PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Prediction models of desirable levels of birdsong and water sound in a noisy environment: A laboratory experiment based on virtual reality

Zhen-Ting, Ong[1]; Joo-Young, Hong[1]; Bhan, Lam[1]; Kenneth Ooi[1]; Woon-Seng, Gan[1]; Samuel, Yeong[2]; Irene, Lee[2]; Sze-Tiong, Tan[2]

[1] School of Electrical & Electronic Engineering, Nanyang Technological University, 639798, Singapore

[2] Building & Research Institute, Housing & Development Board, 738973, Singapore

ABSTRACT

Augmenting pleasant natural sounds in the noisy urban environment is a key strategy in soundscape design. There have been numerous studies on the positive effects of natural sounds on soundscape quality. However, little attention has been directed to predictive models that suggest appropriate levels of natural sounds at specific ambient noise levels. These models provide a blueprint for practical soundscape design. This study, thus, aims to develop prediction models of desirable natural sound levels (birdsong and water sound) to enhance the soundscape quality through laboratory experiments based on virtual reality. The laboratory test consists of two steps (I and II). In step I, participants were instructed to evaluate traffic sound scenes, ranging from 60 to 70 dB, in terms of perceived loudness of noise (PLN) and overall soundscape quality (OSQ). In step II, the participants were instructed to adjust the audio levels of bird or water sounds augmented to each traffic scene to the most desirable levels while considering the preceding PLN of traffic and OSQ. Based on the results, the soundscape predictive models were developed using acoustic indicators to predict the desirable natural sound levels corresponding to traffic noise levels.

Keywords: Soundscape, Virtual Reality, Ecological Validity, Parametric decoding, Ambisonics

1 INTRODUCTION

Numerous studies have demonstrated the viability of adding natural sounds to a noisy urban environment as a soundscape design strategy [1–4]. Of the plethora of natural sounds, water sounds and bird songs have been found to be more suitable in an urban context [5]. Despite a fairly large number of subjective studies on the effects of natural sounds in noisy urban environments, there has been little focus on the development of predictive models to determine the appropriate levels of natural sounds to be augmented in an urban environment. Moreover, these studies have relied on subjective assessments on a pre-determined range of natural sound levels.

To gain further insight into the most preferred sound levels, the method of adjustment method is adopted in this study to allow the participants to control the natural sound stimulus. Since traffic noise is the most prevalent [6], this study focuses on the most persistent source of traffic noise – expressway noise. As a preliminary study, an urban residential location flanked by a minor road and an expressway was recorded and reproduced in cinematic virtual reality (VR) for evaluation. The natural sounds were spatialised and augmented to the VR scene acoustically (i.e. through headphones) and visually (i.e. rendered 3D objects in video). The influence on the preferred natural sound levels of both a generic and natural visual representation of the sound sources were investigated.

To achieve increased ecological validity, spatial audio recording and reproduction techniques have been increasingly adopted in VR-based soundscape evaluation [7–9]. Owing to its reproduction-format-agnostic flexibility, the ambisonic format has emerged as a popular choice. However, linear decoding of the first-order representation of the ambisonic format (FOA) possess inherent issues such as timbral coloration in the higher frequencies and poor localization of directional sounds. Adoption of high-order formats of at least third- up to fifth-order ambisonics have yielded statistically higher timbral quality and localisation accuracy [10–13]. Moreover, perceptually motivated parametric decoding methods, such as the COMPASS framework, have been developed to further enhance the perceptual quality of ambisonic recordings [10,14]. To achieve a balance between cost-effectiveness and perceptual quality, the third-order ambisonic format decoded to head-tracked binaural with the parametric COMPASS framework is adopted in this study.

Therefore, this study investigates the preferred levels of two types of natural sounds (bird and stream) augmented to different sound levels of traffic in a cinematic VR environment of a recorded urban space. Participants were instructed to adjust the levels of the presented natural sounds to the most desirable levels while considering both the perceived loudness of traffic and the overall soundscape quality in terms of pleasantness. The subjective responses were then analysed to form a basic framework for a predictive model to determine the most suitable natural sound levels corresponding to different traffic noise levels.

2 METHOD

2.1 Experimental Design

Participants were presented with an audio-visual recording of an outdoor traffic scene through a VR HMD and headphones in a quiet listening room. For each traffic sound level, the participants were chronologically instructed to: (I) evaluate the traffic scene, (II) adjust the sound levels of a natural sound that is augmented to the same traffic scene to the most desirable level. These two steps were repeated a total of 12 times (3 traffic scenes × 2 natural sounds × 2 visual stimuli) in a randomised order, as described in Table 1.

Table 1 – Description of the 12 test cases

Traffic Sound Levels (dBA)	Natural Sounds	Sound source stimuli
60	Bird	Speaker
60	Bird	Bird
60	Water	Speaker
60	Water	Fountain
65	Bird	Speaker
65	Bird	Bird
65	Water	Speaker
65	Water	Fountain
70	Bird	Speaker
70	Bird	Bird
70	Water	Speaker
70	Water	Fountain

2.2 Stimuli

The audio-visual environment an outdoor traffic scene was recorded in the Nanyang Technological University campus with a spherical panoramic camera (Garmin VIRB 360 Action Camera, USA) and a third-order ambisonics microphone array (ZYLIA ZM-1, Poland). The audio-visual capturing system was mounted on a tripod and fixed at approximately 1.6 m off the ground. The videos were recorded in 4K resolution and post-processed for white balance and exposure compensations (Adobe Premiere Pro CC 2019). The final video rendered maintains the 4K resolution but encoded in the H.265 format with a bit rate of 25 Mbps at 30 frames per second.

2.2.1 Visual Stimuli

The traffic scene recorded was an open area between two student hostels next to a minor road, as shown in Figure 1(a). The minor road is adjacent to an expressway hidden by greenery in the background. The visual stimuli in Figure 1(a) is presented through a VR HMD (Pimax 5K Plus, China) via the GoPro VR player to the participants during the evaluation of the recorded traffic scene.

To investigate the effects of augmenting the traffic scene with natural sounds, a 3D object representing the respective sound source is rendered into the scene. Effects of both a generic speaker sound object and corresponding natural objects were investigated, as shown in Figure 1(b) through 1(c). Therefore, each natural sound audio stimulus is either paired with the generic speaker or its corresponding natural object (i.e. water sounds is paired with a water fountain).

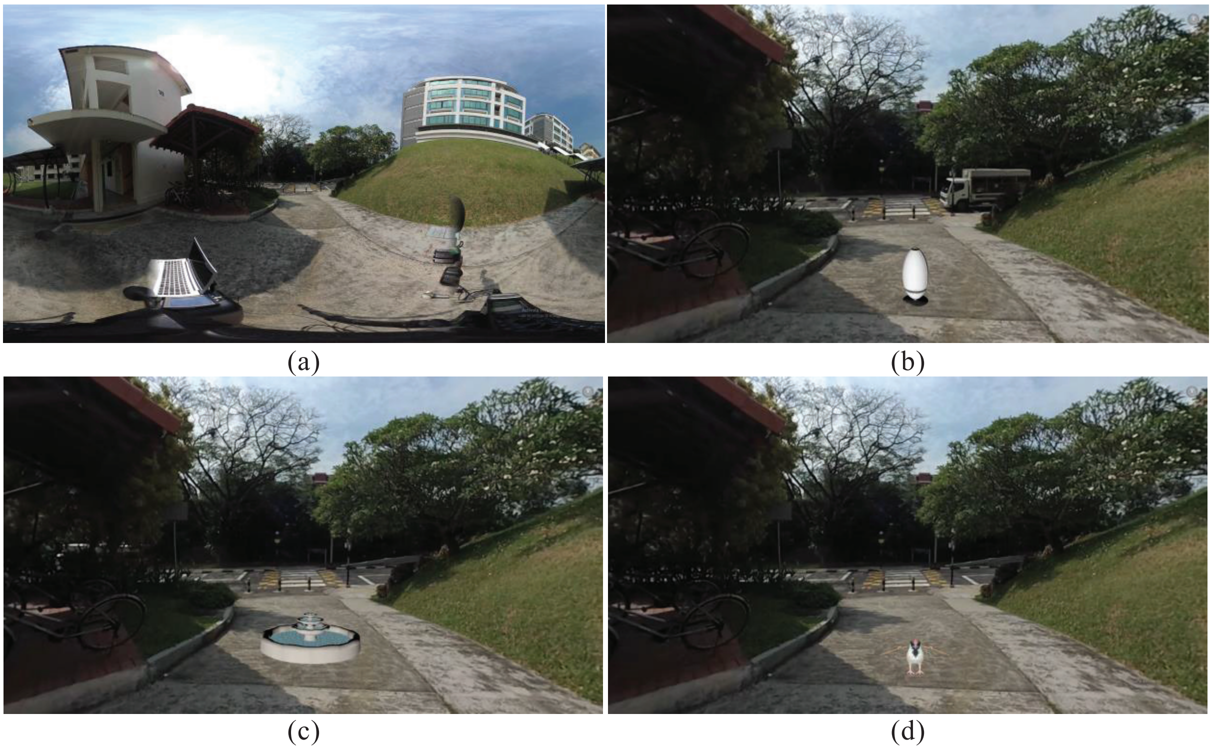

Figure 1 – (a) Equirectangular video still of the traffic scene. Video stills of the same traffic scene with 3D objects rendered into the scene showing (b) a generic speaker, (c) a water fountain, and (d) a sparrow.

2.2.2 Audio Stimuli

To achieve a higher spatial fidelity, both the recorded traffic and spatialised natural sound ambisonic tracks in this study are encoded natively in third-order ambisonics. The recorded traffic scene is encoded in third-order ambisonics from a 19-channel track (Zylia Ambisonics Converter Plugin, Poland). Both the natural sounds employed, a stream sound and a sparrow chirping, are mono tracks that were spatialised into third-order ambisonic tracks via a plugin (Facebook Spatial Workstation, USA). The spatialised natural sound tracks were oriented to the location of the sound sources (i.e. speaker, bird or fountain), where the azimuth and elevation were set to zero and the distance from the listening position was 2 m. After encoding to third-order ambisonics, the tracks are decoded parametrically using the COMPASS framework to head-tracked binaural tracks presented over headphones. The subjective experiment was administered through a digital audio workstation (Reaper, USA) linked via open sound control to the GoPro VR player using a custom workflow.

The 1-min equivalent sound pressure level of the recorded traffic scene in 2.2.1 was calibrated to three levels (60 dB, 65 dB and 70 dB) using a KEMAR artificial head. Hence, participants evaluated the same traffic visual stimulus at three sound levels.

To prevent discomfort and to provide a substantial upper limit (SNR of +10dB at 70 dB traffic level), the 1-min equivalent sound levels of both natural sounds were designed to attain a maximum of 80 dB during the adjustment stage (i.e. step II in 2.1).

After each participant has completed the subjective test, the natural sound levels chosen by the method of adjustment is played through the same headphones and measured with the KEMAR artificial head. The preferred signal-to-noise ratio of each participant is then determined.

2.3 Procedure

Five participants with normal-hearing (0.125, 0.5, 1, 2, 3, 4, 6, and 8 kHz average pure-tone thresholds < 20 dB HL) and exposure to the field of audio or acoustics research were recruited for this study. Formal ethical approval was granted by the institutional review board of Nanyang Technological University (NTU) for this study (IRB-2018-02-024). Informed consent was obtained from each participant prior to the start of the experiments.

For each of the 12 natural sound and 3D object pairs at different sound levels, the participants were instructed verbally to provide their subjective assessments in two steps as described in 2.1. In step I, the participants evaluated the overall soundscape quality and perceived loudness of the traffic sound for a duration of 1 min. To judge the perceived loudness of the traffic sound, participants were asked: "What is your overall impression of the traffic sound in terms of its loudness?". The overall soundscape quality was assessed based on the perceived pleasantness of the sound environment, where participants were asked: "How pleasant is the overall sound environment?".

In step II, a randomly chosen 3D object and natural sound pair was projected through the VR HMD and headphones to the participant, respectively. Participants were instructed to adjust the sound levels of the natural sound to their most comfortable levels while considering the preceding overall soundscape quality and perceived loudness of the traffic sound.

3 RESULTS

Mean preferred signal-to-noise ratios (SNRs) as a function of the traffic noise levels for birdsong and water sounds are plotted in Figure 2(a) and 2(b), respectively. No significant differences in the preferred SNR were observed between natural 3D object (i.e., bird/fountain) and the speaker 3D object. The mean preferred SNR were lower than 0 dB across the three different traffic noise levels for both birdsong and water sounds. This supports the findings in previous studies that similar or 3 dB lower sound pressure levels (SPLs) of natural sounds were preferred in combination with urban noises [15,16]. Overall, there were moderately negative correlations between the preferred SNRs and traffic noise for birdsong and water sounds.

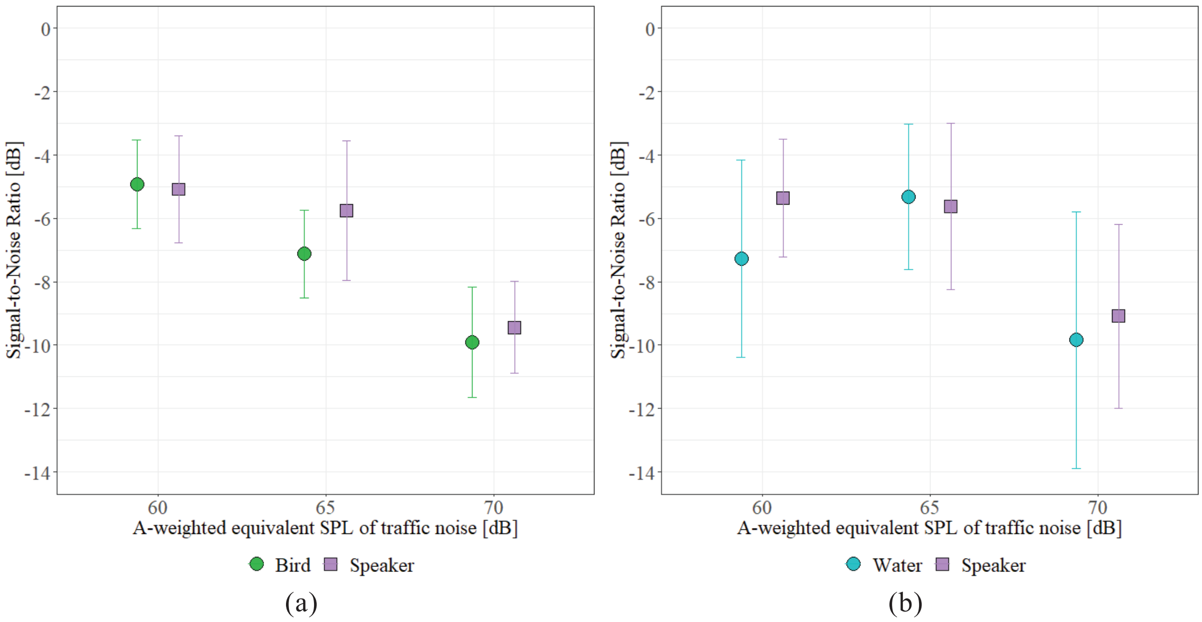

Figure 2 – Signal-to-Noise Ratio as a function of traffic noise levels: (a) birdsong and (b) water sound. Error bars indicate standard errors.

Mean rating scores of the PLN for traffic noise-alone, traffic noise combined with birdsong and water sounds as a function of the traffic noise levels were plotted in Figure 3(a) and 3(b), respectively. The PLN increased as the traffic noise increased. It was found that introducing natural sounds (i.e., birdsong or water sounds) could reduce the PLNs of traffic noises at 60 dB and 65 dB, while the effect of natural sounds on reducing the PLN was relatively small when the traffic noise was 70 dB, which corresponds well with findings of a previous study [1]. In terms of the type of 3D object, there were no significant differences in reduction of the PLN between actual sound hologram (i.e., bird / fountain) and speaker hologram.

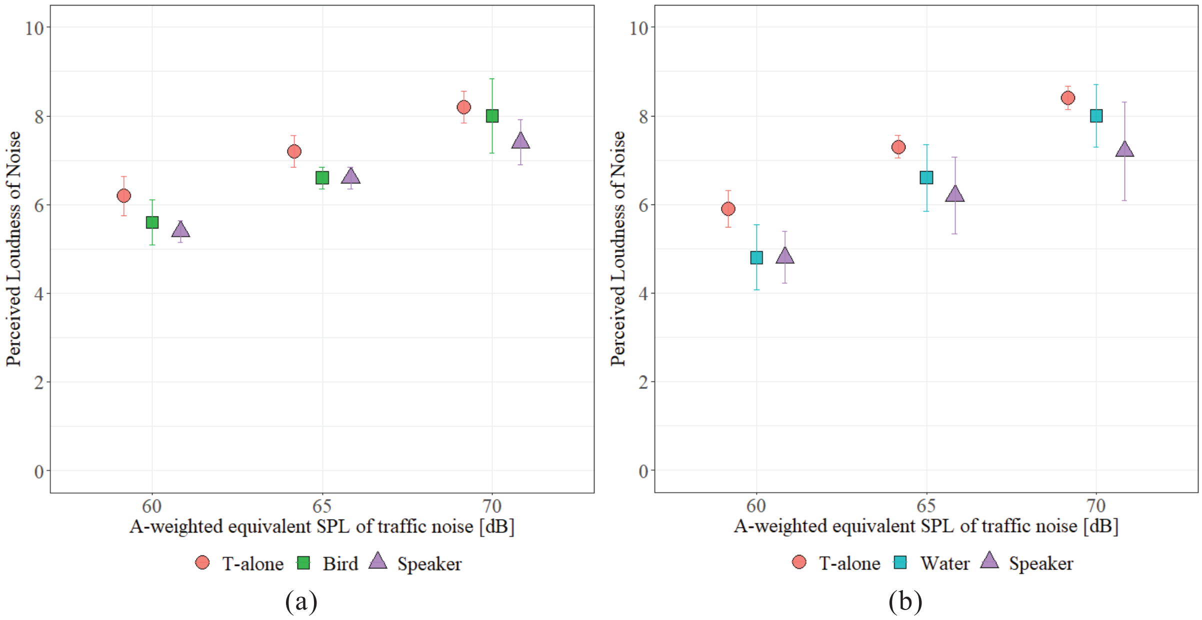

Figure 3 – Mean rating score of perceived loudness of noise as a function of traffic noise levels: (a) birdsong and (b) water sound. Error bars indicate standard errors.

Figure 4 shows the mean rating scores of the OSQ for the acoustic stimuli (traffic noise-alone, traffic noise combined with birdsong and water sounds) across three different traffic noise levels. As expected, the OSQ scores decreased as the traffic noise increased. In comparison, it seems that the effect of natural sounds on OSQ was relatively greater than that on PLN. It was observed that the effects of birdsong and water sounds were significant for enhancing the OSQ for all the traffic noise levels. There were no significant differences between birdsongs and water sounds as shown in Figure 4(a) and 4(b), respectively. Similarly, no significant difference between the natural 3D object and speaker 3D object were found in OSQ.

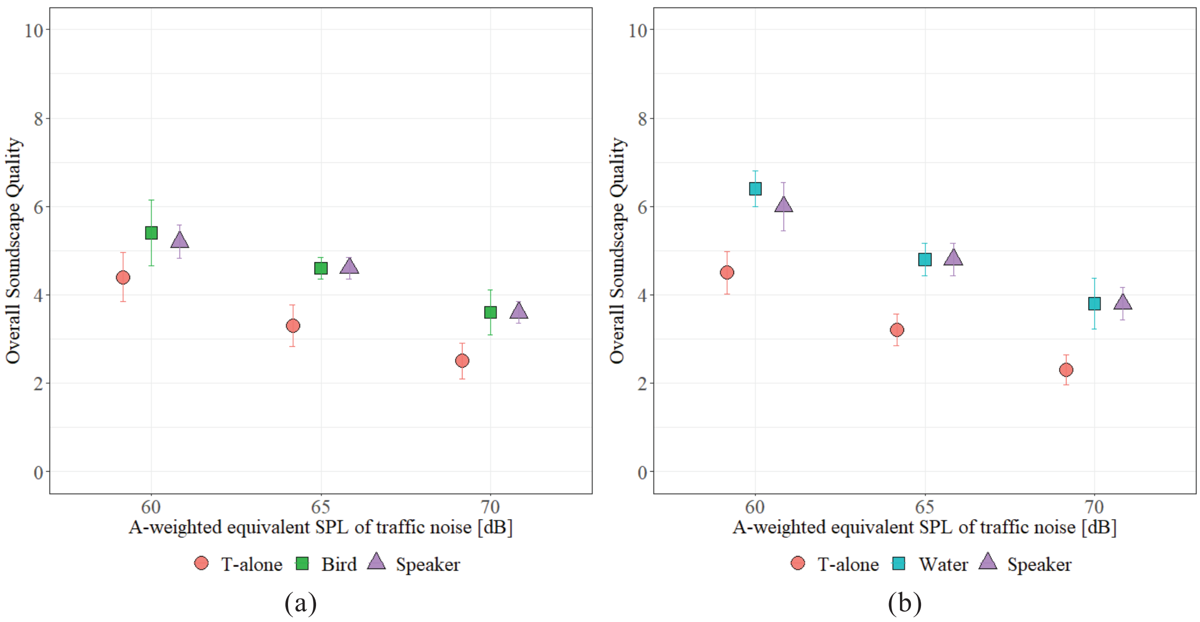

Figure 4 – Mean rating score of overall soundscape quality as a function of traffic noise levels: (a) birdsong and (b) water sound. Error bars indicate standard errors.

4 CONCLUSIONS

Although there were limited participants were recruited for this preliminary study, the trends observed in this VR-based evaluation with parametric spatial audio reproduction are in agreement with previous audio-only evaluations. The participants experienced an overall reduction in perceived loudness and improvement in soundscape quality across all traffic noise levels and with both natural sound types. A lower natural sound level (i.e., negative SNR) was also preferred across all traffic noise levels.

There are also encouraging signs indicating that the introduction of man-made sound sources (i.e., speakers) do not affect the perceived loudness and pleasantness of the acoustic environment. Besides increasing the number of subjective responses, a larger range of traffic sound levels and locations could be included to obtain a deeper insight.

ACKNOWLEDGEMENTS

This material is based on research/work supported by the Singapore Ministry of National Development and National Research Foundation under L2 NIC Award No.: L2NICCFP2-546 2015-5. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Singapore Ministry of National Development and National Research Foundation, Prime Minister's Office, Singapore. Joo Young Hong is currently supported by the 2017 Lee Kuan Yew Postdoctoral Fellowship (Grant no. 50025000).

REFERENCES

1. Hong, J.Y.; Lam, B.; Ong, Z.; Gupta, R.; Gan, W.; Chong, S.H.; Feng, J. Appropriate levels of natural sounds to enhance soundscapes in urban areas. *INTER-NOISE NOISE-CON Congr. Conf. Proc.* **2017**, 49–53.
2. De Coensel, B.; Vanwetswinkel, S.; Botteldooren, D. Effects of natural sounds on the perception of road traffic noise. *J. Acoust. Soc. Am.* **2011**, *129*, EL148-L153.
3. Axelsson, Ö.; Nilsson, M.E.; Hellström, B.; Lundén, P. A field experiment on the impact of sounds from a jet-and-basin fountain on soundscape quality in an urban park. *Landsc. Urban Plan.* **2014**, *123*, 46–60.
4. Hao, Y.; Kang, J.; Wörtche, H. Assessment of the masking effects of birdsong on the road traffic noise environment. *J. Acoust. Soc. Am.* **2016**, *140*, 978–987.
5. Hong, J.Y.; Lam, B.; Ong, Z.T.; Gupta, R.; Gan, W.S. Suitability of Natural Sounds To Enhance Soundscape Quality in Urban Residential Areas. In Proceedings of the 24th International Congress on Sound and Vibration; London, UK, 2017.
6. World Health Organization Regional Office for Europe *Environmental Noise Guidelines for the European Region*; The Regional Office for Europe of the World Health Organization: Copenhagen, Denmark, 2018; ISBN 978 92 890 5356 3.
7. Guastavino, C. Ecological Validity of soundscape reproduction. *Acta Acoust. United With Acoust.* **2004**, *91*, 0–19.
8. Hong, J.Y.; Lam, B.; Ong, Z.-T.; Ooi, K.; Gan, W.-S.; Kang, J.; Feng, J.; Tan, S.-T. Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications. *Build. Environ.* **2019**, *149*, 1–14.
9. Hong, J.Y.; He, J.; Lam, B.; Gupta, R.; Gan, W.-S. Spatial Audio for Soundscape Design: Recording and Reproduction. *Appl. Sci.* **2017**, *7*, 1–21.
10. *Parametric Time-Frequency Domain Spatial Audio*; Pulkki, V., Delikaris-Manias, S., Politis, A., Eds.; John Wiley & Sons, Ltd: Chichester, UK, 2017; ISBN 9781119252634.
11. Thresh, L.; Armstrong, C.; Kearney, G. A Direct Comparison of Localisation Performance When Using First, Third and Fifth Order Ambisonics For Real Loudspeaker And Virtual Loudspeaker Rendering. In Proceedings of the Audio Engineering Society Convention 143; Audio Engineering Society: New York, NY, USA, 2017; pp. 1–9.
12. Bates, E.; Dooney, S.; Gorzel, M.; O'Dwyer, H.; Ferguson, L.; Boland, F.M. Comparing Ambisonic Microphones: Part 1. In Proceedings of the Audio Engineering Society Conference on Sound Field Control; Audio Engineering Society: Guildford, UK, 2016.
13. Bates, E.; Dooney, S.; Gorzel, M.; O'Dwyer, H.; Ferguson, L.; Boland, F.M. Comparing Ambisonic Microphones: Part 2. In Proceedings of the 142th Audio Engineering Society Convention; Audio

Engineering Society: Berlin, Germany, 2017.
14. Politis, A.; Tervo, S.; Pulkki, V. COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); IEEE, 2018; pp. 6802–6806.
15. Jeon, J.Y.; Lee, P.J.; You, J.; Kang, J. Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds. *J. Acoust. Soc. Am.* **2010**, *127*, 1357–1366.
16. Galbrun, L.; Ali, T.T. Acoustical and perceptual assessment of water sounds and their use over road traffic noise. *J. Acoust. Soc. Am.* **2013**, *133*, 227–237.