PAN
ARCHIVES OF ACOUSTICS
Vol. **46**, No. 2, pp. 279–288 (2021)
DOI: 10.24425/aoa.2021.136582

Research Paper

Assessment of Audio-Visual Environmental Stimuli. Complementarity of Comfort and Discomfort Scales

Jan FELCYN[(1)*], Anna PREIS[(1)], Marcin PRASZKOWSKI[(1)], Małgorzata WRZOSEK[(2)]

[(1)] *Department of Acoustics*
Faculty of Physics
Adam Mickiewicz University
Poznań, Poland
*Corresponding Author e-mail: janaku@amu.edu.pl

[(2)] *Institute of Philosophy*
Szczecin University
Szczecin, Poland

(*received April 2, 2020; accepted February 4, 2021*)

The aim of the study was to examine how the wording of a question about audio, visual and audiovisual stimuli can affect the assessment of the environment. The participants of the psychophysical experiments were asked to rate, on a numerical scale, audio and visual information both separately and together, combined into mixes. A set of questions was used for all the investigated audio, visual, and audio-visual stimuli. The participants were asked about the comfort or the discomfort caused by the perceived stimuli presented at three different sound levels.

The results show that there are no statistically significant differences between the assessment of comfort and discomfort associated with visual samples. Actually, the comfort and discomfort ratings are equivalent to the extent that a discomfort rating can be represented as the opposite to the comfort rating, i.e. the discomfort rating is equal to the 10 minus comfort rating.

In general, the results obtained for audio and audio-visual samples were the same, with only a few exceptions that were dependent on sound level. No statistically significant differences were found for the loudest stimuli, but there were some exceptions for the softener cases. Based on the results, we show that only for visual stimuli both scales are totally interchangeable. When presenting audio and audio-visual samples, only one scale should be applied – either discomfort or comfort, depending on the context and the character of the stimuli.

Keywords: audio-visual interaction; environment assessment; discomfort; comfort; environmental perception; environmental quality.

1. Introduction

Perception of the world is multisensory. All the senses work together, at the same time. What is more, they can influence each other, thereby contributing to more sophisticated sensations and reactions. This cross-modality is widely-described in the literature, e.g. (Barutchu *et al.*, 2019; Gu *et al.*, 2019; Kattner *et al.*, 2019; Spence, Zampini, 2006; van Stokkom *et al.*, 2018).

For sound perception, the main field of interest is the interaction between sight and hearing. There is some evidence that adding an image to the sound can influence its perception – e.g. the ventriloquism effect (Wallace *et al.*, 2004) or the McGurk effect (McGurk, MacDonald, 1976). This interaction has been widely described in recent years for various audio-visual stimuli, including e.g. the perception of wind turbine noise (Schäffer *et al.*, 2019; Szychowska *et al.*, 2018). Although the most common way of presenting stimuli is based on screen and loudspeakers/headphones, the development of augmented and virtual reality has recently allowed scientists to develop new ways of studying audio-visual interactions (Asakura *et al.*, 2019).

Audio-visual interactions are particularly important for assessing the quality of the surroundings or public spaces. Not only images (landscape) but also sounds (soundscape) are key factors in our perception of a place. Nowadays, besides traditional assessment methods (like *in situ* survey research), people can use smartphone technologies when assessing the quality of the environment. For landscapes, smartphone apps can be used to assess specific aspects of public space, such as the presence of green areas (Ladle *et al.*, 2018; Vich *et al.*, 2019) or places conducive to physical activity (Hoffimann *et al.*, 2018). On the other hand, mobile apps can be used to rate urban soundscapes (Aspuru *et al.*, 2016; Herranz-Pascual *et al.*, 2016) or to improve the process of noise mapping (Guillaume *et al.*, 2016; Murphy, King, 2016; Zuo *et al.*, 2016).

All aspects of the given place can be reduced to one simple (but not easy) question: how good/bad is it? Audio, visual or audio-visual samples of an environment can be assessed by different means. For the noise assessment there is the standardized ICBEN scale (Fields *et al.*, 2001) used to rate annoyance. For more complex stimuli, other words and scales are used.

Jian Kang and his colleagues used to refer to the 'acoustic comfort' of the public space or in the context of interactions, the 'audio-visual environment comfort' (Kang, 2006; Liu, Kang, 2018). This term is also used in other recent papers (Lee *et al.*, 2020; Zhang *et al.*, 2018). On the other hand, it is not uncommon to ask people about the pleasantness of the surroundings (Filipan *et al.*, 2019; Haapakangas *et al.*, 2020; Lee, Lee, 2020). However, this term sometimes relates to the overall environment assessment (not only its audio component). In this context two scales are widely used: 5-point and 11-point.

The landscape assessment process is more complex. There are many indicators used for different aspects of landscapes, and even several hundreds of them can be used, divided into certain categories. Such a division can be found in (Cutaia, 2016); there are 8 different subgroups: visual, morphological, historical/cultural/architectural, physical geography, naturalistic-environmental, regional land use, actions to protect/improve, and socio-economic (for more explanation see (Cutaia, 2016) at page 52). However, there are some indices to globally describe landscape quality. One of them is the landscape character assessment (LCA) defined as 'a method based on the process of describing, mapping and evaluating different and distinctive characters of the landscape' (Atik *et al.*, 2017). The other index was proposed by Alampi Sottini *et al.* (2018): the Visual Quality Index based on people's answers and principal component analysis. Nevertheless, this approach is time-consuming and requires answers to many different questions about the character of a landscape.

When assessing the audio-visual stimuli in the environment we are interested to study how each component, audio or visual, contributes to the 'total' perception. To study such interaction we need to assess each component separately, but using the same scale and wording. However, the solution is not so obvious. For example, it is quite natural to ask about soundscape annoyance, but it seems strange to ask about landscape annoyance. In the latter case, pleasantness assessment seems to be more appropriate. This thinking leads us to consider the term comfort in the case of landscape assessment, and the term discomfort for soundscape assessment. Of course, these two words are not equivalent – neither are the scales: they are opposite. Thus, in this research we want to find out if two opposite scales can be used interchangeably – i.e. can we say that the discomfort rating is equal to the 10 minus comfort rating? If this is the case, one scale could be simply transformed and then the results for both scales can be easily compared.

In the next sections of this paper we describe the method and the way of formulating questions. Then we analyze the data and try to identify the similarities and differences between both scales.

2. Material and methods

2.1. Aim

Our hypothesis in this paper is that the comfort and discomfort ratings are equivalent in such a way that a comfort rating can be represented as the opposite of the discomfort rating, i.e. discomfort rating is equal to the 10 minus comfort rating. To test the hypothesis, two experiments were conducted in a laboratory with a total of three different conditions: (a) audio samples only, (b) audio-visual samples, and (c) visual samples only.

In all the experimental conditions, an analog of an 11-point numerical ICBEN scale (Fields *et al.*, 2001; Preis *et al.*, 2003) – from 0 to 10 – was used to rank the comfort and discomfort. Both comfort and discomfort assessments were collected in the same experiments and evaluated by the same people. To avoid any order effects, there was a minimum two-day pause between the sessions where comfort and discomfort data were collected. The comfort ratings have already been used in our previous publication (Preis *et al.*, 2015). The discomfort data are new data. This means, however, that the participants and the stimuli are the same as in the previous and the present studies. The comfort data published previously are used here only as reference data to compare the ratings in their current form, and in order to test the hypothesis of this study.

2.2. Participants

Two groups of participants took part in two experiments. Experiment I involved 17 people, including 8 men and 9 women aged 22 to 30 years. 18 participants took part in Experiment II, although reliability analysis of their results led to the exclusion of 4 of them (more details about it can be found in the 'Results' section below). Finally, data from 14 participants, comprising 9 males and 5 females, aged 22 to 26 years, was analyzed in Experiment II. All the participants had normal hearing (the inclusion criterion was 15 dB maximum allowable hearing loss at hearing threshold).

2.3. Stimuli

Seven different locations in Poznań (Fig. 1) were selected for recording with both audio and visual information. Places representing different types of public space were chosen: more natural (park), more crowded (with a lot of people) and transport type (streets). All of them were chosen carefully, to minimalize the risk that respondents would be familiar with them – we wanted to eliminate the possible influence of knowledge on the results.

Fig. 1. Snapshots from seven different places in the Poznań area where audio and video samples were collected.

A short description of each stimulus can be found below:

- Market – a small market in the center of a district, many people passing by, small talk, background music from a small radio, sometimes low noise from a car in the street nearby.
- Pedestrian Zone – in the center of Poznań, closed for road traffic, with many shops, pubs, restaurants etc.; sound of heels clicking, people talking, children playing, opening doors/gates.
- Busy Street – a busy four-lane street with a tramline in the middle of it; sound of cars passing-by (including light and heavy ones).
- Park near busy street – a park located near a very busy street; almost no natural sounds (single birds tweets), heavy background road traffic and a circular saw used by someone in the park.
- Park – a park located several hundred meters from a street and close to Lake Malta; many birds tweeting, distant stationary road traffic.
- Roundabout – a roundabout between two streets with medium traffic flow, road traffic with light and heavy vehicles, but distinguishable to separate pass-bys.
- Side Street – a small street near a residential area consisting of detached houses, single pass-bys of passenger cars.

The audio-visual samples, each of 10-minute duration, were recorded with a Sony HDR-XR200 camera and a TEDS 4101 binaural microphone, together with a Bruel & Kjaer PULSE v.12.6.0.255 system. From these recordings, after careful analysis, 10-second audio-visual samples were created. To investigate the possible effect of sound level on assessing ambient comfort and discomfort, additional soundscape samples were created with the sound level value lower (−6 dB) and higher (+6 dB) than the original recorded sound level. All stimuli are also presented as changes of sound level in time in Fig. 2; Table 1 presents the equivalent sound levels of samples.

The process of creating the audio-visual stimuli used Adobe Premiere Pro CS5 software. As aforementioned, not only the original sound level, but also levels 6 dB higher and lower were used for the presentation of audio only or audio-visual stimuli. Thus, 21 audio samples (7 places × 3 sound levels) and a combination of 21 (7 places × 3 sound levels) original audio-visual samples were used in Experiment I. Each stimulus was presented three times. In total, the participants in Experiment I evaluated 126 different stimuli (42 audio and audio-visual samples, each presented three times). In Experiment II, 7 visual stimuli were presented.

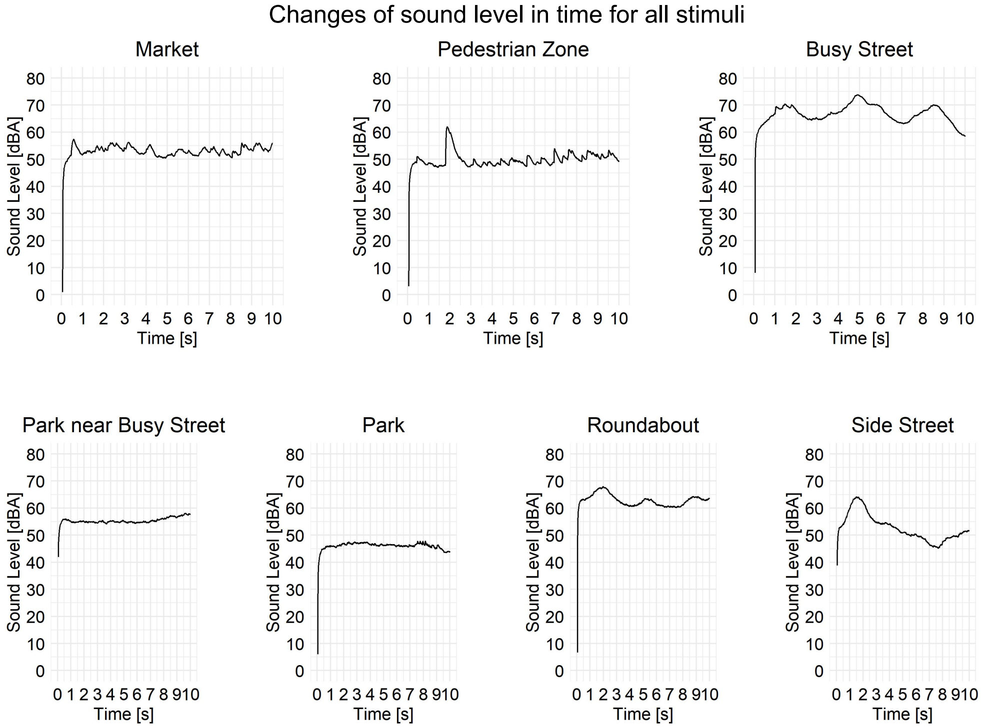

Fig. 2. Relation between sound level values and time for audio parts of all stimuli (all charts made for the 0 dB condition, i.e. for the original sound levels).

Table 1. Sound levels of stimuli presented in audio and audio-visual conditions.

Stimulus	Sound level [dBA]		
	−6 dB	0 dB	+6 dB
Market	47	53	59
Pedestrian Zone	47	53	59
Busy Street	62	68	74
Park near busy street	50	56	62
Park	40	46	52
Roundabout	57	63	69
Side Street	49	55	61

2.4. Procedure

Experiment I used two experimental conditions: audio and audio-visual samples. Participants were asked to rate the degree of comfort and discomfort on a numerical scale from 0 to 10 while imagining they were in such an environment. The participants were sitting in front of a computer screen (PC computers with 17′ LCD screens), watching and listening (Beyerdynamic DT-150 headphones) to the audio and audio-visual samples. For each of the conditions both comfort and discomfort were rated by each participant. The same approach was used in Experiment II, in which only video samples were presented. Again, participants rated both the comfort and discomfort caused by presented stimuli.

The instruction given to the participants in Experiment I was as follows:

'You will be presented with audio and audiovisual samples from various places in Poznań. Imagine that you are in that place and would like to relax. Assess your feeling of comfort/discomfort using the scale from 0 to 10, where 0 means total lack of comfort/discomfort, 10 means total comfort/discomfort. Please focus on the positive/negative feelings associated with the environment'.

In Experiment II, the same instruction was given to the participants, but a small change was applied at the beginning. The words 'Audio and audio-visual samples (...)' were replaced by 'Visual samples (...)'. In Experiment II, the comfort/discomfort assessment of the investigated places was based on visual samples. In both experiments each sample was presented three times to each participant.

In Experiment I, all stimuli (audio and audio-visual, at all three sound levels) were presented in a random order. There were two experimental sessions, first one with the comfort ratings and the second with discomfort ratings. In Experiment II, only visual stimuli were presented to the participants in a random order. The same approach was applied in this experiment i.e. there were two experimental sessions, in the first one comfort ratings and in the second one discomfort ratings were collected.

Please note that according to Fields *et al.* (2001) 'The numeric scale is felt to provide greater assurance that the scale points are equally spaced and thus meet the assumptions for linear regression and similar powerful analysis techniques that can represent the continuous range of responses to noise'.This approach to treat 11-point scale as continuous and quantitative and to compute mean values is widely used in literature (Ioannidou *et al.*, 2016; Klein, 2015; Sung *et al.*, 2017; Haapakangas *et al.*, 2020). Thus, it was also used in this paper.

3. Results

3.1. The analysis of individuals who took part in Experiment I and II

To find out if there are any inconsistencies among participants' answers we conducted reliability analyses for both experiments – Cronbach's alpha values were calculated for each group as well as for each participant in the 'when an item dropped' approach. We used the function 'alpha' from the R package 'psych'.

All participants in Experiment I correlated well with each other, thus no one was excluded from further analyses. On the other hand, 4 participants were excluded from the analyses of the experiment II based on alpha values and the fact that they negatively correlated with the other subjects. To better illustrate the process, below we provide results of reliability analysis before and after the exclusion (Table 2). Please note that 'R std' means the correlation between the subject and the total score when scales were standardized. 'R cor.' stands for the R value corrected for item overlap.

Table 2. Values of standardized and corrected R coefficients for each subject before and after exclusion of four participants in Experiment II.

All participants			4 participants excluded		
Subject	R std	R cor0.	Subject	R std	R cor0.
1	0.62	0.66	1	0.76	0.79
2	0.00	−0.06			
3	0.48	0.43	3	0.53	0.47
4	0.64	0.65	4	0.57	0.57
5	−0.08	−0.14			
6	0.72	0.76	6	0.81	0.84
7	0.06	0.02	7	0.14	0.04
8	0.72	0.76	8	0.83	0.87
9	−0.14	−0.19			
10	0.22	0.14	10	0.16	0.07
11	0.75	0.78	11	0.73	0.74
12	0.62	0.62	12	0.61	0.59
13	−0.28	−0.33			
14	0.33	0.25	14	0.44	0.37
15	0.22	0.12	15	0.13	0.02
16	0.31	0.24	16	0.27	0.20
17	0.67	0.71	17	0.82	0.85
18	0.57	0.57	18	0.53	0.52

3.2. The results for soundscapes only (Experiment I) and landscapes only (Experiment II)

We restate that the hypothesis of this paper is that the discomfort and comfort ratings are equivalent in such a way that a discomfort rating can be represented as the opposite of the comfort rating, i.e. discomfort rating is equal to the 10 minus comfort rating. To test this statement, the comfort data from experimental conditions (a) and (c) were recalculated as the '10 – comfort rating'. In next sections of this article we refer to the original discomfort ratings as 'original discomfort', and the values computed from the formula the '10 – comfort rating' is to be referred to as 'calculated discomfort'.

As it has already been mentioned, all stimuli were rated three times by each participant. In this way, the data was aggregated before statistical analyses in such a way that the median value was calculated from each of the three values and then used as a listener's response. One can ask why we used medians instead of means. Three presentations of a given stimulus are not many, but some fluctuations in people's answers were revealed. However, the differences between medians and means in the majority of cases were not high – not higher than 1. In Fig. 3 we present the differences computed for both scales. As they are not high we preferred to use medians at this stage. However, next analyses are made using mean ratings for both scales.

Moreover, for each stimulus and participant (as each stimulus was presented three times) we computed differences between minimal and maximal value for each scale separately. The analysis revealed that for comfort scale 72% of differences were not higher than 2; for discomfort it was 73%. Results are presented in Fig. 4. The similarity between both histograms also suggests that people's answers are distributed the same, no matter which scale was used.

The results were calculated in R environment using bootstrapped confidence intervals (functions from the package 'boot'). The differences between the means were calculated using a 'WRS2' package and a bootstrapped version of t-tests. In Figs 5a and 5b the averaged results for both scales, the original and calculated discomfort for the audio and visual stimuli (separately), are presented. In Fig. 5a there are three different colours of the points and error bars. Each colour represents a different relative sound level (−6 dB, 0 dB – original sound level and +6 dB) at which the stimuli were presented to the listeners. The asterisk mark means that there is a statistically significant difference

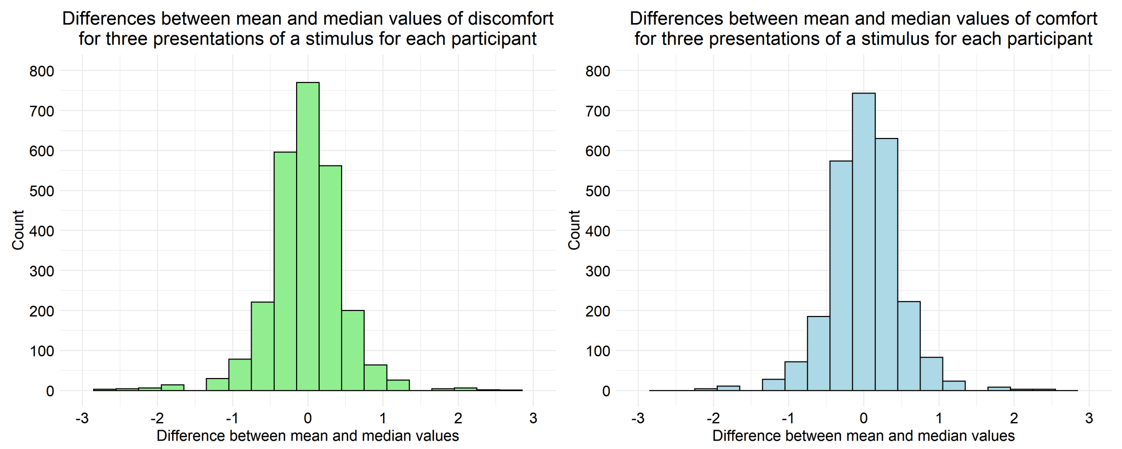

Fig. 3. Histograms of the differences between means and medians for individuals' ratings of discomfort and comfort.

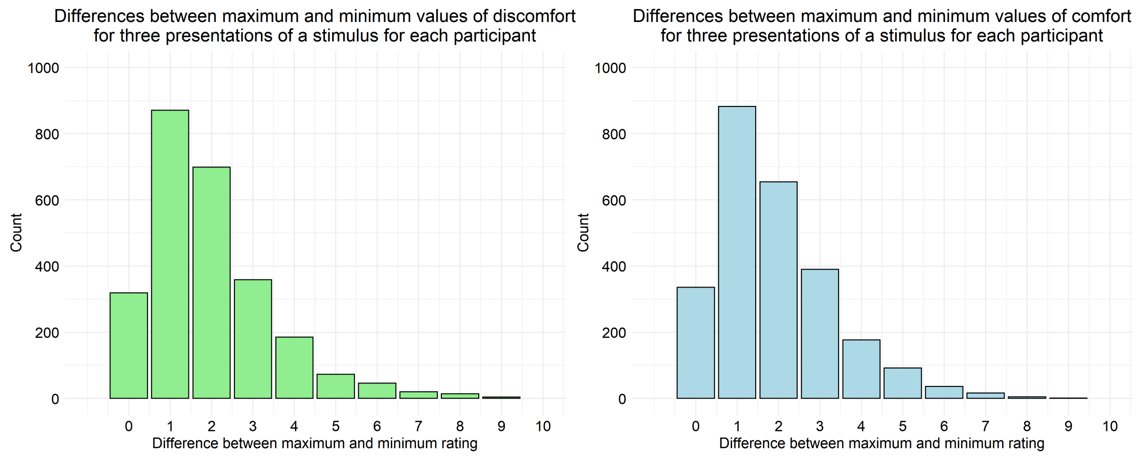

Fig. 4. Histograms of the differences between maximal and minimal values for individuals' ratings of discomfort and comfort.

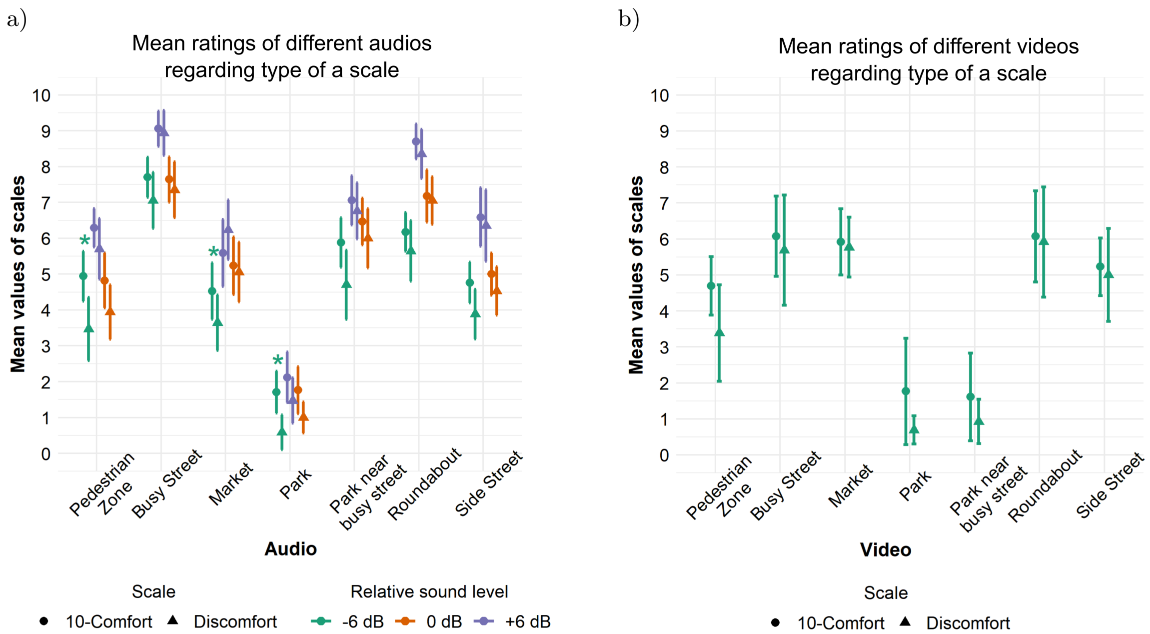

Fig. 5. Averaged assessment of the discomfort scale and 10 minus comfort scale for 7 soundscapes (a) and 7 landscapes (b). Different colours signify the sound level of the presented audio. Statistically significant differences between both scales are marked with asterisks. Error bars represent 95% confidence intervals.

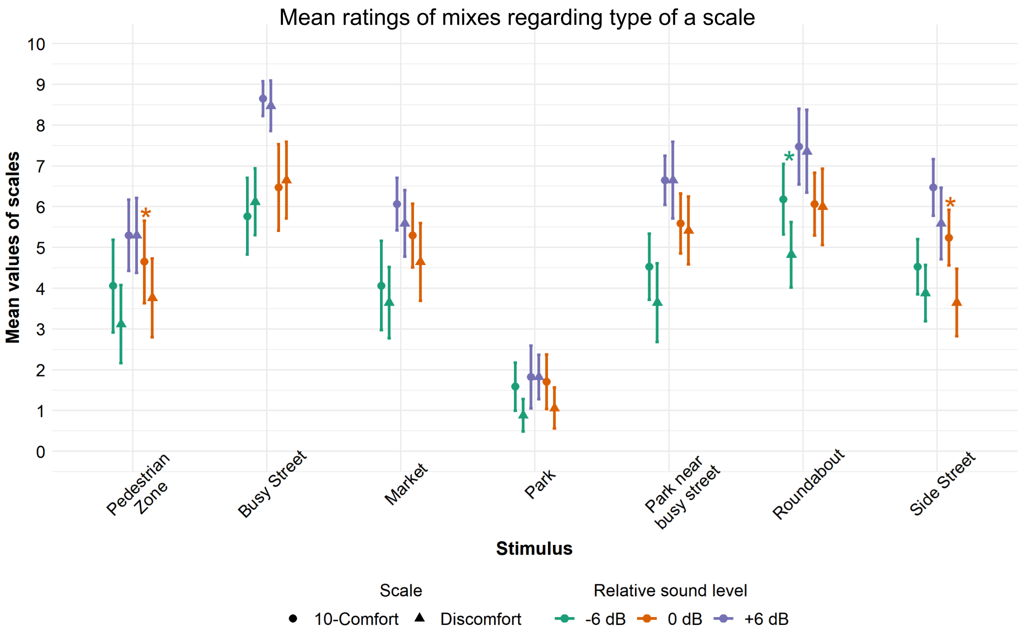

Fig. 6. Averaged assessment of the discomfort and 10 minus comfort scales for the investigated mixes in 7 places in the city of Poznań. Different colours signify the sound level of the presented audio part of the mix, statistically significant differences between both scales are marked with asterisks. Error bars represent 95% confidence intervals.

between the two analyzed scales – regarding their mean values.

It can be seen that for all the audio samples presented at both 0 and +6 dB sound levels, there are no statistical differences between the two scales. This also applies to all the investigated visual stimuli presented in Fig. 5b. On the other hand, there are three cases for −6 dB relative sound level when the mean values of original and computed discomfort ratings are different: this is observed for 'Pedestrian Zone' ($t = 6.325$, $p = 0.0275$, effect size, $es = 0.4987$), 'Market' ($t = 4.4415$, $p = 0.0469$, $es = 0.4777$) and 'Park' ($t = 10.2334$, $p = 0.0038$, $es = 0.7152$).

For video stimuli, both scales are the same in all cases – however some not statistically significant differences can be observed in the range of confidence intervals, especially in case of both 'Park' stimuli.

3.3. The results of the mixes (Experiment I)

Based on the results obtained for the cases of a single modality, audio only and video only, our expectation was that a similar trend should be observed in the originally recorded audio-visual stimuli. The comparison of the mean values of the two discomfort (original and computed) scales is presented in Fig. 6.

As shown in Fig. 6, only if the original stimuli were presented at a sound level of +6 dB there are no significant differences for any of these scales. In contrast, there are statistically significant differences between two scales for two conditions: −6 dB and 0 dB. For the −6 dB condition, a statistically significant difference was found for 'Roundabout' ($t = 7.1806$, $p = 0.0157$, $es = 0.5697$). For the 0 dB condition, two differences were observed for 'Pedestrian Zone' ($t = 4.1856$, $p = 0.0496$, $es = 0.4337$) and 'Side Street' ($t = 12.5182$, $p = 0.002$, $es = 0.6399$). When each mix is analyzed separately, for 'Busy Street', 'Market', 'Park', and 'Park near busy street' there are no significant differences between either discomfort scale for all different relative sound levels.

4. Discussion

With regard to analyses of means (Figs 5 and 6), several interesting cases can be observed. In general, in almost all cases, regardless of whether they were single or bimodal, discomfort was assessed as less (original discomfort) when compared with the results of the subtracted comfort scores from 10 (calculated discomfort) – but this difference is very small and does not show any statistical significance. A possible interpretation is that when we ask about discomfort, it is usually rated higher than comfort. In other words, it may be easier to give higher ratings for discomfort than the same numbers for comfort – i.e. people's requirements to rate something as comfortable are higher than those used to rate discomfort. For example when something is theoretically in the middle of both scales (rated as 5 on both scales), indeed discomfort would be rated as 5 but comfort would be rather between 4 and 5 – giving a higher rate for the calculated discomfort scale. However, it has to be remembered that these differences are small, so we can only speak about some tendencies, not statistically significant differences.

If video stimuli alone were presented to participants, there were no statistically significant differences in either case. However, it should be stated that in the case of both 'natural' stimuli, i.e. 'Park' and 'Park near busy street', the values of the original discomfort ratings had smaller confidence intervals than those on the calculated discomfort scale. These results suggest that by asking people about the discomfort felt in natural areas, they agree that there is almost no discomfort in it. But when we ask them how to rate comfort, answers are more complex and the variety is also wider. Nevertheless, it could be treated merely as a tendency, without statistical significance.

For audio stimuli, it should be noticed that differences were found for the −6 dB condition – but there were only three such cases. When the sound level increased, all differences disappeared – for both the original (0 dB) and +6 dB conditions. A possible interpretation is that when the sound level is lower than in a real environment, people may be confused about their perception of a given soundscape in terms of comfort and discomfort. As the sound level increases, their ratings become more consistent, no matter which scale is used.

With regards to the results of the analyses of mixes, there are three cases in which the mean values of both scales are statistically different – one case for −6 dB and two cases for the 0 dB condition. So this time, there is no rule that differences disappear with the sound level increase. However, only three such cases cannot be treated as a stable tendency and it is hard to interpret that. Maybe for those cases there is some confusion among listeners: on the basis of visual stimuli, they expect sound with other sound values than those that were actually used. But definitely more data should be gathered to test this explanation reliably.

The analysis of means provides a global point of view, without taking into account individual differences between subjects. However, in psychophysical experiments these differences are always present and cannot be eliminated. One way of catching it is through the listeners' reliability analysis made before the main analysis. As we have mentioned already, based on Cronbach's alpha values and correlation between subjects, we excluded some of them from the dataset – because they negatively correlated with the others.

Summing up, we can say that both scales are totally interchangeable only for visual stimuli – no matter which one is used, it can be transformed into the other using the formula '$10 - x$, where 'x' is the rating on a given scale. On the other hand, some statistically significant differences were found for both audio and audio-visual conditions. In this case, it is better to assume that both scales are not the same, and only one should be used. We propose to use the comfort scale for more natural and peaceful environmental samples. The discomfort scale should be better for more dense stimuli like urban ones, or those with busy traffic.

5. Conclusions

As both scales were the same for visual stimuli (based on mean values) it is a matter of choice which scale – discomfort or comfort – we apply in the survey because one scale can be easily transformed into the other one. We propose to ask about comfort in such cases.

As some differences between scales were statistically significant for audio and audio-visual stimuli, it seems that both scales are interchangeable only when the sound level is higher than in the original situation (no differences were found for +6 dB cases).

Based on these remarks, we can say that it is generally possible to rate only discomfort for audio, audio-visual and visual stimuli – but in the latter case, discomfort could be also obtained as a simple transformation of the comfort scale, i.e. the 10 minus comfort rating.

Acknowledgments

This research was supported by the grant given by National Science Center, no. 2017/25/N/HS6/01164.The authors also declare no conflict of interest.

References

1. Alampi Sottini V. *et al.* (2018), Urban landscape assessment: a perceptual approach combining virtual reality and crowdsourced photo geodata, *Aestimum*, (73): 147–171, doi: 10.13128/aestimum-24927.

2. Asakura T., Tsujimura S., Yonemura M., Hyojin L., Sakamoto S. (2019), Effect of immersive visual stimuli on the subjective evaluation of the loudness and annoyance of sound environments in urban cities, *Applied Acoustics*, **143**: 141–150, doi: 10.1016/J.APACOUST.2018.08.024.

3. Aspuru I., García I., Herranz K., Santander A. (2016), CITI-SENSE: methods and tools for empowering citizens to observe acoustic comfort in outdoor public spaces, *Noise Mapping*, **3**(1): 37–48, doi: 10.1515/noise-2016-0003.

4. Atik M., IşikliR.C., Ortaçeşme V., Yildirim E. (2017), Exploring a combination of objective and subjective assessment in landscape classification: Side case from Turkey, *Applied Geography*, **83**: 130–140, doi: 10.1016/J.APGEOG.2017.04.004.

5. Barutchu A. *et al.* (2019), Multisensory perception and attention in school-age children, *Journal of Experimental Child Psychology*, **180**: 141–155, doi: 10.1016/j.jecp.2018.11.021.

6. Cutaia F. (2016), *Strategic Environmental Assessment: Integrating Landscape and Urban Planning*,

UNIPA Springer Series, Springer International Publishing, Cham, doi: 10.1007/978-3-319-42132-2.

7. Fields J.M. *et al.* (2001), Standardized general-purpose noise reaction questions for community noise surveys: research and a recommendation, *Journal of Sound and Vibration*, **242**(4): 641–679, doi: 10.1006/jsvi.2000.3384.

8. Filipan K., De Coensel B., Aumond P., Can A., Lavandier C., Botteldooren D. (2019), Auditory sensory saliency as a better predictor of change than sound amplitude in pleasantness assessment of reproduced urban soundscapes, *Building and Environment*, **148**: 730–741, doi: 10.1016/j.buildenv.2018.10.054.

9. Gu J., Liu B., Li X., Wang P., Wang B. (2019), Cross-modal representations in early visual and auditory cortices revealed by multi-voxel pattern analysis, *Brain Imaging and Behavior*, **14**(5): 1908–1920, doi: 10.1007/s11682-019-00135-2.

10. Guillaume G. *et al.* (2016), Noise mapping based on participative measurements, *Noise Mapping*, **3**(1): 140–156, doi: 10.1515/noise-2016-0011.

11. Haapakangas A., Hongisto V., Oliva D. (2020), Audio-visual interaction in perception of industrial plants – Effects of sound level and the degree of visual masking by vegetation, *Applied Acoustics*, **160**: 107121, doi: 10.1016/j.apacoust.2019.107121.

12. Herranz-Pascual K., García I., Aspuru I., Díez I., Santander Á. (2016), Progress in the understanding of soundscape: objective variables and objectifiable criteria that predict acoustic comfort in urban places, *Noise Mapping*, **3**(1): 247–263, doi: 10.1515/noise-2016-0017.

13. Hoffimann E., Campelo D., Hooper P., Barros H., Ribeiro A.I. (2018), Development of a smartphone app to evaluate the quality of public open space for physical activity. An instrument for health researchers and urban planners, *Landscape and Urban Planning*, **177**: 191–195, doi: 10.1016/J.LANDURBPLAN.2018.05.005

14. Ioannidou C., Santurette S., Jeong C.-H. (2016), Effect of modulation depth, frequency, and intermittence on wind turbine noise annoyance, *The Journal of the Acoustical Society of America*, **139**(3), 1241–1251, doi: 10.1121/1.4944570.

15. Kang J. (2006), *Urban Sound Environment*, Taylor & Francis.

16. Kattner F., Samaan L., Schubert T. (2019), Cross-modal transfer after auditory task-switching training, *Memory & Cognition*, **47**(5): 1044–1061, doi: 10.3758/s13421-019-00911-x.

17. Klein A. (2015), *Annoyance indicators for various urban road vehicle pass-by noises and urban road traffic noise combined with tramway noise*, Ecole Nationale des Travaux Publics de l'Etat.

18. Ladle A., Galpern P., Doyle-Baker P. (2018), Measuring the use of green space with urban resource selection functions: An application using smartphone GPS locations, *Landscape and Urban Planning*, **179**: 107–115, doi: 10.1016/J.LANDURBPLAN.2018.07.012.

19. Lee H.M., Lee H.P. (2020), Noise masking in high population country using sound of water fountain, *Applied Acoustics*, **162**: 107206, doi: 10.1016/j.apacoust.2020.107206.

20. Lee H.M., Liu Y., Lee H.P. (2020), Assessment of acoustical environment condition at urban landscape, *Applied Acoustics*, **160**: 107126, doi: 10.1016/j.apacoust.2019.107126.

21. Liu F., Kang J. (2018), Relationship between street scale and subjective assessment of audio-visual environment comfort based on 3D virtual reality and dual-channel acoustic tests, *Building and Environment*, **129**: 35–45, doi: 10.1016/j.buildenv.2017.11.040.

22. McGurk H., MacDonald J. (1976), Hearing lips and seeing voices, *Nature*, **264**(5588): 746–748, doi: 10.1038/264746a0.

23. Murphy E., King E.A. (2016), Smartphone-based noise mapping: Integrating sound level meter app data into the strategic noise mapping process, *Science of The Total Environment*, **562**: 852–859, doi: 10.1016/j.scitotenv.2016.04.076.

24. Preis A., Kaczmarek T., Wojciechowska H., Żera J., Fields J.M. (2003), Polish version of standardized noise reaction questions for community noise surveys, *International Journal of Occupational Medicine and Environmental Health*, **16**(2): 155–159.

25. Preis A., Kociński J., Hafke-Dys H., Wrzosek M. (2015), Audio-visual interactions in environment assessment, *Science of The Total Environment*, **523**: 191–200, doi: 10.1016/j.scitotenv.2015.03.128.

26. Raimbault M. (2006), Qualitative judgements of urban soundscapes: Questionning Questionnaires and semantic scales, *Acta Acustica united with Acustica*, **92**(6): 929–937.

27. Schäffer B., Pieren R., Wissen Hayek U., Biver N., Gręt-Regamey A. (2019), Influence of visibility of wind farms on noise annoyance – A laboratory experiment with audio-visual simulations, *Landscape and Urban Planning*, **186**: 67–78, doi: 10.1016/j.landurbplan.2019.01.014.

28. Spence C., Zampini M. (2006), Auditory contributions to multisensory product perception, *Acta Acustica united with Acustica*, **92**(6): 1009–1025.

29. van Stokkom V.L., Blok A.E., van Kooten O., de Graaf C., Stieger M. (2018), The role of smell, taste, flavour and texture cues in the identification of vegetables, *Appetite*, **121**: 69–76, doi: 10.1016/j.appet.2017.10.039

30. Sung J.H. *et al.* (2017), Influence of transportation noise and noise sensitivity on annoyance: a cross-sectional study in South Korea, *International Journal*

of Environmental Research and Public Health, **14**(3): 322, doi: 10.3390/ijerph14030322.

31. Szychowska M., Hafke-Dys H., Preis A., Kociński J., Kleka P. (2018), The influence of audio-visual interactions on the annoyance ratings for wind turbines, **129**: 190–203, doi: 10.1016/j.apacoust.2017.08.003.

32. Vich G., Marquet O., Miralles-Guasch C. (2019), Green exposure of walking routes and residential areas using smartphone tracking data and GIS in a Mediterranean city, *Urban Forestry & Urban Greening*, **40**: 275–285, doi: 10.1016/J.UFUG.2018.08.008.

33. Viollon S., Lavandier C., Drake C. (2002), Influence of visual setting on sound ratings in an urban environment, *Applied Acoustics*, **63**(5): 493–511, doi: 10.1016/S0003-682X(01)00053-6.

34. Wallace M.T., Roberson G.E., Hairston W.D., Stein B.E., Vaughan J.W., Schirillo J.A. (2004), Unifying multisensory signals across time and space, *Experimental Brain Research*, **158**(2): 252–258, doi: 10.1007/s00221-004-1899-9.

35. Zhang X., Ba M., Kang J., Meng Q. (2018), Effect of soundscape dimensions on acoustic comfort in urban open public spaces, *Applied Acoustics*, **133**: 73–81, doi: 10.1016/j.apacoust.2017.11.024.

36. Zuo J., Xia H., Liu S., Qiao Y. (2016), Mapping urban environmental noise using smartphones, *Sensors (Basel)*, **16**(10): 1692, doi: 10.3390/s16101692.