variation. We tentatively label it Variability Focus. Variability refers to changes, i.e. rhythm and variation across loudness and pitch domains, and Focus refers to the time domain of the listening context, i.e. from slow and global, to fast and local. We may think of sounds with high Variability Focus as sizzling or whizzing, and sounds with low Variability Focus as thumping or booming.

Between these two new dimensions appear quite neatly two pairs of acoustic features, measuring opposing constructs. CminusA estimates the energy difference between C- and A-weighted Leq. We may reason that this difference should be smaller for sounds of smaller Mass because their energy is spread out and resemble the C-weighting curve, and at the same time the relative dominance of low frequencies gives room for slow periodicity, that is, lower Variability Focus. N10m90 measures the relative amount of shorter, louder (more massive) sounds, against the background. We may speculate that sound events "stick out" from the background when they have higher pitch, and have higher zero-cross count. Tempo is quite naturally covarying with low timescale Variability Focus, but may also be more prevalent, and easier to detect for the algorithm, for louder (more massive) sounds. Finally, rolloff quite naturally covaries with both spectspread and zero-cross count, as all 3 measures are likely to increase with sounds that have lots of high-frequency energy.

We tentatively adopt the Mass - Variability Focus plane for acoustic features, and calculate coordinates for the 12 soundscapes.

D. Repeated Measures Analysis of Variance

The dependent variables have been discussed: overall, timespent, Pleasantness, Eventfulness, pos.d2m, and col.d2m. Looking at the correlation matrix in Table 3, we suspected that analyses on overall and Pleasantness would be likely to show similarities.

Table 3. Correlations between dependent variables.

	overall	Pleasantness	Eventfulness	pos.d2m	col.d2m
Pleasantness	0.79				
Eventfulness	-0.29	-0.29			
pos.d2m	0.09	0.11	-0.12		
col.d2m	0.22	0.22	-0.11	0.07	
timespent	-0.06	-0.01	0.03	0.05	-0.02

For the independent variables describing psychological features, we included auralvisual orientation, TIPI (5 dimensions) and POMS (6 dimensions), and for the independent variables describing stimuli features the derived acoustic components: Mass and Variability Focus. We are aware that employing 11 measures for 43 participants risks overfitting; however, the fact that TIPI and POMS are well established instruments makes their inclusion reasonable. The results in Table 4 includes interactions with $p<0.01$.

Table 4. Significant interactions.

percept (dv)	rater feature	stimulus feature	F(1, 449)	p	characteristic associated with more broadly rated differences in the percept feature (dv)
overall	Openness	Mass	313.8	0.000 ***	more Openness
overall	aural-visual	Mass	122.7	0.000 ***	towards aural orientation
overall	Openness	Variability Focus	27.7	0.000 ***	less Openness
overall	aural-visual	Variability Focus	17.6	0.000 ***	towards aural orientation
overall	confusion	Mass	9.1	0.0027 **	less confusion
overall	vigour	Mass	8.4	0.0039 **	low vigour
Pleasantness	Openness	Mass	357.5	0.000 ***	more Openness
Pleasantness	aural-visual	Mass	184.1	0.000 ***	towards aural orientation
Pleasantness	Openness	Variability Focus	31.1	0.000 ***	less Openness
Pleasantness	aural-visual	Variability Focus	27.6	0.000 ***	towards aural orientation
Pleasantness	fatigue	Mass	9.1	0.0027 **	more fatigue
Pleasantness	confusion	Mass	6.7	0.0098 **	less confusion
Eventfulness	aural-visual	Mass	190.0	0.000 ***	more towards aural
Eventfulness	Openness	Mass	117.6	0.000 ***	less Openness
Eventfulness	Conscientiousness	Mass	10.4	0.0013 **	high Conscientiousness
Eventfulness	Emotional Stability	Mass	9.8	0.0019 **	high Emotional Stability
position spread	Openness	Mass	9.4	0.0023 **	less Openness
colour spread	Openness	Mass	13.6	0.000 ***	less Openness
time spent	Openness	Variability Focus	12.0	0.000 ***	less Openness

From the ANOVA results we can make a number of inferences. Note that we are only discussing interaction effects. In the plots below we have centered all variables and divided the participants in three groups with different level on the independent rater feature. For example, the blue line refers to the 33% with lowest vigour, the gray to the middle-vigour 33%, and the red to the 33% with highest vigour score.

1) Overall and Pleasantness

People who were more open-minded perceived larger quality differences in terms of soundscape Mass, but less quality differences in terms of Variability Focus. Those who paid more attention to the sense of hearing generally perceived larger quality differences. The same situation holds for the compound rating of Pleasantness, as expected from Table 3.

See Figure 8 for a plot of the interaction between soundscape Mass and Openness, in their correlation with Pleasantness ratings . To get a high score on TIPI Openness, the respondent would tend to see herself as more open to new experiences, more complex, less conventional and less uncreative. To explain the results, could it be that Openness acts as a moderator in the relationship between soundscape Mass and the perception of Pleasantness? and the same for overall quality.

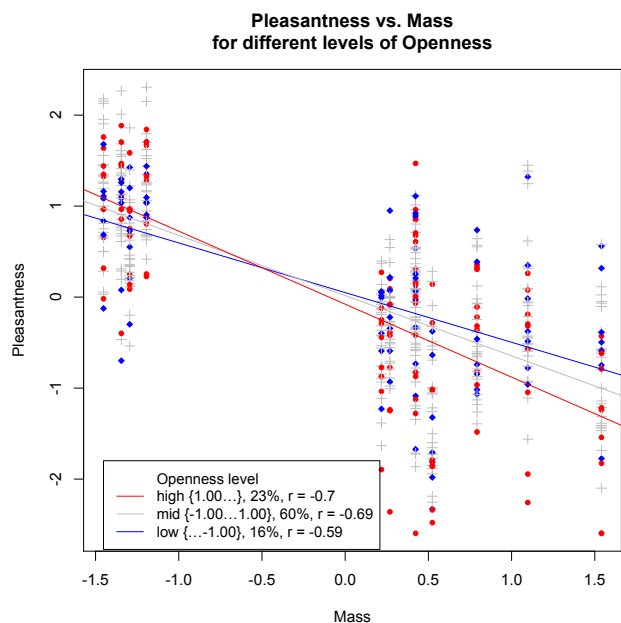The interaction effects involving mood states are more difficult to interpret.

Figure 8. Interaction plot of Pleasantness ~ Mass : Openness

2) Eventfulness

People who paid more attention to the sense of hearing perceived larger differences in Eventfulness of sound Mass. Since our definition of Mass is based on acoustic feature detection of (low-frequency) foreground events, this result confirms that aurally oriented people are more able to tell the difference. See Figure 9 for an interaction plot.

In Eventfulness, open-minded people perceived lesser differences in terms of soundscape Mass. This could be explained by the fact that the two dimensions are

(theoretically) orthogonal, and we are looking at a very small effect.

Finally, high Conscientiousness, Emotional Stability and vigour were all associated with more discriminating ratings towards soundscapes when spaced along the Mass dimension.
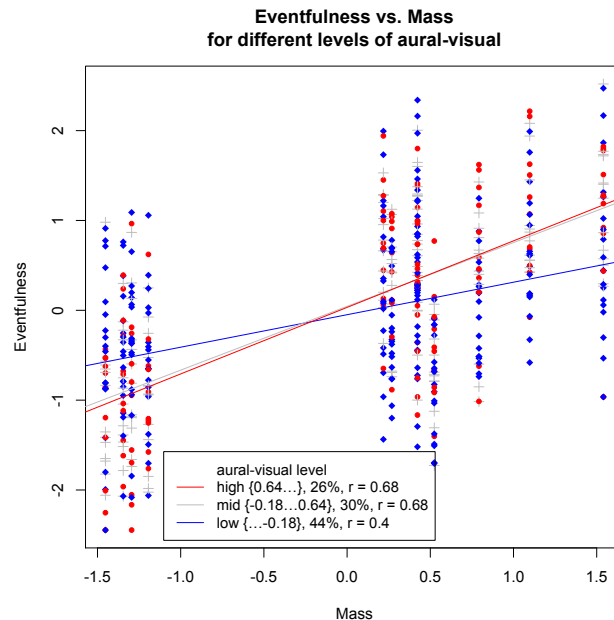

Figure 9. Interaction plot of Eventfulness ~ Mass : auralvisual

3) Blob Position Spread and Colour Spread

People who scored low on Openness showed a tendency to make larger differences in the amount of spreading the blobs out in the 'white-square' interface as a response to soundscape Mass. In parallel, the same pattern applies to their broader useage of the colour palette. To get a low score on TIPI Openness, the respondent would tend to see herself as less open to new experiences, less complex, more conventional and more uncreative. It is remarkable that such a self-image is associated with larger spread in terms of blob position and colours.

4) Time Spent Listening

Again in a similar way, people scoring low on Openness showed a tendency to make larger differences in the amount of time spent on listening in regards to Variability Focus. One may speculate that they were "spreading out" their time in the same way they spread out blobs and colours. See Figure 10 for an interaction plot.

IV.   CONCLUSION

We used the Ten-Item Personality Index, the Profile of Mood States protocol, and a simple measure for aural-visual orientation as psychological features of the raters. Our results from the ratings of soundscape perception along the 8 adjectival dimensions of the Swedish Soundscape Quality Protocal lend support to the 2-dimensional model proposed by Axelsson, Nilsson & Berglund. We found patterns in the way raters made colour associations to soundscapes, and the way they grouped soundscapes according to a spatial metaphor.

We used principal component analysis to determine a set of acoustic features that span a 2-dimensional plane related to latent higher-level features that would be specifically relevant to soundscape perception. We tentatively named these new dimensions Mass and Variability Focus; the first depends primarily on loudness and spectral shape, the second on the relative prominence of amplitude variability at either end of the spectrum.
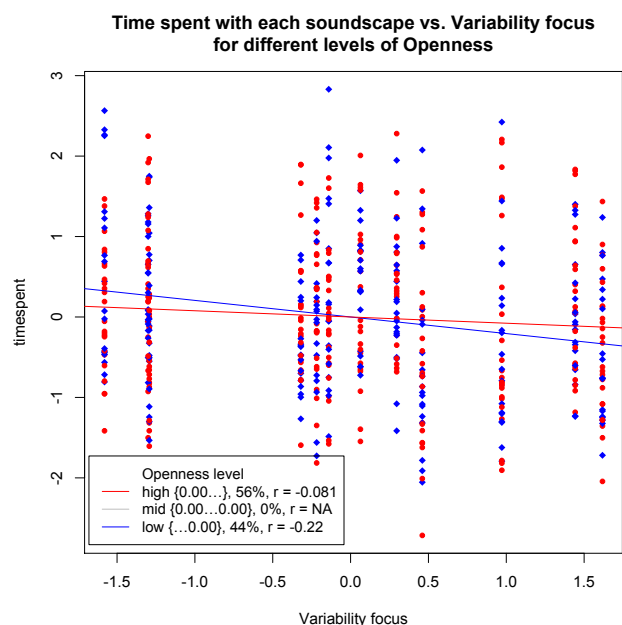

Figure 10. Interaction plot of Timespent ~ Variability Focus : Openness

A series of ANOVA revealed patterns of significant correlations between perception ratings and the derived acoustic features in interaction with personality measures. Several of the interactions were linked to the personality trait Openness and others to aural-visual orientation; Conscientiousness and Emotional Stability were also represented.

Cross-correlation and analysis of mean pairwise correlation indicated that the TIPI dimensions were relatively independent and more reliable than the POMS measures. The interaction results involving personality traits were relatively straightforward to interpret, while those with mood were not. We conclude that personality traits are more promising in regards to soundscape perception studies.

When it comes to the perceptual ratings as a whole, we aim to pursue the analysis by using multidimensional scaling (MDS) of overall (1D), blob position (2D), colour (3D), content (5D) and quality (8D). These 5 dissimilarity matrices can be cross-correlated, to reveal further patterns in the way participants use space, colour and semantic descriptions of soundscapes.

When it comes to the acoustic features, it remains to be shown if soundscape perception can be adequately described with a 2-dimensional model, perhaps with dimensions such as the proposed Mass and Variability Focus. In future work, we will certainly need larger numbers of soundscape stimuli, perhaps of shorter duration, and a larger group of rater-participants. It is also possible to investigate computational methods for feature selection optimisation, as discussed in (Eerola, Lartillot & Toivainen 2009), who suggest using

Projection to Latent Structure (PLS) regression in this kind of situation. PLS is a multivariate regression whereby a predictor-variable space is projected onto a space with smaller dimensionality. The problem with 'number crunching' methods in general is that the explanatory dimensions are hard to grasp, and it becomes a challenge to express their meaning verbally in simple terms.

APPENDIX

SSER soundscape stimuli and raw response data are available at http://www.permagnus.net.

REFERENCES

Adams, M., Bruce, N., Davies, W., Cain, R., Carlyle, A., Cusack, P., Hume, K., Jennings, P. & Plack, C. (2008). "Soundwalking as methodology for understanding soundscapes" In Proceedings of the Institute of Acoustics Spring Conference 2008 – Widening Horizons in Acoustics, Reading UK, April 2008, pp 552-558

Andringa, Tjeerd C. (2010). "Soundscape and core affect regulation". Proceedings of Inter-noise 2009, Portugal.

Axelsson, Östen (2011). Aesthetic Appreciation Explicated. PhD Thesis, Stockholm University.

Axelsson, Östen, Nilsson, Mats E. & Berglund, Birgitta (2010). "A principal components model of soundscape perception". Journal of the Acoustical Society of America #128 (5), November 2010.

Bodin, Lars-Gunnar (2012). "Meaning and Meaningfulness in Electroacoustic Music". Keynote address to Conference of Electroacoustic Music studies (EMS) 2012, Stockholm.

Bresin, Roberto (2005). "What is the color of that performance?". International Computer Music Conference (ICMC), Barcelona, 2005.

Cain, R., P. Jennings, J. Poxon, A. Scott (2009), "Emotional dimensions of a soundscape". In Proceedings of InterNoise 2009, 23-26th August, Ottawa, Canada.

Cain, R., P. Jennings, M. Adams, N. Bruce, A. Carlyle, P. Cusack, W. Davies, K. Hume and C. Plack (2008), "SOUND-SCAPE: A framework for characterising positive urban soundscapes", In Proceedings of Acoustics 08 – Euronoise, the European conference on noise control, Paris France, June 2008, pp 1019-1022

Cain, R. & P. Jennings (2007), "Developing best practice for lab-based evaluations of urban soundscapes", In Proceedings of Inter-Noise 2007, Istanbul, August 2007.

Davies, W. and M. Adams, N. Bruce, R. Cain, A. Carlyle, P. Cusack, K. Hume, P. Jennings, C. Plack (2007), "The Positive Soundscape Project", In Proceedings of the 19th International Conference on Acoustics, Madrid, September 2007.

Davies, William J. and Adams, Bruce, Marselle, Cain, Jennings, Poxon, Carlyle, Cusack, Hall, Hume & Plack (2009). "The positive soundscape project: A synthesis of results from many disciplines". Proceedings of Inter-noise 2009, Canada.

Eerola, Lartillot & Toivainen (2009). Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. 10th International Society for Music Information retrieval Conference (ISMIR 2009).

Friberg, A., Schoonderwaldt, E., & Juslin, P. N. (2007). CUEX: An algorithm for automatic extraction of expressive tone parameters in music performance from acoustic signals. Acta Acustica United with Acustica, 93, 411-420.

Genesis S.A. (2009). Loudness Toolbox, for MatLab. http://www.genesis.fr (accessed 14 May 2012).

Gosling, Samuel D., Rentfrow, Peter J. & Swann Jr., William B. (2003). "A very brief measure of the Big-Five personality domains". Journal of Research in Personality 37 (2003) 504–528.

Jennings, P. & Cain, R. (2009), "A Framework for assessing the change in perception of a public space through its soundscape", In Proceedings of InterNoise 2009, 23-26th August, Ottawa, Canada.

John, Oliver P. & Srivastava, Sanjay (1999). "The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives". Chapter 4, pp. 102-38 in Handbook of Personality. Theory and Research. 2nd edition. Pervin, Lawrence A. & John, Oliver P. (Eds). The Guilford Press 1999.

Lartillot, Olivier, Eerola, Tuomas, Toiviainen, Petri & Fornari, Jose (2008). "Multi-feature modeling of pulse clarity: Design, validation, and optimization", International Conference on Music Information Retrieval, Philadelphia, 2008.

Lartillot, Olivier (2011). MIRtoolbox 1.3.2, for MatLab. http://www.mathworks.com/matlabcentral/fileexchange/24583-mirtoolbox (accessed 14 May 2012).

Lindborg, PerMagnus (2010b). "Aural and Visual Perceptions of a Landscape". Unpublished pilot study.

Lindborg, PerMagnus (2010a). "Perception of emotion portrayal in cartoons by aurally and visually oriented people". Proceedings of the 11th International Conference in Music Perception and Cognition (ICMPC11, University of Washington, USA): Australia/USA: Causal Productions.

Maisonneuve, Nicolas, Matthias, Stevens, Niessen, Maria E., Hanappe, Peter & Steels, Luc (2009). "Citizen Noise Pollution Monitoring". The Proceedings of the 10th International Digital Government Research Conference.

Maisonneuve et al. (2008-11). NoiseTube. http://www.noisetube.net (accessed 14 May 2012).

McCrae, Robert R. & Costa, Paul T. (1999). "A Five-Factor Theory of Personality". Chapter 5, pp. 139-53 in Handbook of Personality. Theory and Research. 2nd edition. Pervin, Lawrence A. & John, Oliver P. (Eds). The Guilford Press 1999.

MIREX, The Music Information Retrieval Evaluation eXchange. Webpages at http://www.music-ir.org/mirex/wiki/MIREX_HOME (accessed 14 May 2012).

Morfeld, Matthias, Petersen, Corinna, Krüger-Bödeker, Anja, Mackensen, Sylvia von & Bullinger, Monika (2006). "The assessment of mood at workplace - psychometric analyses of the revised Profile of Mood States (POMS) questionnaire". Psychosoc Med. 2007; 4: Doc06. Published online in May 2007 and available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2736534/ (accessed 14 May 2012).

Nilsson, Mats (2007. Soundscape quality in urban open spaces. In Proceedings of INTER-NOISE 2007, Istanbul, Turkey.

Nygren, Paul (2009). Loudness calculation according to ITU-R BS. 1770-1, for MatLab.

Nyuk, Hien Wong & Wy, Leng Seow Jan (2003). "Total building performance evaluation of academic institution in Singapore". Building and Environment 38 (2003) 161 – 176. Elsevier Science Ltd.

Schwarz, Diemo (2004): Data-Driven Concatenative Sound Synthesis. PhD Thesis in Acoustics, Computer Science, Signal Processing Applied to Music, Université Paris 6 - Pierre et Marie Curie.

Schwarz, Diemo (2008). Principles and Applications of Interactive Corpus-Based Concatenative Synthesis JIM 2008.

Terry, P. C., Lane, A. M., & Fogarty, G. J. (2003). "Construct validity of the POMS for use with adults". Psychology of Sport and Exercise, 4 (2), 125-139.

Terry, Peter C. , Lane, Andrew M. , Lane, Helen J. and Keohane, Lee(1999) "Development and validation of a mood measure for adolescents". Journal of Sports Sciences, 17: 11, 861 — 872.

Vuoskoski, Jonna K. & Eerola, Tuomas (2011). "The role of mood and personality in the perception of emotions represented by music". Cortex 47 (2011) 1099-1106

Vuoskoski, Jonna K. & Eerola, Tuomas (2012). "Measuring music-induced emotion : A comparison of emotion models, personality biases, and intensity of experiences." 2011 15: 159 Musicae Scientiae.