EFFECT OF AUDIO-VISUAL INTERACTION ON SUBJECTIVE ASSESSMENT OF INDUSTRIAL SOUNDSCAPES

J Smyrnowa Environment Agency, UK
G Brown Environment Agency, UK

1 INTRODUCTION

Audio-visual interaction plays a crucial role in the overall environmental and soundscape assessment[1,2]. The presence of natural sounds such as birdsong or flowing water in audio-visual scenes have been linked with higher soundscape appraisal[3], while scenes containing views of abundant water and vegetation were rated higher compared to the audio only presentations of the same scenes[2]. However, there is contradictory evidence on the effect of the visibility of a sound source on noise annoyance. Studies have shown both a decreased[4] (road traffic) and increased (roads, railways and wind turbines)[5,6] annoyance when the noise source was visible compared to the audio only scenes when it was not.

Few studies have considered industrial sound sources as the subject of audio-visual studies of noise annoyance and environmental/soundscape assessment. In one study, partial masking of an industrial plant by vegetation was demonstrated to affect rating of the pleasantness of a scenery[7]; however, no significant effect was observed in terms of noise annoyance[7]. Another laboratory study showed that the visibility of a chiller had no effect on perceived loudness and noise annoyance of an audio-visual scene[8]. A forest soundscape containing both sound and sight of a substation and surrounding pylons was found to be significantly less pleasant, comfortable, natural and calming compared to the soundscape in a part of the same forest with no industrial sound sources present[9].

This paper investigates the effects and interactions of sound source visibility and audibility on industrial source identification and appraisal in virtual settings. In an online experiment, subjects were presented with visual only and audio-visual scenarios recorded in-situ. Some of the visual scenes contained industrial sound sources - partially hidden by vegetation or fully exposed, while almost all audio files had industrial sources present alongside other environmental sounds. The industrial sound sources varied in terms of the overall level and the presence of tonal or impulsive components. In some of the audio-visual scenes, the audio information was deliberately swapped with other scenes while preserving the original video information.

The study also investigated the effect of visual and auditory information on the assessment of comfort, willingness to stay and overall quality of the presented environment, and acceptability of the industrial source to the scene.

2 SITES AND LOCATIONS

Four sites of different audio-visual context were selected for the study. These are referred to as S1, S2, S3 and S4. Three of the sites (S1 – S3) were near industrial plants, while S4 was in a residential area. At S1, recordings were made at five locations around Abbey Forged Products, Sheffield. Locations S1L1, S1L2 and S1L3 were at decreasing distances to the north-east of the plant, while locations S1L4 and S1L5 were at northern and south-western sides, respectively. Location S2 was to the south-east of the Abbey Forged site boundary on a pavement in front of residential houses. The residential area and the plant were separated by a river, a 45-metre-wide tree-belt and a road.

Two recordings were made at this location, but at different periods of time, named as S2L1 and S2L2. Location S3 was situated on a footpath at the rear of William Cook Cast Production Plant, Sheffield, approximately 20 metres from the plant building. Location S4 was in the residential area of Catcliffe, Sheffield, approximately 100 metres from the M1 motorway. Audio-visual features of each of the sites and measurement locations, including dominant sound sources, are described in Table 1, while the visual representations are shown in Figure 1.

Table 1: Audio-visual features of the sites and recording locations.

Site/ Location	Visual description	Sound sources
S1L1	180m from Abbey Forged Products, among the woods; no industrial site is visible	Birdsong, aircraft noise, a clearly audible tonal sound from the plant
S1L2	80m from Abbey Forged plant, among the woods, industrial plant (Abbey Forged) is clearly visible in a distance but not dominant	A low freq. tonal sound (with varying loudness) from the plant, occasional bangs, distant aircraft, birdsong
S1L3	Next to Abbey Forged Products; the plant is the dominant visual source	A low freq. sound from Abbey Forge is dominant, many bangs and hisses
S1L4	50m to the north of Abbey Forged; among bushes and trees, the site building is visible	Birdsong, aircraft noise, a clearly audible tonal sound from the plant
S1L5	On a bridge to the south-west of Abbey Forged; the river, bridge and the site building (it is not obvious what kind of building it is) are the dominant visual sources	Sound of the river flow, a few bangs from the plant, some traffic noise
S2L1	Near a road with trees, the plant building (Abbey Forged) is largely hidden by the trees but still is visible, passing cars and a pedestrian with a dog	A pedestrian walking, passing cars, a slightly perceptible tonal sound from Abbey Forged
S2L2	Near a road with trees, the plant building (Abbey Forged) is largely hidden by the trees but still is visible	Birdsong, a very distant low freq. noise (traffic, aircraft and plant rumbling noise) and a clearly (but not very loud) tonal sound from Abbey Forged plant
S3	Rear of William Cook industrial plant dominates; a big stack, fence, bushes	Very loud broadband sounds from the plant, intermittent hisses (pressure releases), occasional bangs
S4	Houses and a side road; the motorway is not visible	General hum from the motorway, occasional passing cars on the side road

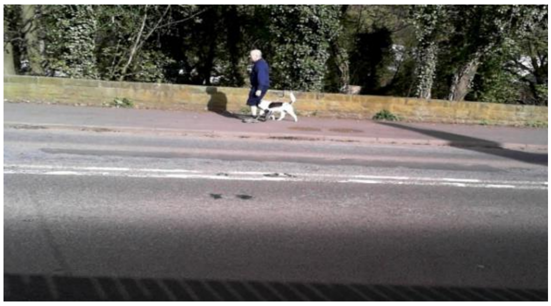

S1L1


S1L2


S1L3

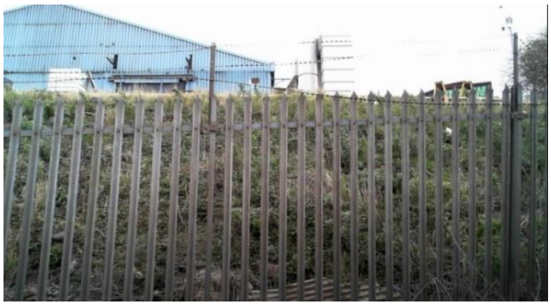

S1L4

S1L5



S2L1

S2L2



S3

S4

Figure 1: Visual representations of the scenes used in the online experiment.

At each location, binaural recordings (24-bit rate and 48 kHz resolution) were made with the SQobold acquisition system and a binaural headset, both of HEAD acoustics[10,11]. The recordings were supplemented with a GPS track recordings and simultaneous video recordings via camera attached to the SQobold recorder.

Additionally, measurements of objective acoustic parameters and monaural audio recordings (24-bit rate and 48 kHz resolution) were carried out with a B&K 2250 SLM. The headset was worn by one of the authors of this paper and thus was at approximately 1.6 metres above the ground. The SLM was positioned at 1.2 metres above the ground. The recordings were made in April 2020. Each of the recordings lasted 1-2 minutes.

3 ONLINE EXPERIMENTS

3.1 Experiment design

An online experiment was set up in PsyToolkit - a free-to-use toolkit for demonstrating, programming, and running cognitive-psychological experiments and surveys[12,13]. All recordings were post-processed and saved on a Google Drive to be used during the experiment. ArtemiS SUITE software of HEAD acoustics[11] and a free version of Reaper software[14] were used for the post-processing of the recordings. The results of the survey were saved on the server for further downloading. Nobody except of the authors of this paper had access to the survey results.

In order to achieve equal levels of sound at the user`s end, a 1kHz tone was generated in Audacity software[15]. Prior to the experiment, the participants were asked, first, to set the volume on their computer to zero; then, while listening to the tone, increase the volume gradually until the tone becomes just audible. They were also warned to not adjust the volume any further before the end of the experiment. This ensured that all sounds played during the survey were at a level similar to the real environments. It also ensured that the sound levels would not be harmful to hearing.

3.2 Audio-visual files and survey

During the experiment participants were asked to watch audio-visual files and answer questions after each presentation. All files used in the experiment were of 1-minute duration. In order to investigate the effect of the visual scene on what people expected to hear from the scene, the video only files (with no audio) were played first and were then followed by the audio-visual scene (file) of the same location. In some audio-visual presentations the audio part was swapped for the audio from a different scene. Overall, nine video only files followed by audio-visual files (with the same video part) were presented to the participants. The sequence of the files used during the experiment is shown in Table 2. The table also shows the overall sound pressure levels and special acoustic features of the sound files used in each experiment. Prominence of tonality, impulsivity and associated penalties in the audio files were identified using BZ5503 software of B&K. Intermittency features were rated subjectively based on the guidance in BS 4142 UK standard[16].

Table 2: Scene sequence with associated audio and video files used during the experiments; SPLs and special acoustic features of the audio files.

Scene number	Video file	Audio file	L_{Aeq}, dB	Prominent Tonality		Impulsivity		Intermittency
				Hz	Penalty, dB	Level, dB	Penalty, dB	Penalty, dB
1	S1L1	S1L1	51.5	790.6	3.8	-	-	-
2	S1L5	S1L5	59.1	150	0.2	-	-	-
3	S2L1	S2L1	67.8	-	-	-	-	-
4	S1L4	S1L1	51.5	790.6	3.8	-	-	-
5	S3	S3	66.2	92.18	0	69.0	3.4	3*
6	S1L2	S1L2	59.9	790.6	6	-	-	-
7	S2L2	S2L2	44.1	1193	3.2	-	-	-
8	S4	S3	63.0	N/A	N/A	N/A	N/A	3*
9	S1L3	S1L2	59.9	790.6	6	-	-	-

* penalty identified subjectively

The survey was anonymous, although at the start of the experiment participants had to sign a consent form. The questionnaire consisted of four parts. The first part was pre-survey questions on the participant`s age, gender, current residence type (urban, suburban or rural), any hearing impairment, general sensitivity to noise, and any experience or knowledge in noise assessment or acoustics. Questions in part two followed the video only files and aimed to investigate the auditory expectation participants might have had from a visual scene. Questions from parts three and four were asked after watching the same video but with added audio track. Part three was focused on the identification of sound sources through auditory and visual sensations. Part four was to investigate the overall perceptive qualities of the audio-visual scene This part of the survey was adapted from our earlier pilot soundwalks study[17]. Table 3 shows questions from parts two, three and four of the survey. The questions in part two had a binominal scale, where 0 represented “No” and 1 – “Yes”. The questions in parts three and four used a five point scale, with 0 representing absence of a sound source (part three) or the lowest perceived quality (part four) and 4 was to indicate that the source was either the most dominant (part tree) or of the highest perceptive quality (part four). The answer choices had both numerical and descriptive ratings scales as recommended by soundscape standard ISO/TS 12913 Part 3[18].

Table 3: Questions from parts two, three and four of the survey.

Part two: Q1: What sound sources do you expect to hear at this location?	Traffic noise (e.g. cars, buses, trains, airplanes)	Sounds of human activity (e.g. footsteps, conversation)	Natural Sounds (e.g. singing birds, flowing water, wind in vegetation	Industrial sound (sources)
Part three: Q2: To what extent did you hear the following sounds? Q3: To what extent did you see the following sound sources?				
Part four	Q4: To what extent would you like to stay in the presented environment?	Q5: To what extent did you feel comfortable in the presented environment?	Q6: To what extent do you think the industrial sound is acceptable in the presented environment?	Q7: Overall how would you describe the presented sound environment?

3.3 Participants

All participants were unpaid volunteers, mostly from the Environment Agency but also included some external participants. 50 participants took part, however only 34 responded to all questions. From those, 31 reported no known hearing impairment, while two suffered from a mild tinnitus. Responses to the other questions from the first part of the survey are shown in Table 4.

Table 4: Descriptive statistics from the pre-survey questions where numbers in bold represent number of the responses.

Age	Gender	Residency	Acoustic knowledge	Sensitivity to noise
18-25: **1** 26-35: **10** 36-45: **8** 46-55: **10** 55+: **5**	Male: **18** Female: **16**	Urban: **18** Suburban: **16** Rural: **0**	IOA Certificate in Env. Noise Measurements: **7** EA Noise Regulator: **4** IOA Diploma: **1** Expert Witness dealing with noise: **1** Subjective Jury Testing: 1	A lot: **8** Normal: **25** Little: **1**

4 RESULTS

4.1. Aural expectations from the video scenes

Python scripts were developed in Spyder IDE[19] for data analysis. The two-sample Kolmohorov-Smirnov nonparametric statistical test[20] was applied to compare the sample distribution from the answers to “What sound sources do you expect to hear at this location?” asked after watching video only files of the scenes. Comparisons were made between the visual scenes that had the same audio information, including: S1L1 vs S1L4, S2L1 vs S2L2, S3 vs S4 and S1L2 vs S1L3. The statistical test was applied to understand if the visibility of a source (particularly of an industrial sound source) affected the audio expectations. Table 5 shows numbers of “Yes” responses to Q1; it also shows if there is a statistically significant difference between the scenes, considering a 0.95 confidence interval ($p<0.05$).

For industrial sounds, a statistically significant difference has been found between all pairs, except for S2L1 vs S2L2. The difference in the auditory expectations seemed to be affected by the visual presence of a plant. This is shown by S1L4 (the plant is clearly visible) and S3 (the view of the plant is dominant), compared to S1L1 and S4 (no plant is visible in both scenes), respectively. A degree of visibility of the industrial sound source seemed also to affect the auditory expectations, observed for the pair of the scenes S1L2 (a plant is slightly visible) vs S1L3 (the plant is clearly visible).

Table 5: Numbers of "Yes" answers to question Q1 What sound sources do you expect to hear at this location?

Sounds	Pairs of visual files							
	S1L1	S1L4	S2L1	S2L2	S3	S4	S1L2	S1L3
Traffic noise (e.g. cars, trains, aircraft)	33	21	34	33	17	32	25	24
p-value	0.02		0.99		0.003		0.81	
Sounds of human activity	25	19	32	20	18	28	25	26
p-value	0.6		0.04		0.04		0.99	
Natural Sounds	33	33	23	28	23	28	34	31
p-value	1		0.81		0.0005		0.99	
Industrial sound	2	34	20	21	30	1	20	34
p-value	0.0005		0.99		0.0006		0.003	

4.2. Effect of audio-visual interactions on industrial source identification and appraisal

The Kolmohorov-Smirnov nonparametric test was applied to compare the sample distribution from the responses to "To what extent did you hear the following sounds?" (Q2) with the responses to "To what extent did you see the following sound sources?" (Q3), both asked after watching the audio-visual representation of the same scene. The statistical test was applied only for industrial sounds and aimed to investigate if the *view* of the industrial source affected its aural identification, and also if the *sound* of an industrial source influenced the visual identification of the source. Correlations between the answers to Q2 and Q3 were also investigated. Numbers of the responses to both questions in each scene are shown in Figure 2. Statistically significant differences (p<0.05) between the answers to Q2 and Q3 distributions have been found for scenes 6, 7, and 8.

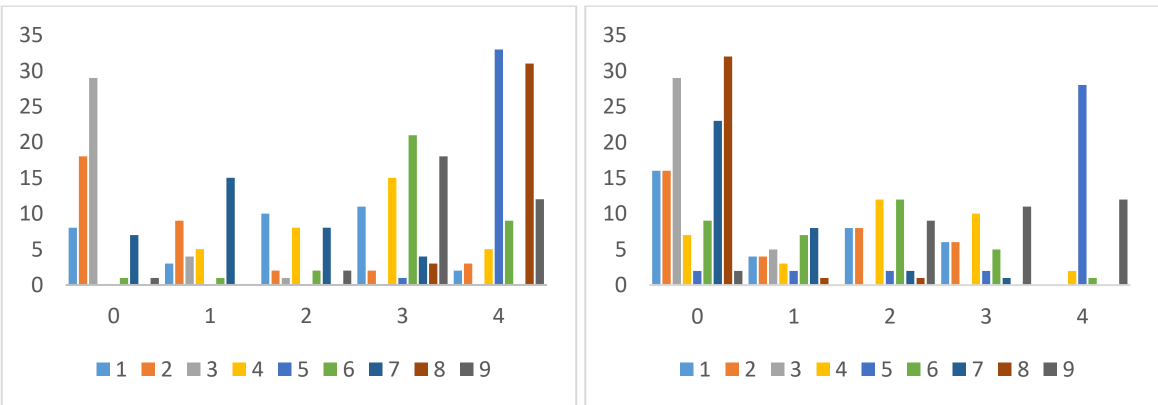


Figure 2: Numbers of responses to questions Q2 (left) and Q3 (right), where 0 – not at all, 1 - a little, 2 – moderately, 3 – a lot and 4 – dominates completely. The colours represent the scene number (see Table 2).

For scene 6, 21 participants replied that they heard the industrial sound source "a lot" and 9 as "dominated completely", while the visibility of the source was distributed between "not at all", "a little" and "moderately". In this scene the industrial plant was indeed audible but not visually dominant (see Table 1). For scene 7, on the contrary, in terms of hearing the industrial source, most of the responses were "a little" and "moderately", while 23 respondents said they did not see the source at all. In this scene the plant (mostly its tonal sound) was a little bit audible but the view was nearly totally masked by the trees. Finally, for scene 8, 31 participants indicated that the industrial source was the dominant audio source but 32 of them did not see it at all. In this scene the original audio file (mostly traffic noise) was swapped for a loud industrial plant (S3).

Results of a linear regression applied to the responses to Q2 and Q3 have shown a good correlation for scenes 3 and 5 (R = 0.99). In scene 3, the industrial source was nearly totally visually masked by vegetation, while the sound was totally masked by traffic. This was reflected in the responses since most of the participants neither saw nor heard the plant. In scene 5 the plant was both visually and aurally dominant and this was indicated by the respondents.

It is interesting to note that for scene 6 there is a very low correlation (R = 0.18) between Q2 and Q3 compared to scene 9 (R=0.6), even though the same audio file was used in both scenes. In scene 9, the plant was visually much closer and more dominant than in scene 6. This indicates that visual information overpowered the aural information: in scene 6 only 9 indicated that it was "the most audible source" while in scene 9 there were 12 answers in the same category.

4.3. Perceptual qualities of audio-visual scenes

Figure 3 shows distributions of the responses to questions Q4 – Q7 related to the overall perceptual quality of the audio-visual scenes.

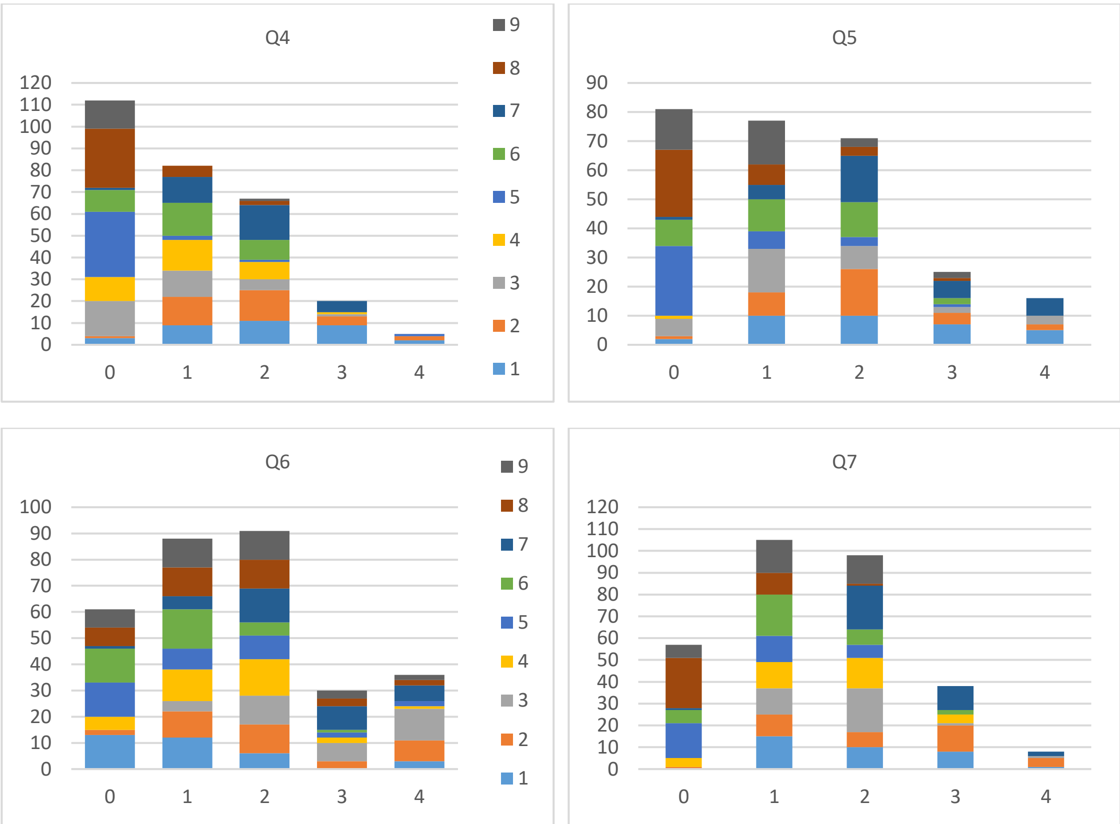


Figure 3: Percentages of the distributions of answers to Q4 – Q7, where colour represents the audio-visual scenes.

The distributions of the responses to Q4: "To what extent would you like to stay in the presented environment?" and Q5: "To what extent did you feel comfortable in the presented environment?" are skewed towards the left (low perceptual quality) for all scenes. This might be related to the relatively high sound levels which were above 50dBA at all the scenes (see Table 2). The responses to Q6 "To what extent do you think the industrial sound is acceptable in the presented environment?” and Q7: “Overall how would you describe the presented sound environment?” are also skewed toward the low end but with more responses as “bad” and “neither good nor bad” rather than “very bad”. Correlation coefficients between the responses to questions Q4 – Q7 are presented in Table 6.

Table 6: Correlation coefficients between Q4 - Q7 related to the overall perceptual qualities of the audio-visual scenes.

Scene	**1**				**2**				**3**				**4**				**5**			
/Quest	Q4	Q5	Q6	Q7	Q4	Q5	Q6	Q7	Q4	Q5	Q6	Q7	Q4	Q5	Q6	Q7	Q4	Q5	Q6	Q7
Q4	1.0				1.0				1.0				1.0				1.0			
Q5	0.9	1.0			0.8	1.0			0.7	1.0			0.8	1.0			0.2	1.0		
Q6	0.3	0.3	1.0		0.7	0.6	1.0		0.1	0.1	1.0		0.6	0.6	1.0		0.2	0.1	1.0	
Q7	0.8	0.7	0.4	1.0	0.6	0.7	0.7	1.0	0.6	0.6	0.1	1.0	0.6	0.7	0.8	1.0	0.3	0.6	0.5	1.0
Scene	**6**				**7**				**8**				**9**							
/Quest	Q4	Q5	Q6	Q7	Q4	Q5	Q6	Q7	Q4	Q5	Q6	Q7	Q4	Q5	Q6	Q7				
Q4	1.0				1.0				1.0				1.0							
Q5	0.8	1.0			0.7	1.0			0.8	1.0			0.6	1.0						
Q6	0.5	0.6	1.0		0.3	0.4	1.0		0.7	0.6	1.0		0.2	0.3	1.0					
Q7	0.7	0.6	0.7	1.0	0.6	0.5	0.7	1.0	0.8	0.7	0.7	1.0	0.5	0.6	0.4	1.0				

For scenes 1, 5 and 9, a relatively low correlation has been found between acceptability of the industrial plant (Q6) and the responses to all other three questions ($R \leq 0.5$). For scenes 2 and 4, acceptability of the industrial plant (Q6) seems to correlate well with answers to Q4, Q5 and Q7. For scene 6, Q6 correlates with Q5 and Q7, while for scene 8 it correlates with Q4 and Q5 ($R \geq 0.6$).

In scenes 1, 5 and 6, the responses for what extent the industrial sound source was acceptable (Q6) were mostly as “0 – not at all” and “1 – a little”. It is interesting that the overall level of scene 1 was 51.5dBA, while in scenes 5 and 6 it was 66dBA and 60dBA, respectively. In scenes 1 and 6 the views were mostly of the greenery and trees, while in scene 5 the view of the plant was dominant.

In scene 3, the responses skewed toward the higher acceptability rates – the scene had a generally low level of background sound and a faint tonal component of industrial sound with an overall level of about 68dBA. No dominant view of the industrial plant was present.

It seemed that the visual and aural information clearly affected the subjective judgements of the acceptability of the industrial plant. However, it was not clear if the participants made their judgements because it was already audible and/or visible, or based on the overall suitability to the environments.

5 DISCUSSION AND CONCLUSIONS

This paper examined the effects and interactions of sound source visibility and audibility on source identification and soundscape appraisal in virtual settings. Emphasis has been placed on the assessment of industrial sound sources. The effect of the visual and auditory information on overall perceived quality of a scene, and the acceptability of the industrial source within the scene were also investigated.

In terms of the audio expectations from a visual scene only, most of the participants identified the sound sources that were clearly visible in those visual scenes. For example, a person walking a dog, cars passing, roads, vegetation or a clearly visible plant building, However, this was not the case when the industrial source was either visually absent or partially masked by vegetation – the participants were less sure here and the answers were within the 50:50 range.

With regards to Q2 vs Q3, the results have shown that when the industrial source was both the most visible and most aurally dominant, the responses to the questions correlated well. It was also noted that when an industrial source was not at all visible but was most dominant aurally, the participants still identified it as the most aurally dominant (scene 5 vs scene 8).

All questions in Part 4 tended to be scored low for nearly all of the scenes. This may be due to the fact that the overall sound levels of all scenes were above 50dBA. More studies are needed to investigate the appraisal of soundscapes with lower sound levels. Additionally, the impact of tonal and impulsive components of industrial sound sources on soundscape assessment should be also examined.

The effect of the visual and aural scenes on the acceptability of the industrial plant should also be investigated further, distinguishing between scenes where the industrial sound source is clearly dominant, where it is absent and where visually masked by vegetation or other environmental or urban features.

The limitations of this study: participants were asked to access the scenes online using standard headphones, not in laboratory conditions with high-quality headphones. Differences in the quality of visual displays used and background sounds may have also affected the results.

Acknowledgements

The author of this paper would like to thank HEAD acoustics for lending their equipment and software and Anthony Shepperson and Miguel Garcia Pedroche of HEAD acoustics UK Ltd for in-depth technical advice and many fruitful conversations.

6 REFERENCES

1. H. L. S.-K. Lau, A review of audio-visual interaction on soundscape assessment in urban built environments, Applied Acoustics, Vol. 166, 107372, (Sept. 2020).
2. J. Luis Carlesa, I. Lopez Barrioa and J. Vicente de Lucio, Sound influence on landscape values, Landscape and Urban Planning 43, 191-200, (1999).
3. W. Yang and J. Kang, Soundscapes and sound preferences in urban squares, J. of Urban Design, 10, 69–88, (2005).
4. Z. Bangjun, S. Lili and D. Guoqing, The influence of the visibility of the source on the subjective annoyance due to its noise, Applied Acoustics, 64(12), 1205-1215, (2003).
5. L. Maffei, M. Masullo, F. Aletta and M. Di Gabriele, The influence of visual characteristics of barriers on railway noise perception, Science of the Total Environment, 445, 41- 47, (2013).
6. K. Sun, B. De Coensel, G. M. E. Sanchez, T. Van Renterghem, and D. Botteldooren, Effect of interaction between attention focusing capability and visual factors on road traffic noise annoyance, Applied Acoustics, 134, 16-24, (2018).
7. F. Aletta, M. Masullo, L. Maffei and J. Kang, The effect of vision on the perception of the noise produced by a chiller in a common living environment, Noise Control Engineering Journal, 64(3), 363-378, (2016).
8. A. Haapakangas, V. Hongisto and D. Oliva, Audio-visual interaction in perception of industrial plants – Effects of sound level and the degree of visual masking by vegetation, Applied Acoustics, Volume 160, 107- 121, (2020).
9. A Leiper, W. J. Davies and D. C. Waddington, Using soundscape assessment tools to determine the impact of industrial noise in quiet areas, 13th ICBEN Congress on Noise as a Public Health Problem, Sweden, (June 2021).
10. https://www.head-acoustics.com/eng/nvh_overview.htm.
11. HEAD acoustics, SQobold: www.head-acoustics.com/eng/nvh_sqobold.htm / ArtemiS SUITE: https://www.head-acoustics.com/eng/nvh_artemis_suite.htm
12. G. Stoet, PsyToolkit - A software package for programming psychological experiments using Linux, Behavior Research Methods, 42(4), 1096 - 1104, (2010).
13. G. Stoet, PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. Teaching of Psychology, 44(1), 24-31, (2017).

14. https://www.reaper.fm/
15. https://www.audacityteam.org/
16. BS 4142: 2014+A1:2019 Methods for rating and assessing industrial and commercial sound (2019).
17. G. Brown and J. Smyrnowa, Soundwalks in industrial areas: A pilot study, Acoustics Bulletin, 32-44, (July/August, 2020).
18. PD ISO/TS 12913-3:2019, Acoustics — Soundscape — Part 3: Data analysis.
19. https://www.spyder-ide.org/
20. F. J. Massey, Jr., The Kolmogorov-Smirnov Test for Goodness of Fit, J. Am. Statistical Association, 46 (253), 68 - 78, (March 1951).