How does visual context of urban open spaces drive soundscape pleasantness and eventfulness: A laboratory experiment using a higher order Ambisonics speaker system and an immersive virtual reality (IVR) head-mounted display (HMD)

Tin Oberman[1]
University College London, Institute for Environmental Design and Engineering
14 Upper Woburn Place, WC1H 0NN, London, United Kingdom

Andrew Mitchell[2]
University College London, Institute for Environmental Design and Engineering
14 Upper Woburn Place, WC1H 0NN, London, United Kingdom

Francesco Aletta[3]
University College London, Institute for Environmental Design and Engineering
14 Upper Woburn Place, WC1H 0NN, London, United Kingdom

Jian Kang[4]
University College London, Institute for Environmental Design and Engineering
14 Upper Woburn Place, WC1H 0NN, London, United Kingdom

ABSTRACT

Perception of sound environments is influenced by the context we experience them in. A big portion of the contextual information comes from the visual domain, including our understanding of what a place is but also from its visual features. A laboratory study combining questionnaire responses and eye-tracking tools was designed to investigate if the soundscape outcomes and participants' behaviour inside the simulation can be explained by the perceptual outcomes defined by visual information. 360 degree videos and First Order Ambisonics audio recordings of 27 different urban open spaces taken from the International Soundscape Database were used as the stimuli delivered via a Virtual Reality Head-Mounted Display and a Higher Order Ambisonics speaker array, while a neutral grey environment without sounds being reproduced represented the baseline scenario. A questionnaire tool was deployed within the IVR simulation to collect participants' responses describing their perception of the reproduced environments corresponding to the circumplex model featured in the Method A of the ISO/TS 12913-2. The results revealed a good coverage of the two-dimensional perceptual circumplex space and significant differences between perceptual outcomes driven by sound and those driven by visual stimuli.

[1] t.oberman@ucl.ac.uk

[2] a.mitchell@ucl.ac.uk

[3] f.aletta@ucl.ac.uk

[4] j.kang@ucl.ac.uk

1. INTRODUCTION

Perception of sound environments is influenced by the context we experience them in. This can include objective factors such as the building and/or landscape materials that influence sound propagation, other sounds that create a background for prominent sounds in a complex auditory scene but also personal factors such as our intentions in a specific place. A big portion of these contextual information comes from the visual domain and our understanding of what a place is. Context is one of the key elements in the definition of soundscape (1–3).

In the literature, many and various approaches can be found where the researchers looked at determining the extent and the nature of this relationship. These were mostly based on investigating associations between a set of perceptual soundscape metrics and another set of features that describe the context – either using objective indicators or subjective aspects of landscape quality.

Following the development of the standardised tools for soundscape data collection – ISO/TS 12913-2 (4), the questionnaire presented in the Method A, featuring the circumplex model and the 8 attributes describing soundscape quality, has emerged as the most common and effective soundscape assessment tool for providing perceptual metrics (5). It features eight scales to assess to what extent a soundscape is: pleasant, vibrant, eventful, chaotic, annoying, monotonous, uneventful or calm. The eight attributes mentioned can then be mapped onto a two-dimensional circumplex model defined by the two main orthogonal axes describing valence (pleasant vs annoying) and arousal (eventful vs uneventful), as developed by (6) and confirmed by (7). Besides the eight-scales the questionnaire also features assessment of: 1) the dominance of the sound source types, 2) the overall soundscape quality, 3) soundscape appropriateness. The assessment of the context, however, is not covered by this tool.

The key idea was that the Method A questionnaire can be used together with the Soundscapy visualisation tool (8) to investigate the effect which visual context has on soundscape perception in the most direct way. An experiment was designed where the Method A questionnaire was used to collect perceptual data in the following three scenarios: 1) no visual stimuli, audio stimuli only; 2) no auditory stimulus, visual stimuli only, 3) both visual and auditory stimuli. This was made possible thanks to the auralisation laboratory with a speaker array to provide auditory stimuli and an Immersive Virtual Reality (IVR) head-mounted display to provide visual context, as it was considered to be the most ecologically valid way of soundscape presentation (9).

2. METHODOLOGY

A laboratory study combining questionnaire responses and eye-tracking tools was prepared using 360 degree videos and calibrated First Order Ambisonics (FOA) audio recordings of urban open spaces from London (UK), Granada (ES), Venice (IT), Shenyang (CN) and Shenzhen (CN), recorded following the Soundscape Indices Protocol (10) and taken from the International Soundscape Database (11).

2.1. Pilot studies

Two pilot experiments were conducted: 1) Investigating the applicability of the Method A questionnaire to assess the perceived affective quality of the video scenes without the audio stimuli (12); 2) Optimising the stimuli selection to cover the full circumplex model associated with the eight dimensions employed in the Method A.

The first pilot experiment featured 39 different scenes and revealed a good coverage of the calm and eventful space. The annoying and uneventful dimensions were not covered which could mean that either visual context cannot generate a monotonous outcome in an urban space, or that the dataset is missing more stimuli that would address that segment of the circumplex space (12).

This led to further twelve field recordings made specifically to target the annoying, monotonous and uneventful area of the circumplex model.

2.2. Stimuli selection

The second pilot study was focused on providing three scenes per each section of the circumplex space in a three by three matrix. Three researchers conducted independent assessments of the 51 videos (the combined audio and video condition) to select the most representative 27 stimuli. After the independent assessment, the outcome was discussed until an agreement was reached. The selected scenes are presented in Table 1.

Table 1: Overview of the scenes used in the experiment. The L_{Aeq15s} values reported are in dB, as measured in the laboratory.

Session ID	Location	Description	Dominant sound sources	L_{Aeq15s}
BidderSt1	Bidder Street, London (UK)	Street, no trees, warehouse area	Traffic, ventilation	61.5
BlairSt1	Blair Street, London (UK)	Residential street	Aircraft, traffic	70.2
BlundellSt1	Blundell Street, London (UK)	Street, no trees, warehouse area	Engine noise	68.1
CaledonianPark1	Caledonian Park, London (UK)	Urban park	Birdsong, traffic	49.4
CamdenTown4	Camden Town undeground station entrance, London (UK)	Exit / entrance to the underground train station	Busker, traffic, people talking	79.5
CarloV2	Carlo V Palace entrance, Granada (ES)	Square in front of a historical palace	People talking	57.5
DadonSq3	Dadong Square, Shenyang (CN)	Square in a residential area	People talking, traffic	51
EustonTap3	Euston Square, London (UK)	Square next to a pub and a busy street	Traffic, siren, people talking	69.1
HereEspresso1	Here East, East Bay Lane, London (UK)	Outdoor caffée across primary	Children playing,	63.4

		school playground	people talking	
IvesRd2	Ives Road, London (UK)	Road, industrial area	Scrapyard crane	69.2
KingsfordRow1	Kingsford Row, London (UK)	Residential pedestrian street	Wind rustling in the trees	54.1
LianhuashanParkEntrance1	Lianhuashan Park, Shenzhen (CN)	Entrance to the large park	Music over PA, singing	76.1
MarchmontGardens4	Marchmont Gardens, London (UK)	Pocket park	Traffic, footsteps	51.4
NewRiverWalk1	New River Walk, London (UK)	Urban park	Birdsong	41.5
OlympicSq3	Olympic Square, Shenyang (CN)	Large waterfront park	People talking	47.8
PancrasLock2	St Pancras Lock, London (UK)	Promenade next to a canal lock	Waterfall	65.4
PingshanSt1	Pingshan Street, Shenzhen (CN)	Residential street	People talking	62.4
PlazaBibRambla1	Plaza Bib Rambla, Granada (ES)	Urban square, paved	People talking, footsteps	61.9
RegentsParkFields2	Regent's Park Broadwalk, London (UK)	Large urban park	People talking, traffic	46.7
RegentsParkJapan2	Regent's Park Japanese Garden, London (UK)	Small garden inside a large urban park	Birdsong, waterfall	47.9
RiverLeeSchools1b	Capital Ring, Olympic Park, London (UK)	River front	Aircrafts, birdsong	48.9
SanMarco1	Piazza San Marco, Venice (IT)	Urban square, paved	Music, people talking, footsteps	67.3
StephensonSt1	Stephenson Street, London (UK)	Road next to a railway,	Trucks passing by	77.2

		industrial area		
TateModern3	Tate Modern Garden, London (UK)	Waterfront promenade	Music, people talking and cheering	61.6
TorringtonSq4	Byng Place, London (UK)	Paved square	Music, traffic, people talking	61.5
WineOfficeCt1	Wine Office Court, London (UK)	Pedestrian street, no trees	Constructi on noises, footsteps	61.9
ZhongshanPark5	Zhongshan Park, Shenyang (CN)	Large urban park	Music	56

2.3. Laboratory setup

The experiment was conducted in a listening chamber featuring twelve coaxial active speakers (Genelec 8331) arranged to provide Second Order Ambisonic playback, while the visual stimuli were delivered via an Immersive Virtual Reality (IVR) Head-Mounted Display (HMD), as shown in Figure 1. The participant was required to be sitting on a highchair with their head in a "sweet spot" – as close as possible to the point that is equally distant from the centre of all the twelve speakers.

The questionnaire was deployed within the IVR simulation created in Unity and operated by a participant via joystick.

Figure 1: A participant taking on the experiment.

2.4. Procedure

The experiment consisted of three blocks within which the 27 stimuli were presented in random order, including a neutral grey environment without any sounds being reproduced which represented the baseline scenario. First block featured only auditory stimuli with a neutral grey environment presented in the HMD, the second block featured only 15s long video samples similar to (12), while the third block featured both audio and visual recordings combined. The participants were advised to take frequent breaks, one within each block but more were allowed upon request.

3. RESULTS

The preliminary results reveal a good coverage of the two-dimensional perceptual circumplex space and significant differences between perceptual outcomes driven by sound and those driven by visual stimuli. The perceptual outcomes in the two-dimensional circumplex space were calculated from the eight scales, relying on the recommendations outlined in the ISO/TS 12913-3(13) and visualised in the Figure 2 using the Soundscapy tool (8).

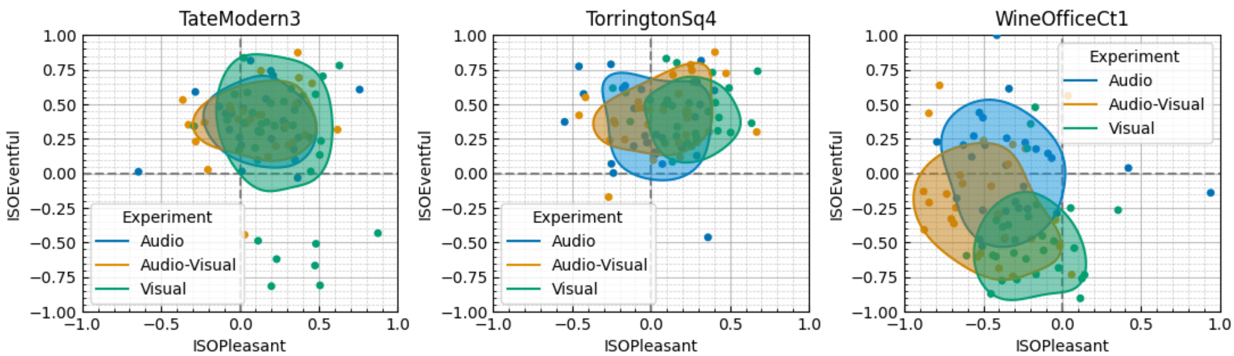

Figure 1: A sample from the results showing the perceptual outcomes for three locations. The blue, orange and green shapes represent the 50th percentile curves of the participants' responses for the audio-only block, audio-visual combined block and the visual-only block, respectively.

4. FINAL COMMENTS AND CONCLUSIONS

The results revealed a good coverage of the two-dimensional perceptual circumplex space and significant differences between perceptual outcomes driven by sound and those driven by visual stimuli. However, different trends were observed in how the auditory and visual stimuli interacted, changing the perceptual outcome in terms of: 1) direction in the two-dimensional circumplex space, 2) magnitude, 3) spread of responses. This has proven the importance and the complexity of the concept of context and that it cannot be reduced to a single environmental feature such as the amount of greenery at a location. A holistic understanding of non-auditory factors contributing to soundscape perception is necessary for effective and sustainable soundscape planning.

ACKNOWLEDGEMENTS

This research was funded through the European Research Council (ERC) Advanced Grant (no.740696) on "Soundscape Indices" (SSID). We want to thank Ms Xinrui Xu, Mr Daniel Perski and Ms Hui-Zhong Zhang for their support with data collection and experimental setup.

REFERENCES

1. International Organisation for Standardization. ISO 12913-1:2014 Acoustics — Soundscape — Part 1: Definition and conceptual framework [Internet]. International Organization for Standardization; 2014. Available from: https://www.iso.org/standard/52161.html

2. Kang J, Aletta F, Gjestland TT, Brown LA, Botteldooren D, Schulte-Fortkamp B, et al. Ten questions on the soundscapes of the built environment. Building and Environment. 2016 Nov;108:284–94.

3. Brown AL. A Review of Progress in Soundscapes and an Approach to Soundscape Planning. International Journal of Acoustics and Vibration. 2012;17(2):73–81.

4. International Organisation for Standardization. ISO/TS 12913-2:2018 Acoustics — Soundscape — Part 2: Data collection and reporting requirements [Internet]. 2018. Available from: https://www.iso.org/standard/75267.html

5. Aletta F, Torresin S. Adoption of ISO/TS 12913-2:2018 Protocols for Data Collection From Individuals in Soundscape Studies: an Overview of the Literature. Curr Pollution Rep. 2023 Oct 28;9(4):710–23.

6. Axelsson Ö, Nilsson ME, Berglund B. A principal components model of soundscape perception. The Journal of the Acoustical Society of America. 2010 Nov;128(5):2836–46.

7. Torresin S, Albatici R, Aletta F, Babich F, Oberman T, Siboni S, et al. Indoor soundscape assessment: A principal components model of acoustic perception in residential buildings. Building and Environment. 2020 Sep;182:107152.

8. Mitchell A, Aletta F, Kang J. How to analyse and represent quantitative soundscape data. JASA Express Letters. 2022 Mar 1;2(3):037201.

9. Xu C, Oberman T, Aletta F, Tong H, Kang J. Ecological Validity of Immersive Virtual Reality (IVR) Techniques for the Perception of Urban Sound Environments. Acoustics. 2020 Dec 25;3(1):11–24.

10. Mitchell A, Oberman T, Aletta F, Erfanian M, Kachlicka M, Lionello M, et al. The Soundscape Indices (SSID) Protocol: A Method for Urban Soundscape Surveys—Questionnaires with Acoustical and Contextual Information. Applied Sciences. 2020 Apr 1;10(7):2397.

11. Mitchell A, Oberman T, Aletta F, Erfanian M, Kachlicka M, Lionello M, et al. The International Soundscape Database: An integrated multimedia database of urban soundscape surveys -- questionnaires with acoustical and contextual information [Internet]. Zenodo; 2022 [cited 2024 Feb 2]. Available from: https://zenodo.org/record/6331810

12. Xu X, Oberman T, Mitchell A, Huang S, Aletta F, Kang J. Effects of Visual Stimulus on Soundscape Perception: A Pilot Study Using Immersive Virtual Reality (IVR). In: Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023 [Internet]. Turin, Italy: European Acoustics Association; 2024 [cited 2024 Feb 2]. p. 1047–54. Available from: https://dael.euracoustics.org/confs/landing_pages/fa2023/000732.html

13. International Organisation for Standardization. ISO/TS 12913-3:2019 Acoustics — Soundscape — Part 3: Data analysis [Internet]. 2019. Available from: https://www.iso.org/standard/69864.html