

## *A mixed-reality approach to soundscape assessment of outdoor urban environments augmented with natural sounds<sup>^</sup>*

Joo Young HONG<sup>1</sup>, Bhan LAM<sup>2\*</sup>, Zhen-Ting ONG<sup>2</sup>, Kenneth OOI<sup>2</sup>, Woon-Seng GAN<sup>2</sup>, Jian KANG<sup>3</sup>, Samuel YEONG<sup>4</sup>, Irene LEE<sup>4</sup> and Sze-Tiong TAN<sup>4</sup>

### **Abstract**

To investigate the effect of augmenting natural sounds in noisy environments, an in-situ experiment was conducted using a mixed-reality head-mounted display (MR HMD). Two outdoor locations close to an expressway were selected for the experiment. A natural sound (birdsong or stream) along with a hologram (sparrow/fountain or loudspeaker) was projected through the MR HMD. Participants were asked to adjust the natural sound levels to their preferred level under ambient traffic noise conditions at each location. Participants also assessed the perceived loudness of traffic (PLN) and overall soundscape quality (OSQ) in conditions with and without the augmented natural sounds. The results showed that both natural sounds significantly reduced the PLN and enhanced the OSQ. No significant differences in subjective responses were found between the loudspeaker and visual representations of the natural sound source as holograms. Analysis on the preferred signal-to-noise ratio (SNR), i.e. ratio of natural sound to traffic levels, indicated a strong negative correlation between the preferred SNRs and ambient traffic noise levels. Overall, the preferred SNR of the birdsong was significantly higher than that of the water sound. Among the acoustic parameters tested, the A-weighted traffic noise level was the strongest predictor for the preferred SNR of both the birdsong and water sound. However, the correlation for the water sound was relatively higher than the birdsong. This was due to the larger variance in the subjective evaluation for the birdsong.

**Keywords:** Soundscape; Soundscape intervention; Acoustic environment; Natural sounds; Augmented reality; Mixed reality

Submitted to Elsevier Building and Environment on 30 July 2020, Revised 19 January 2021, Accepted 6 February 2021, Available online 12 February 2021.

---

<sup>^</sup> The research protocols used in this research were approved by the institutional review board of Nanyang Technological University (NTU), Singapore [IRB-2018-02-024].

<sup>1</sup> Department of Smart City Architectural Engineering, Chungnam National University, Daejeon 34134, Republic of Korea

<sup>2</sup> School of Electrical & Electronic Engineering, Nanyang Technological University, 639798, Singapore.

\* Send correspondence to: Bhan Lam (blam002@e.ntu.edu.sg)

<sup>3</sup> UCL Institute for Environmental Design and Engineering, The Bartlett, University College London, Central House, 14 Upper Woburn Place, London WC1H 0NN, United Kingdom.

<sup>4</sup> Building & Research Institute, Housing & Development Board, 738973, Singapore

## 1. *INTRODUCTION*

As noise exposure adversely affects human health and well-being, addressing noise pollution has become critical to the development of a sustainable urban environment [1–3]. Current environmental policies on noise generally focus on mitigation or regulation based on sound pressure level (SPL) measurements. However, there is a growing body of evidence showing that reduction in SPL neither necessarily satisfies people nor improves their quality of life [4,5]. Furthermore, conventional noise control measures (e.g., low noise pavement, noise barriers, or earth mounds) are not always feasible and cost-effective [4].

Recently, the soundscape approach, which considers sound as a resource to design an acoustic environment in the context of a place, has gained traction as a new paradigm for sustainable urban sound management [4,6–8]. In particular, one of the representative soundscape design approaches is by employing pleasant natural sounds as acoustic design elements to mask environmental noise in outdoor areas.

In the past decade, numerous studies have investigated the effects of water sounds [9–15] and birdsongs [16–20] for enhancing acoustic comfort in urban areas. Some have concluded that the sound level of these natural sounds is a significant indicator of a desirable soundscape. Jeon et al. [11] and You et al. [14] found that similar or 3 dB lower natural sound levels were preferred when natural sounds were augmented with urban noise at levels of 55 dB or 75 dB. Galbrun and Ali also reported similar observations, in which the participants preferred the

augmented water sound levels to be lower than traffic noise at 55 dB [15]. In addition, Hong et al. [17] showed that appropriate natural sound levels considering perceived loudness and soundscape quality could differ according to noise levels; when urban noises are 55 and 65 dB, similar or 3 dB higher natural sound levels were evaluated as appropriate, whereas when urban noise became 75 dB, natural sound levels 6 dB lower than the noise were the most desirable [17].

### **1.1 *Ecological validity of investigations into the effects of natural sounds***

To date, a large proportion of studies have investigated the effects of natural sounds on soundscapes based on auditory experiments conducted in controlled laboratory conditions. These auditory experiments are limited in reflecting real-world settings because the soundscape perception is multisensory [21,22]. Even though a laboratory experiment provides a controlled setting to study experimental variables, which allows such studies to yield more consistent results and thus allows for enhanced repeatability, these results have suboptimal generality to real-world scenarios, thereby resulting in low ecological validity. In contrast, in-situ experiments can provide a realistic representation of real-world settings, which can guarantee high ecological validity of results obtained [23–26]. Since it is generally difficult to control independent variables such as natural sound levels or the type of natural sounds in situ, few studies have adopted in-situ experiments to explore the effects of natural sounds on soundscape [9,27,28]. Axelsson et al. [9] conducted a field experiment to examine the effect of water

sound from a real water fountain on perceived soundscape quality in an urban park by comparing the soundscapes when the fountain was turned on or off. Cerwén [28] also investigated the effect of forest sounds played via a loudspeaker installed in an arbor on an urban street by comparing situations with and without the forest sounds playing. Both studies showed that natural sounds had positive effects on enhancing soundscape quality in real-world settings. In these field studies, traffic noises and natural sound levels were varied by changing the position of respondents from sound sources (traffic noise, or natural sound sources). For in-situ studies, however, there are still inherent limitations in independently controlling each sound variable (i.g., traffic noise, or natural sound levels) to explore their relationships.

Furthermore, soundscape intervention by augmenting pleasant natural sounds to existing soundscape can be implemented in two ways: deployment of real sound sources (e.g., vegetation or fountains) [9,18,29] or by installing active loudspeaker systems [27,28,30]. Although soundscape design using real sources would be the most natural way to introduce pleasant sounds, deployment of water features is not always feasible and planting vegetation does not guarantee the presence of birdsongs. Hence, adding natural sounds via a loudspeaker system could be a viable soundscape design strategy. However, the loudspeaker system may result in incongruency in the audio-visual environment. Many studies have reported that overall audio-visual coherence in the context of the place is a critical factor affecting soundscape [22,31,32]. Particularly, Hong et al. [33] have found that the

perceptions of natural sounds (e.g., pleasantness and appropriateness) are influenced by the coherency of audio-visual elements. To date, there have been no reports comparing the effects of natural sounds originating from real sound sources (e.g., bird/fountain) and loudspeaker systems in situ. Hence, it is important to investigate whether there are differences in terms of perceptions of natural sounds between presenting loudspeaker and real sound source images.

## **1.2 Adoption of immersive display technologies in soundscape assessment**

To achieve higher ecological validity in laboratory conditions, there has been a steady adoption of virtual reality (VR) in soundscape research [31], especially in combination with spatial audio [23,26,34,35]. In general, VR refers to an immersive three-dimensional environment that is either entirely computer-generated, a cinematic reproduction from an omnidirectional camera, or a fusion of both.

While VR content is increasingly consumed on VR head-mounted displays (HMDs), these devices have recently been marketed alongside augmented reality (AR) or mixed reality (MR) HMDs. For clarity, the definitions and characteristics of VR, AR, and MR in the recent CTA-2069 standard [36] are henceforth adopted throughout this paper, as listed in Table A.1 in Appendix A.

Since audio-visual interaction significantly influences the perception of soundscape [31] and landscape [35], it is important to understand the visual characteristics of these emerging immersive display technologies. In particular, AR is differentiated from MR in that the digital audiovisual components in AR

are superimposed on rather than integrated into the user's environment (see Table A.1). The ability of MR displays to seamlessly blend the real-world environment with digital content that merges into and interacts with the real world provides an avenue for immersive, in-situ soundscape design, and renovation [31].

### **1.3 *Mixed reality approach***

As there are still few MR HMDs in the market, there have been few reported cases of the MR approach in soundscape research [37,38]. Nevertheless, an MR HMD potentially enables the replication of laboratory-based investigations of natural sounds in in-situ environments with high ecological validity. Therefore, this study adopts the MR approach to investigate the effects of virtual audio-visual integration of natural sounds to in-situ soundscapes.

Specifically, the effects of natural sound types (birdsong or water sounds), ambient traffic noise levels, and visual images of sound sources (images of real sound source or loudspeaker) on the subjective assessment of soundscape attributes are investigated. Three soundscape attributes were evaluated: the perceived loudness of the soundscape, the soundscape quality, and the preferred natural sound level. These attributes have been considered as important perceptual attributes of soundscapes in previous studies [11,17,24]. Particularly, relationships between the preferred natural sound level and the acoustic characteristics of the background traffic noise are explored. This is because generating appropriate natural sound levels in response to ambient noises is

hypothesized as a key design factor for soundscape intervention by natural sounds.

#### **1.4 Aims of the study**

Therefore, three research questions are formulated: (1) Do the three factors (i.e., natural sound types, background traffic noise levels, and visual images of the sound source) influence the evaluation of soundscape attributes (i.e., perceived loudness noise, overall soundscape quality, and preferred natural sound level)? (2) What are the critical acoustic parameters of ambient traffic noise to predict the preferred natural sound levels? (3) Are there differences in assessment of the soundscape attributes between in-situ and laboratory conditions? To answer the questions, an in-situ experiment was conducted using a MR HMD and acoustic measurement devices. The methodology for conducting this experiment is presented in Section 2. The results pertaining to the first and second research questions are analyzed in Sections 3.1 to 3.3, and 3.4, respectively. Furthermore, regarding the third research question, the results of this study are compared with previous studies conducted in laboratory experiments in Section 4.1. Lastly, the limitations and implications of this study are discussed in Sections 4.2 and 4.3.

## **2. METHOD**

The characteristics of the study area are first described, followed by an analysis of its background traffic noise levels. Subsequently, the formulation of the audio-visual stimuli, equipment used, subjective evaluation, and experimental design for

the mixed-reality HMD platform is described in detail. Lastly, the participant information, subjective experiment procedure, and data analysis methods are provided.

## 2.1 Study area

Two outdoor locations, A and B, in the vicinity of student hostels within the campus of Nanyang Technological University, Singapore, were chosen for the in-situ experiment, as shown in Figure 1. Both locations were flanked by a minor road (2 lanes) and a major expressway ( $2 \times 4$  lanes), with varying background traffic noise levels. Location A was a pedestrian walkway beside a busy road parallel to an expressway, whereas location B was an open area facing a small section of a minor road parallel to an expressway. Locations A and B were approximately 50 m and 30 m away from the expressway, respectively.

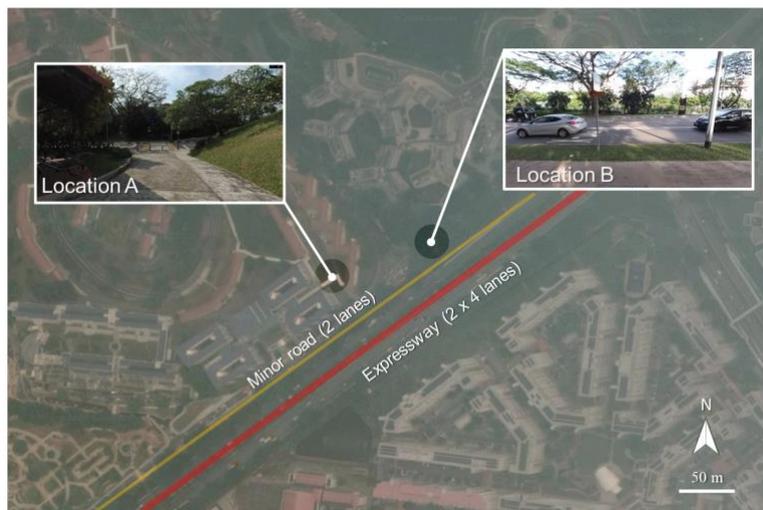


Figure 1 The selected locations for the in-situ experiment (source: Google Maps). The coordinates of Location A were 1.3447874N, 103.6878366E, whereas the coordinates of Location B were 1.3453807N, 103.6889984E.

These locations were decided based on pilot measurements that indicated a relatively stable ambient sound pressure level. This stability allows the traffic noise to be treated as the control across all participants.

## 2.2 Background traffic noise

Since road traffic is one of the major sources of noise pollution in urban environments [29,39,40], traffic noise was designated as the target noise to be augmented by natural sounds. To examine the in-situ background noise experienced by each participant in locations A and B, decibel-based indicators and psychoacoustic parameters were calculated from 132 (i.e. 4 recordings each for 33 participants) 3-min binaural recordings for each location. Five decibel-based parameters, namely the 3-min A-weighted equivalent SPL ( $L_{Aeq, 3\text{-min}}$ ), three percentage exceedance levels (i.e.,  $L_{A10}$  [10%],  $L_{A50}$  [50%], and  $L_{A90}$  [90%]), and the difference between  $L_{A10}$  and  $L_{A90}$  (denoted as  $L_{A10}-L_{A90}$ ), were calculated in accordance with ISO 1996-1 [41]. Similarly, percentage exceedance values of psychoacoustic parameters *loudness* (N) and *sharpness* (S) were calculated according to ISO 532-1 [42] and DIN 45692 [43], respectively. We denote by  $N_{10}-N_{90}$  the difference between  $N_{10}$  and  $N_{90}$ , and by  $S_{10}-S_{90}$  the difference between  $S_{10}$  and  $S_{90}$ . All parameter calculations were performed via a commercial software package (ArtemiS Suite, HEAD acoustics GmbH, Germany), and are summarized in Table 1. Statistical analyses (i.e. minimum, maximum, mean, and standard deviation) were performed on the arithmetic means of both left and right channels

of recorded binaural signals for all acoustic parameters.

Across both locations, the  $L_{Aeq, 3-min}$  obtained from the 132 recordings ranged from 63.6 to 78.7 dB. Notably, the mean  $L_{Aeq, 3-min}$  of location A (67.6 dB) was approximately 6.0 dB lower than that of location B (73.6 dB). Overall, the percentage exceedance levels at location A were relatively lower than those at location B as well. The standard deviations of the  $L_{Aeq, 3-min}$  were 1.8 dB and 1.6 dB at locations A and B, respectively, which indicated that there was low temporal variation in the traffic noise within each location. In addition, the mean  $L_{A10}-L_{A90}$  was 5.5 dB and 6.7 dB at locations A and B, respectively, which implied that fluctuations in sound levels during over the measurement period for the two locations were similar.

The psychoacoustic *loudness* exhibited similar trends with the decibel-based indicators across both locations. In terms of spectral characteristics, the *sharpness* values of the traffic noises at location B were slightly higher than those at location A. These differences in *sharpness* values might be attributed to the distances of the locations from the expressway. Since location B was closer to the expressway than location A, the traffic noises at location B contained slightly higher energies at high frequencies than those at location A. The measured acoustic indicators demonstrate that the acoustic environments in locations A and B showed a distinctive difference in terms of ambient traffic noise levels and the difference was consistent across the participants.

Table 1. Statistics of the calculated acoustic parameters of ambient traffic noises including noise indicators, *loudness* (N), and *sharpness* (S) values at the locations. Numbers in parentheses represent numbers of 3-min audio recordings.

Location		Noise indicators [dB]					<i>Loudness</i> [sone]					<i>Sharpness</i> [acum]				
		$L_{Aeq,3-min}$	$L_{A10}$	$L_{A50}$	$L_{A90}$	$L_{A10}-L_{A90}$	$N_{3-min}$	$N_{10}$	$N_{50}$	$N_{90}$	$N_{10}-N_{90}$	$S_{3-min}$	$S_{10}$	$S_{50}$	$S_{90}$	$S_{10}-S_{90}$
A (n=132)	Min	63.63	65.23	63.22	59.72	2.56	21.10	20.45	17.70	14.00	4.35	1.12	1.23	1.13	0.94	0.15
	Max	71.98	74.32	69.09	67.29	11.81	50.05	37.70	26.00	22.95	20.50	1.34	1.51	1.33	1.23	0.36
	Mean	67.59	69.49	66.36	64.04	5.45	30.77	27.45	21.85	18.49	8.96	1.21	1.35	1.20	1.10	0.24
	SD	1.77	2.04	1.36	1.50	1.63	5.51	3.60	1.83	1.71	2.94	0.04	0.05	0.04	0.04	0.04
B (n=132)	Min	69.57	71.13	69.11	66.36	1.86	31.55	30.50	26.10	21.30	6.20	1.20	1.29	1.18	1.08	0.10
	Max	78.71	80.98	77.36	75.04	11.53	64.05	50.95	41.70	37.45	25.25	1.52	1.69	1.51	1.38	0.33
	Mean	73.56	75.75	71.80	69.04	6.72	45.16	40.28	31.26	25.98	14.30	1.32	1.44	1.31	1.21	0.23
	SD	1.60	1.73	1.24	1.44	1.63	5.78	3.96	2.51	2.56	3.51	0.05	0.06	0.05	0.05	0.04
Combined (n=264)	Min	63.63	65.23	63.22	59.72	1.86	21.10	20.45	17.70	14.00	4.35	1.12	1.23	1.13	0.94	0.10
	Max	78.71	80.98	77.36	75.04	11.81	64.05	50.95	41.70	37.45	25.25	1.52	1.69	1.51	1.38	0.36
	Mean	70.58	72.62	69.08	66.54	6.08	37.97	33.87	26.56	22.24	11.63	1.27	1.39	1.25	1.16	0.24
	SD	3.43	3.66	3.02	2.90	1.75	9.15	7.46	5.20	4.34	4.19	0.07	0.07	0.07	0.07	0.04

### 2.3 Audio-visual stimuli

Two natural sounds, a birdsong (sparrow) and a water sound (stream), which were used in a previous study [17], were selected as the acoustic stimuli to augment the traffic noise. Both the birdsong and water sound were evaluated as the most pleasant among various natural sounds in a previous study [33]. As shown in Figure 2(a), the birdsong predominantly contains high-frequency components above 2.5 kHz, whereas Figure 2(b) shows that the water sound was dominated by mid to high frequency components from 500 Hz to 8 kHz. In terms of temporal characteristics, the water sound was constant with low temporal variability, whereas the birdsong was intermittent with high temporal variability.

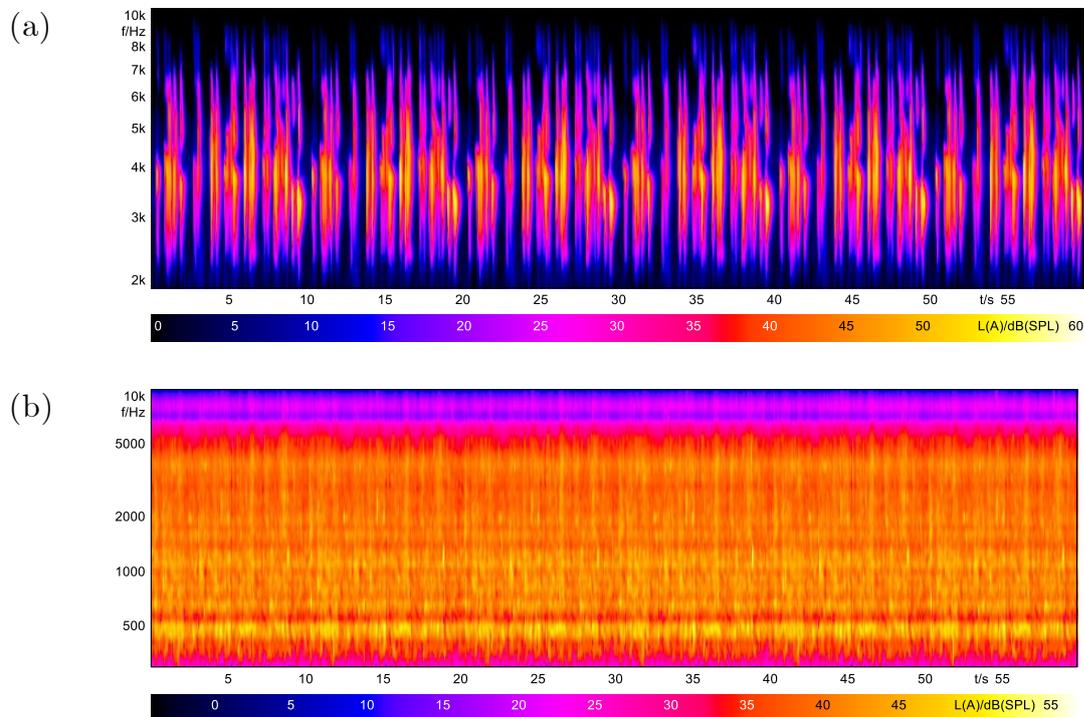


Figure 2. Spectrograms of the (a) birdsong and (b) water sound presented to all participants.

The natural sounds were presented to the participants in conjunction with realistic visual projections as holograms through a MR HMD (Hololens, Microsoft Corporation, USA), mimicking the presence of sound sources in the outdoor environment, as shown in Figure 3. Two types of visual projections were employed in this study, a hologram corresponding to a speaker and another corresponding to the natural sound source (i.e., a water fountain for the water sound and a sparrow for the birdsong) to examine the effects of visual sound sources on soundscape assessment.

The holographic audio-visual stimuli were spatialised to emulate a point sound source anchored on the ground 2 m in front ( $0^\circ$  azimuth) of the participant's evaluation position (Microsoft Spatializer Plugin, USA). At each location, a total of four combinations of natural sounds and holograms were presented to each participant in random order through the MR HMD, as depicted in Figure 3.

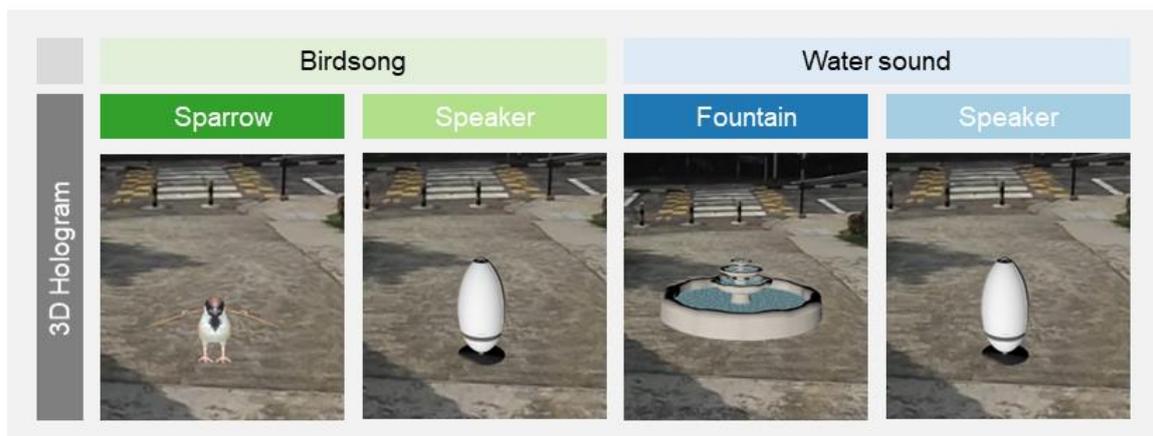


Figure 3. Holograms of a bird, a fountain, and a speaker, combined with their corresponding natural sounds, projected through an MR HMD in situ.

## 2.4 *Equipment*

All participants were required to put on a MR HMD (Microsoft HoloLens, USA), a calibrated binaural microphone (Brüel & Kjær TYPE 4101-B, Denmark), and a portable acoustic data acquisition system (SQobold, HEAD acoustics, Germany), as shown in Figure 4. The visual projection of the stimuli was displayed through the holographic lens of the MR HMD, while the sounds were presented spatially [23,36] via its downward-firing speakers such that the virtual hologram was seemingly integrated with the real environment. Head-tracked audio reproduction was applied in this study. The MR HMD rendered audio-visual elements in six degrees-of-freedom (6DoF) indicating that the audio-visual elements were anchored onto the real-world and the audio stream was rendered such that it was relative to the 6DoF translational and head movements of the participants. The sound source and visual objects were always anchored to the same in-situ physical locations. Specifically, the acoustic stimuli were spatialized as point sources 2 m away from the listener at 0 degrees azimuth.

The binaural microphones recorded the in-situ acoustic environment throughout the experiment at a sampling rate of 48 kHz, as required by ISO 12913-2 [44]. In addition, a Bluetooth keyboard (Logitech K380) was introduced as a reliable interface for participants to adjust the volume levels of the audio stimuli.

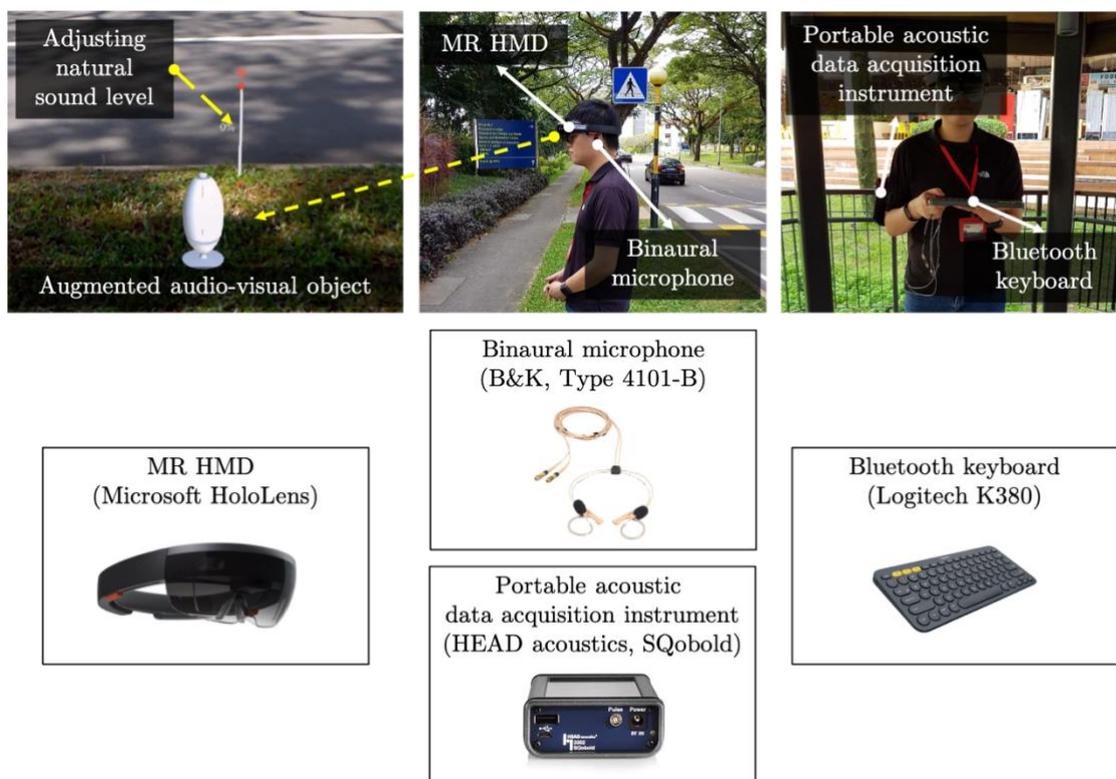


Figure 4. Equipment setup for the in-situ MR experiment

## 2.5 Subjective evaluation

Both the perceived loudness of noise (PLN) and overall soundscape quality (OSQ) were assessed for both the condition with only traffic noise and the condition where the traffic noise was augmented with a natural sound. The PLN is defined as the subjectively judged auditory loudness of the target traffic noise, and the OSQ is defined as the hedonic value of the sound in context [24]. In accordance with ISO 16832 [45], PLN was assessed using an 11-point rating scale (0: Not heard and 10: Extremely loud). Similar to previous studies [17,46], OSQ was also evaluated using an 11-point rating scale (0: Extremely bad and 10: Extremely good).

In terms of optimal natural sound levels, participants were asked to adjust the

playback level (via Bluetooth keyboard) of the natural sounds to their own preferred level with respect to the current in-situ ambient traffic noise levels. The sound levels of the natural sounds generated from the MR HMD were adjusted based on the normalised sound volume levels from 0 (silence) to 100 (loudest), where 100 represents a 1-min A-weighted equivalent sound level of approximately 80 dB for birdsong and water sounds. The scale bar (0 to 100) of the natural sound volume level was projected alongside the holograms as a visual reference, as shown in Figure 4.

## **2.6 Experimental design**

A two-way repeated-measure factorial design was applied to investigate the effects of audio-visual stimuli, locations, and their interaction on soundscape attributes (PLN and OSQ) and preferred natural sound-to-noise ratio (SNR). Audio-visual stimuli consisted of four different audio-visual combinations: Birdsong + bird hologram ( $B_{\text{bird}}$ ), Birdsong + speaker hologram ( $B_{\text{speaker}}$ ), Water sound + fountain hologram ( $W_{\text{fountain}}$ ), and Water sound + speaker hologram ( $W_{\text{speaker}}$ ). These four audio-visual stimuli combinations were presented to the participants in random order at both locations A and B.

For each of the four sound-hologram combinations (2 natural sounds  $\boxtimes$  2 visual stimuli) at each location, the participants were instructed to provide subjective assessments in three steps: In step 1, the participant was asked to listen to the ambient environment (mainly traffic noise) for 3 minutes while a binaural

recording of the environment was captured with a binaural microphone worn by the participant. The participant then evaluated the PLN and OSQ based on what they heard in the past 3 minutes. In step 2, a randomly chosen natural sound and hologram pair was projected through the MR HMD to the participant. The participant was then instructed to adjust the sound levels of the natural sound to their most preferred level while considering the OSQ and PLN. Lastly, in step 3, participants were asked to assess the OSQ and PLN of the scene with augmented audio-visual stimulus. These three steps were repeated for a total of 4 times (2 natural sounds  $\times$  2 visual stimuli) in a randomised order at each location. A summary of the three procedural steps is shown in Table 2.

Table 2. A summary of procedural steps of the in-situ experiment

Step	Stimuli (delivery mode)	Instruction to participant	Remarks
1	Traffic (ambient)	a. Listen to ambient environment for 3 min. b. Evaluate PLN and OSQ of past 3 min	
2	Traffic (ambient) Natural sound (MR HMD) Hologram (MR HMD)	Adjust sound levels of natural sound stimuli to most preferred levels based on its effects on OSQ and PLN	Natural sound and hologram pair are presented at random
3	Traffic (ambient) Natural sound (MR HMD) Hologram (MR HMD)	Evaluate PLN and OSQ of the ambient environment augmented with the audio-visual stimulus	The audio stimulus is presented at the levels set in step 2

## 2.7 Participants

To determine the minimum required sample size for a 95% statistical power level, a priori statistical power analysis was performed for the within-subject two-way repeated-measures analysis of variance (RM ANOVA) using G\*Power

calculator v3.1. The result suggested that 22 participants were required to detect an effect size of  $f(U) = 0.5$ ,  $\alpha = 0.05$ , and  $(1-\beta) = 0.95$ . In total, 33 participants (17 male and 16 female) took part in the experiment, which exceeded the required minimum sample size. The participants' ages ranged from 20 to 41 yrs. (Mean = 26.5, SD = 5.0). Hearing tests for the participants were conducted using an audiometer (Interacoustics AD629, Denmark) prior to the commencement of the experiment to confirm that they had normal hearing.

## **2.8 Procedure**

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.

Through pilot tests in a listening room, it was found that the sound levels from the MR HMD downward-firing speakers at the binaural microphone could vary ( $\pm 3$  dB) among participants although all conditions were kept constant. It was determined that the measured sound levels were dependent on distances between the downward-firing speakers and the binaural microphone position. Thus, the distance between the speaker drivers and the opening of each corresponding ear canal was measured and kept constant for every participant to ensure that the sound levels of the natural sounds were measured reliably.

Before the actual in-situ experiment, a short training session was conducted to

familiarize the participants with the experimental procedure. During the binaural recording, participants were instructed to stand facing the holograms with minimal head movement for consistency of visual environment and to minimise interference. The in-situ experiment took approximately 30 min at each location, including a 15 min break in between locations to relieve boredom and fatigue [47].

After completing the in-situ soundscape assessment, each participant was brought to a quiet listening room with an A-weighted background noise level of 28 dB. The participants were refitted with the test equipment, whilst ensuring the same distance between the downward-firing speakers of the MR HMD and the opening of the ear canal as measured in situ. Then, all the 3-min preferred natural sound levels that the participant had chosen during the in-situ assessment were recorded to calculate the sound levels of the preferred natural sound in the absence of the ambient traffic noise. The natural sound-to-noise ratio (SNR) for each test case defined by

$$SNR = L_{Aeq, 3\text{-min, Natural}} - L_{Aeq, 3\text{-min, Traffic}} \quad (1)$$

where  $L_{Aeq, 3\text{-min, Traffic}}$  refers to the 3-min A-weighted equivalent SPL during step 1, and  $L_{Aeq, 3\text{-min, Natural}}$  denotes the 3-min A-weighted equivalent SPL of natural sound captured in the quiet listening room.

All in-situ experiments were conducted between 10:00 and 12:00 with clear weather from May 2019 to March 2020. The mean temperature and relative humidity across all days of the in-situ experiments conducted were 31.6 °C (SD = 1.3) and 78.4% (SD = 3.8), respectively.

## **2.9 Data analysis**

A two-way RM ANOVA was conducted to investigate the effects of locations and audio-visual stimuli, and their interaction on PLN, OSQ, and participants' preferred SNR. Normality assumptions of the residuals of dependent variables (i.e., PLN, OSQ, and participants' preferred SNR) for each level of independent variables were tested with Shapiro-Wilk's test. The results showed that some datasets violated the normality assumption. However, it is known that ANOVA can yield robust and valid results against violation of the normality assumption [48,49]. In RM ANOVA tests, the assumption of sphericity was examined using Mauchly's test of sphericity. If the assumption was violated, Greenhouse–Geisser correction was then applied to adjust the degrees of freedom of the F-distribution. Post-hoc comparisons were conducted using the least significance difference test. In addition, a multiple regression analysis was conducted to develop models to predict preferred SNRs of natural sounds in response to ambient traffic noise conditions. All statistical analyses were conducted using the statistical software package, IBM SPSS (version 25.0, IBM, USA).

## **3. RESULTS**

### **3.1 Effect of natural sounds on perceived loudness of noise**

Mean PLN scores of the stimuli are plotted in terms of the locations in Figure 5. To examine whether there were significant differences in mean PLN scores between the stimuli and locations, a two-way RM ANOVA was conducted. The

results showed that the main effects of location [ $F(1, 32) = 44.1, p < 0.001$ ] and stimuli [ $F(2.8, 89.1) = 8.4, p < 0.001$ ] were statistically significant on PLN, whereas there was no interaction between location and stimuli as shown in Table 3.

Table 3. Summary of RM ANOVA results for perceived loudness of noise (PLN)

Factors	df <sub>1</sub>	df <sub>2</sub>	<i>F</i>	<i>p</i>	$\eta_p^2$
Location	1.00	32.00	44.10	<0.001	0.58
Stimuli <sup>a</sup>	2.78	89.06	8.44	<0.001	0.21
Interaction	4.00	128.00	0.46	0.76	0.01

<sup>a</sup> Assumption of sphericity was violated and Greenhouse–Geisser correction was applied.

The mean PLN score at location B (mean = 5.5, SD = 1.5) was higher than that at location A (mean = 4.3, SD = 1.6) across the stimuli due to the ambient traffic noise level difference between the two locations. A post-hoc test revealed that both the birdsong and water sound could reduce the PLN as compared to the traffic noise alone case. However, there were no significant differences in PLN between adding birdsong and water sounds ( $p < 0.05$ ). Interestingly, no significant differences between the holograms of real sound source (i.e., bird/fountain) and speaker for both the birdsong and water sound were observed in reducing PLN.

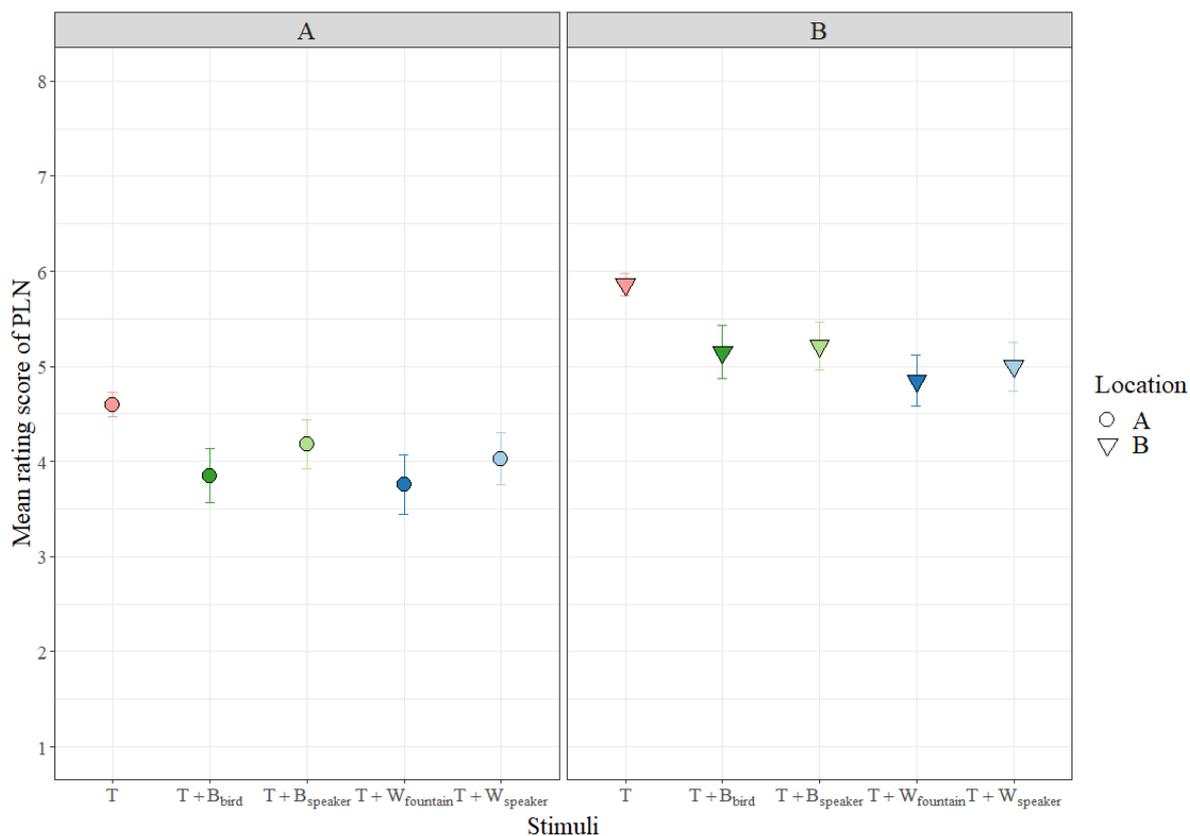


Figure 5. Mean PLN scores as a function of stimuli. The variables ‘T’, ‘B’, and ‘W’ designate the target traffic noise, birdsong, and water sounds, respectively; ‘+’ denotes the combination of stimuli. Error bars indicate 95% confidence intervals and the subscript indicates the hologram type.

### 3.2 Effect of natural sounds on overall soundscape quality

Mean OSQ scores for locations A and B were plotted as a function of stimuli in Figure 6. A two-way RM ANOVA was also performed to examine the effect of locations and stimuli on OSQ, as summarized in Table 4. The results showed that the locations [ $F(1, 32) = 55.8, p < 0.001$ ] and stimuli [ $F(2.8, 90.7) = 8.9, p < 0.001$ ] significantly affected OSQ scores. No interaction between locations and stimuli was found.

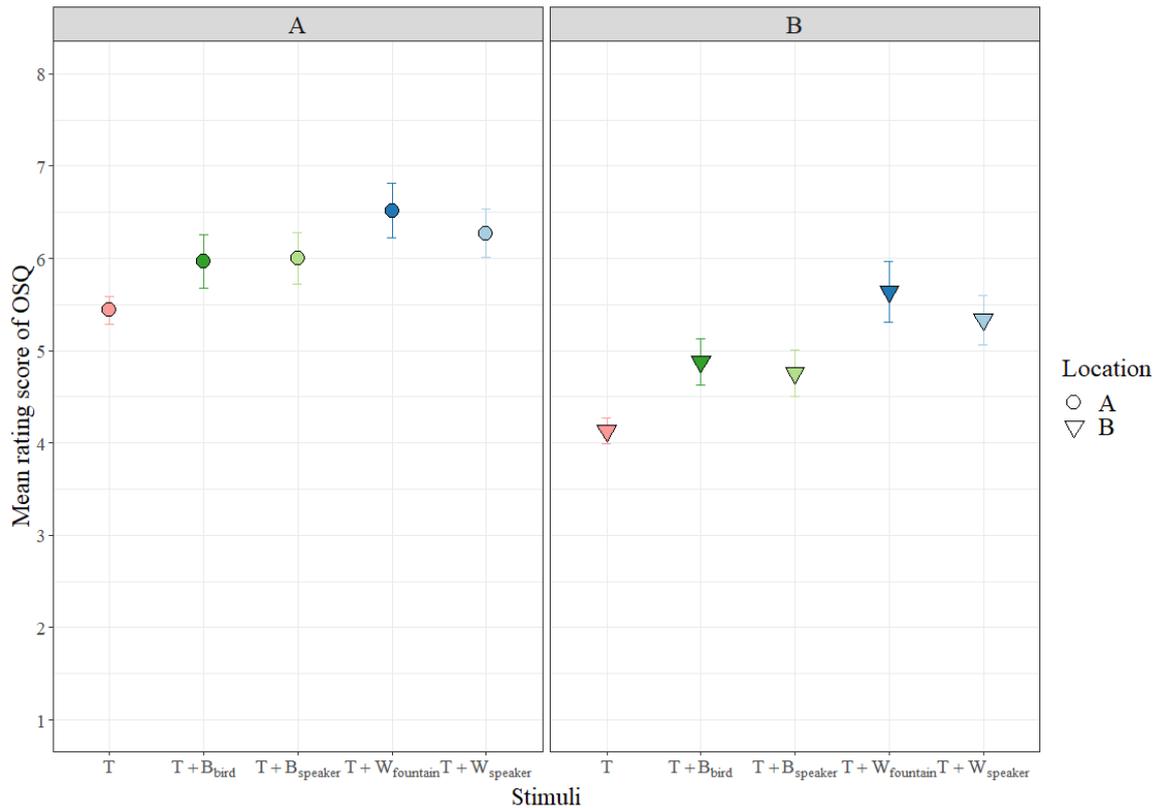


Figure 6. Mean OSQ scores as a function of stimuli. The variables ‘T’, ‘B’, and ‘W’ designate the target traffic noise, birdsong, and water sound, respectively; ‘+’ denotes the combination of stimuli. The subscript indicates the hologram type. The error bars indicate 95% confidence intervals.

As expected, the mean OSQ score at location A (mean = 5.8, SD = 1.7) with lower traffic noise levels was significantly higher than that at location B (mean = 4.6, SD = 1.7) because location A had lower traffic noise levels than location B. A post-hoc test was conducted to compare the mean OSQ among the stimuli. The results showed augmenting water sounds (T+W<sub>fountain</sub> and T+W<sub>speaker</sub>) and birdsong (T+B<sub>bird</sub> and T+B<sub>speaker</sub>) significantly increased the OSQ compared to the traffic-noise-alone condition ( $p < 0.05$ ). In contrast, no statistical mean differences were observed between the water sound and birdsong in terms of OSQ.

Table 4. Summary of RM ANOVA results for overall soundscape quality (OSQ)

Factors	df <sub>1</sub>	df <sub>2</sub>	<i>F</i>	<i>p</i>	$\eta_p^2$
Location	1.00	32.00	55.80	<0.001	0.64
Stimuli <sup>a</sup>	2.84	90.72	8.93	<0.001	0.22
Interaction	4.00	128.00	0.73	0.57	0.02

<sup>a</sup> Assumption of sphericity was violated and Greenhouse–Geisser correction was applied.

### 3.3 Preferred natural sound-to-noise ratios

Figure 7 presents the mean values of 3-min A-weighted equivalent SPLs of traffic and selected preferred natural sound levels at the locations. The mean preferred level for the birdsong was approximately 75 dB at both locations A and B. Meanwhile, the mean preferred level for the water sound was around 71 dB, which is relatively lower than that of the birdsong. Particularly, for the water sound, a lower mean sound level than the ambient traffic noise level was selected by the participants at location B. Similar to PLN and OSQ, there were no significant differences between the holograms of the speaker and the real sound source (i.e., bird/fountain, corresponding to the natural sound used).

The preferred SNRs of birdsong and water sound for the participants were calculated as defined as Eq. (1). Figure 8 shows the mean preferred SNRs of the stimuli at locations A and B. A two-way RM ANOVA showed that there were statistically significant differences in the mean preferred SNR in terms of the

locations [ $F(1, 32) = 109.26, p < 0.001$ ] and stimuli [ $F(1.8, 56.2.7) = 37.2, p < 0.001$ ], as presented in Table 5.

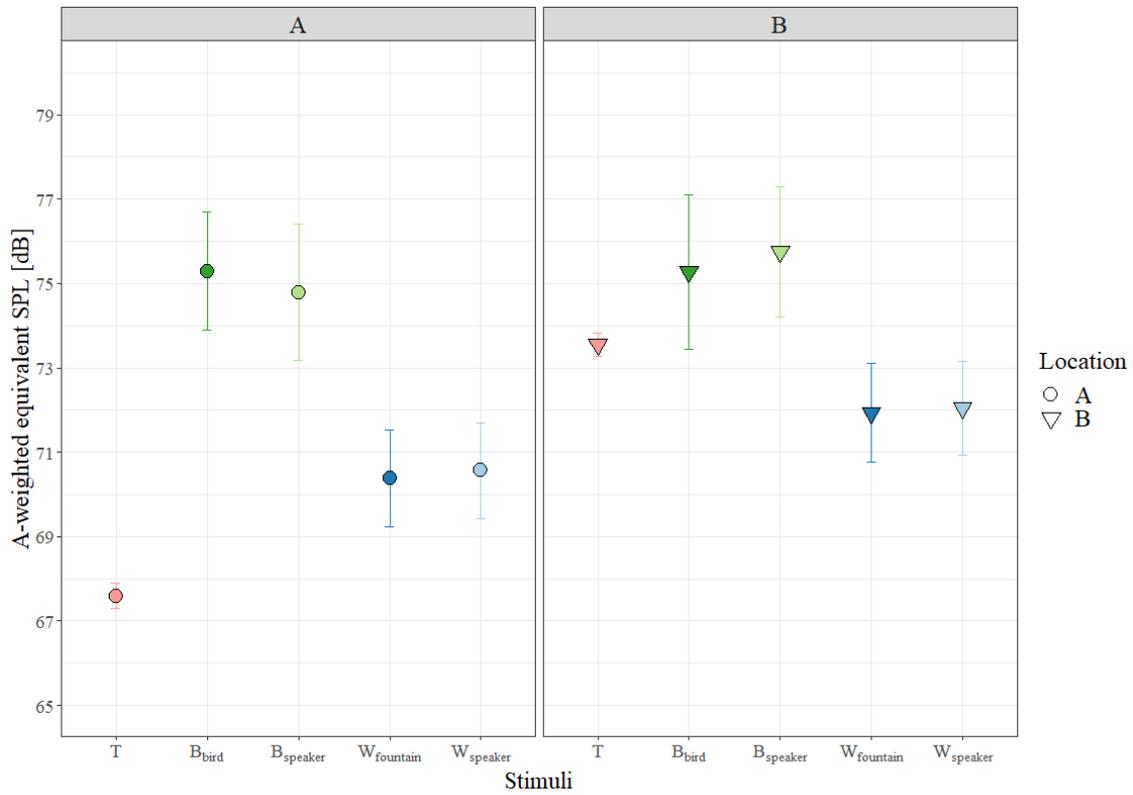


Figure 7. Mean  $L_{Aeq,3-min, Traffic}$  and  $L_{Aeq,3-min, Natural}$  of stimuli as a function of locations. The variables ‘T’, ‘B’, and ‘W’ designate the target traffic noise, birdsong, and water sound, respectively. The subscript indicates the hologram type. The error bars indicate 95% confidence intervals.

Table 5. Summary of RM ANOVA results for preferred natural sound-to-noise ratio (SNR) between natural sound and traffic noise levels

Factors	df <sub>1</sub>	df <sub>2</sub>	<i>F</i>	<i>p</i>	$\eta_p^2$
Location	1.00	32.00	109.26	< 0.001	0.77
Stimuli <sup>a</sup>	1.75	56.15	37.16	< 0.001	0.54
Interaction	3.00	96.00	0.43	0.73	0.01

<sup>a</sup> Assumption of sphericity was violated and Greenhouse–Geisser correction was applied.

Overall, the preferred SNRs at location A were higher than those at location B. This implies that as background traffic noise level increases, the desirable natural sound level tends to decrease. Interestingly, with regards to natural sound types, there was also a significant difference in choosing the preferred SNR between birdsong and water sound. Post-hoc test results showed that the preferred SNR of birdsong was significantly higher than that of water sound at both locations A and B ( $p < 0.05$ ). At location A (mean  $L_{Aeq,3-min, traffic}$  of 67.6 dB), the preferred SNRs for birdsong and water sound were 7.2 dB and 3.1 dB, respectively. At location B (mean  $L_{Aeq,3-min, traffic}$  of 73.6 dB), the participants chose a preferred SNR of 2.0 dB for birdsong, whereas the preferred SNR for water sound was found at -1.6 dB. This demonstrates that preferred SNR could differ depending on the types of natural sounds, as well as the ambient noise level.

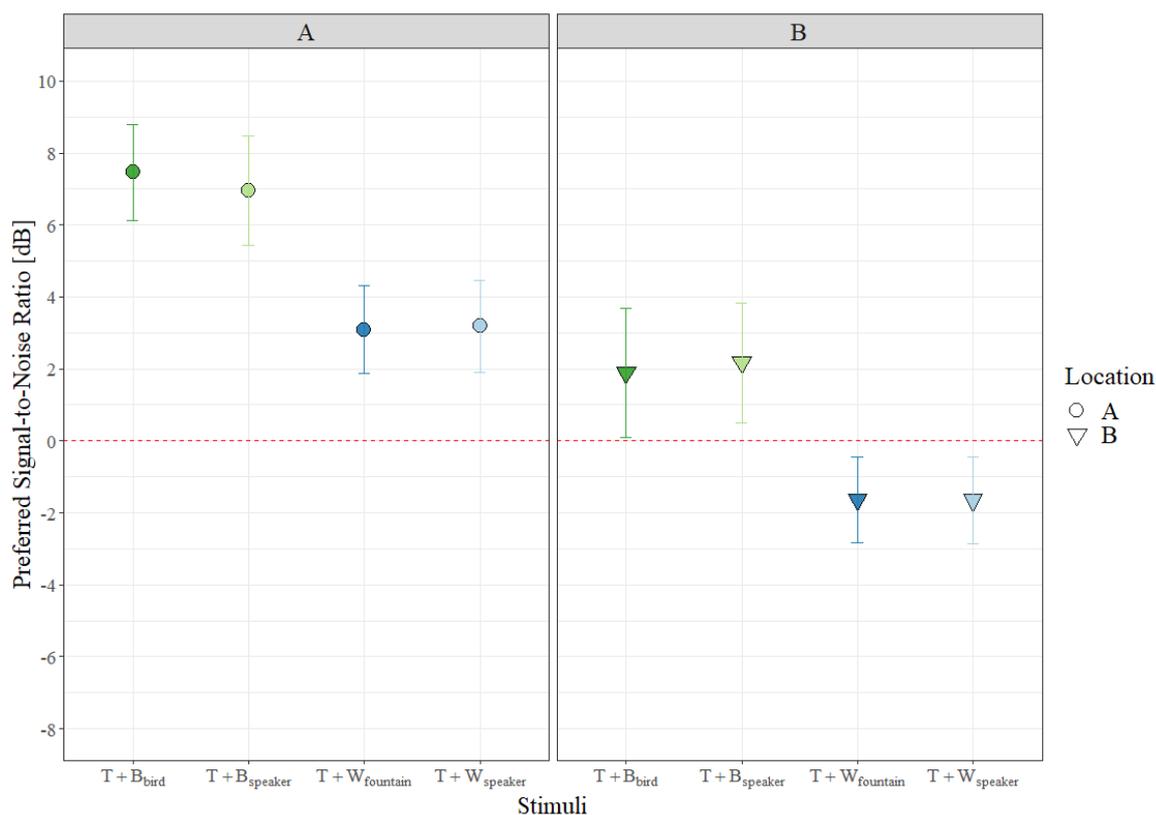


Figure 8. Mean preferred SNRs of each stimulus as a function of locations. The variables ‘B’ and ‘W’ designate the birdsong and water sound, respectively. The error bars indicate 95% confidence intervals, and the subscript indicates the hologram type.

### 3.4 *Acoustic indicators for predicting preferred natural sound-to-noise ratios*

To explore acoustic indicators to predict the preferred SNR for birdsong and water sound, Pearson’s correlation coefficients between the SNR and acoustic parameters of the ambient traffic noise (i.e., noise indicators (3-min A-weighted equivalent SPL and associated percentage exceedance thresholds), *loudness*, and *sharpness*) were calculated as shown in Table 6. Regarding the noise indicators, the SNR for both birdsong and water sound showed statistically significant negative correlations with the A-weighted equivalent SPL and percentile levels of

the traffic noise ( $p < 0.01$ ). This demonstrates that at higher background traffic noise levels, the participants preferred lower natural sound levels of birdsong and water sound and vice versa. This is also evidenced in Figure 9, which shows the individual participants' preferred SNRs for both the birdsong and water sound as a function of the measured  $L_{Aeq, 3\text{-min}}$  of traffic noise levels varying from 63.3 dB to 78.7 dB ( $\sim 15$  dB range) across the two locations.

These results also corroborate the findings of a previous study [17] that participants tended to prefer similar or higher natural sound levels when  $L_{Aeq}$  of traffic noise was 65 dB, whereas when  $L_{Aeq}$  of the traffic noise rose to 75 dB, natural sounds -6 dB lower than the traffic noise was evaluated as most desirable. However, overall the correlations for water sound were relatively greater than those for birdsong. Similarly, mean and percentage exceedance loudness values showed statistically significant negative correlations with the preferred SNR ( $p < 0.01$ ) although the correlation coefficients were slightly smaller than those of decibel-based indicators. The values of  $L_{A10}-L_{A90}$  and  $N_{10}-N_{90}$ , both of which represent temporal variations of the ambient traffic noise, also had negative and moderate correlations with the preferred SNR indicating that larger temporal variation of traffic noise may result in lower preferred SNRs. It was also found that *sharpness* values of the ambient traffic noises were negatively associated with the preferred SNR for both the birdsong and water sound. However, the strength of correlation was lower than those of the noise indicators and *loudness*.

Table 6. Pearson’s correlation coefficients between acoustic parameters of traffic noise and preferred SNR between natural sound and traffic noise levels ( $*p < 0.05$ ,  $**p < 0.01$ )

	$L_{Aeq,3-min}$	$L_{A10}$	$L_{A50}$	$L_{A90}$	$L_{A10}-L_{A90}$
SNR <sub>Bird</sub>	-0.50**	-0.50**	-0.49**	-0.47**	-0.21*
SNR <sub>Water</sub>	-0.65**	-0.66**	-0.60**	-0.58**	-0.47**
	$N_{3-min}$	$N_{10}$	$N_{50}$	$N_{90}$	$N_{10}-N_{90}$
SNR <sub>Bird</sub>	-0.47**	-0.51**	-0.49**	-0.46**	-0.41**
SNR <sub>Water</sub>	-0.65**	-0.65**	-0.61**	-0.59**	-0.57**
	$S_{3-min}$	$S_{10}$	$S_{50}$	$S_{90}$	$S_{10}-S_{90}$
SNR <sub>Bird</sub>	-0.30**	-0.22*	-0.31**	-0.30**	0.10
SNR <sub>Water</sub>	-0.44**	-0.45**	-0.43**	-0.41**	-0.12

Linear regression analyses were performed using acoustic parameters to develop prediction models for the preferred SNR for birdsong and water sound. Prior to the regression analyses, multicollinearity among all acoustic parameters were examined using the variance inflation factor (VIF). In general, VIF values greater than 10 indicate that there are multicollinearity problems. The test results showed that there were multicollinearity problems in the decibel-based indicators, *loudness*, and *sharpness* variables ( $VIF > 10$ ). Thus, based on the correlation analyses, *loudness* and *sharpness* were removed due to multicollinearity, while  $L_{Aeq, 3-min}$  of traffic noise, which had the highest correlation coefficients among the acoustic indicators, was selected as the best predictor in the regression models.

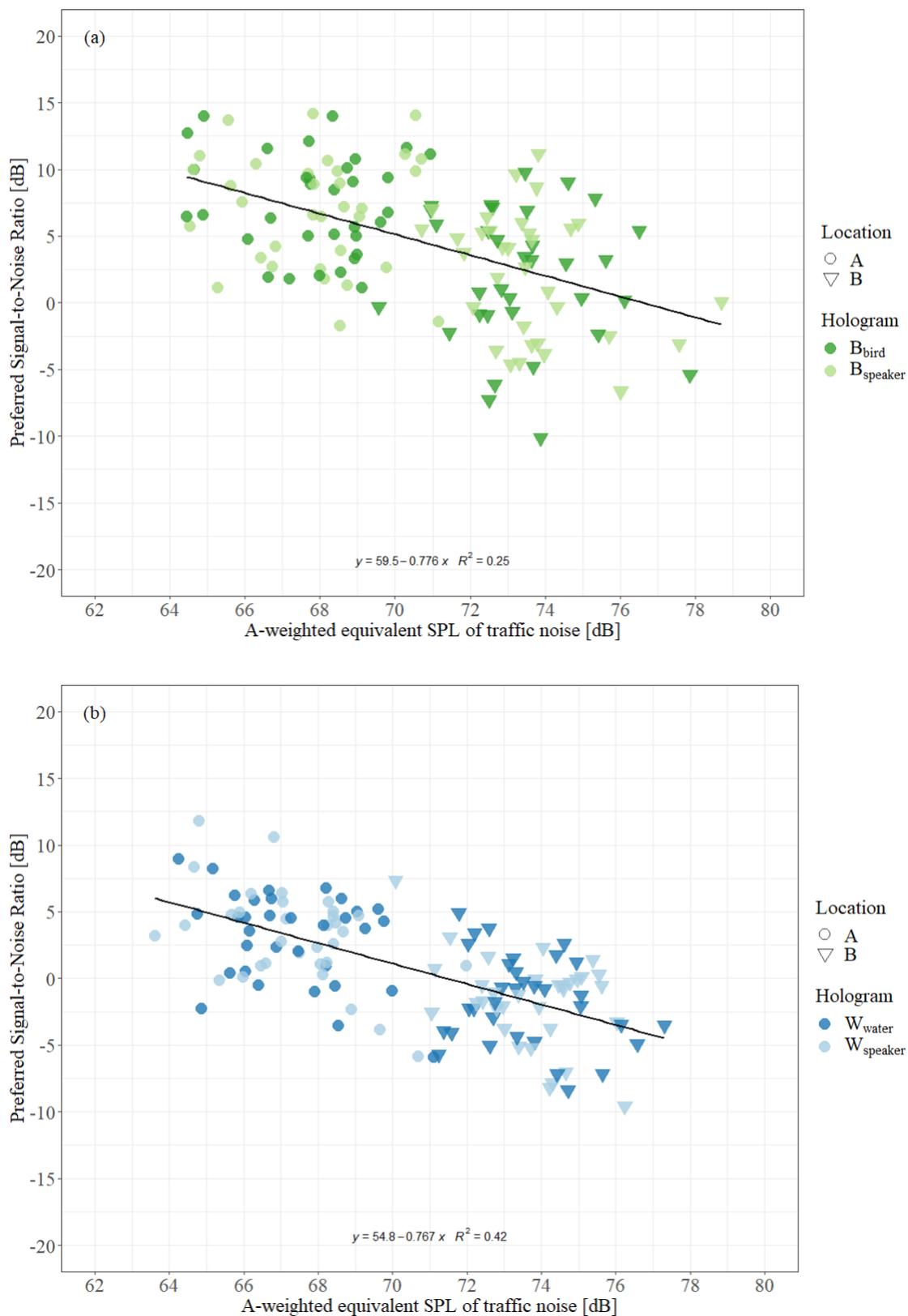


Figure 9. Preferred SNRs for stimuli as a function of 3-min A-weighted equivalent SPL of traffic noise for (a) Birdsong and (b) Water sound

The individual preferred SNRs for the birdsong ( $SNR_{bird}$ ) and water sound ( $SNR_{water}$ ) were plotted as a function of  $L_{Aeq, 3-min}$  of the ambient traffic noise in Figures 9(a) and 9(b), respectively. There were significant negative correlations between the preferred SNRs and A-weighted equivalent SPL of traffic noise for both birdsong and water sound.

The prediction models for the preferred  $SNR_{bird}$  [ $F(1,130) = 43.9, p < 0.001$ ] and  $SNR_{water}$  [ $F(1,130) = 95.2, p < 0.001$ ] were obtained from simple linear regression analyses as shown in Equations (2) and (3), respectively. Overall, the regression coefficients for birdsong and water sound were similar showing approximately -0.78. By comparing the constants, the preferred SNR of the birdsong was approximately 4.7 dB higher than that of the water sound.

$$SNR_{bird} = -0.78 L_{Aeq,3-min, traffic} + 59.5, R^2 = 0.25, p < 0.001 \quad (2)$$

$$SNR_{water} = -0.77 L_{Aeq,3-min, traffic} + 54.8, R^2 = 0.42, p < 0.001 \quad (3)$$

where,  $L_{Aeq, 3-min}$  refers to the 3-min A-weighted equivalent SPL of the ambient traffic noise.

The regression model for  $SNR_{water}$  exhibited better prediction accuracy than the model for  $SNR_{bird}$ ; the coefficients of determination ( $R^2$ ) for the birdsong and the water sound models were 0.25 and 0.42, respectively. This could be attributed to the larger variance in  $SNR_{bird}$  than that in  $SNR_{water}$  as shown in Figure 9. This indicates that the preferred SNRs for the birdsong more largely varied across the

individual participants, whereas the preferred SNRs for the water sound were relatively similar and constant across the participants.

## **4. DISCUSSION**

### **4.1 *Effect of augmenting natural sounds on judged soundscape attributes in laboratory and in situ***

The results of this study reaffirm that adding birdsong and water sounds could significantly decrease the PLN of traffic noise in real-world conditions. It was also found that there was no statistically significant difference in the reduction effect of PLN between the birdsong and water sounds. The result is in line with the findings of a previous study [17] that birdsong, which is unable to energetically mask traffic noise due to its predominantly high-frequency content and an intermittent temporal structure, can still reduce the perceived loudness of traffic noise. This supports the finding that the saliency of natural sounds plays a key role in soundscape assessment [50]. Regarding the effect of natural sounds on soundscape quality, both birdsong and water sound could significantly enhance OSQ compared to traffic noise-alone conditions. In addition, there were no significant differences in OSQ between the birdsong and water sound when they were used to augment the traffic noise. These results are also in line with the findings of previous studies [17]. Although the present study does not aim to quantify the effects of natural sound on PLN and OSQ with equivalent target noise levels, a previous study [46], conducted in a laboratory using the same

acoustic stimuli, revealed that the reduction in PLN and enhancement in OSQ were estimated as equal to approximately 4 dB and 11 dB reductions of the target traffic noise, respectively.

In terms of the preferred SNR, the results of this study are slightly different from previous laboratory experiment findings in [16]. To clearly show the relationship between the SNR and background noise levels in this study, the range of measured A-weighted equivalent SPL of traffic noise was broken into five intervals with a width of 3 dB, and then mean preferred SNR were calculated according to the five intervals. Figure 10 compares the mean preferred SNRs with the present study and that in [17] as a function of target traffic noise levels.

In [17], the auditory test in a laboratory condition was conducted using the same birdsong and water sounds adopted in this study to determine appropriate SNRs between natural sounds and target traffic noise at varying A-weighted SPLs. It was found that when the traffic noise level was 65 dB, the participants most preferred an SNR of 0 to 3 dB. Meanwhile, when the traffic noise was presented at 75 dB, an SNR of 6 dB was evaluated as the most appropriate.

On the contrary, the preferred SNRs obtained from the present in-situ test were approximately 5 to 6 dB higher than those from those in [17] as shown in Figure 10. The discrepancy in the preferred SNR could be attributed to different experimental conditions between in-situ and laboratory conditions. In [17], auditory experiments were conducted without visual stimuli under a controlled laboratory condition, whereas the present study was conducted in in-situ

environments, which provided real multisensory environmental perceptions to the participants. This implies that the participants might have been more attentive to the acoustic stimuli in the absence of visual stimuli in [17] than those in the present study because other environmental factors (e.g., vision, temperature, humidity, etc.) in-situ environments might have reduced the participants' attention on the acoustic stimuli. The results of the present study corroborate with the findings of Sudarsono et al. [51] that acoustic reproduction approximately 10 dB lower than the actual sound level measured in the in-situ experiments could produce similar subjective assessment results to those in-situ experiments.

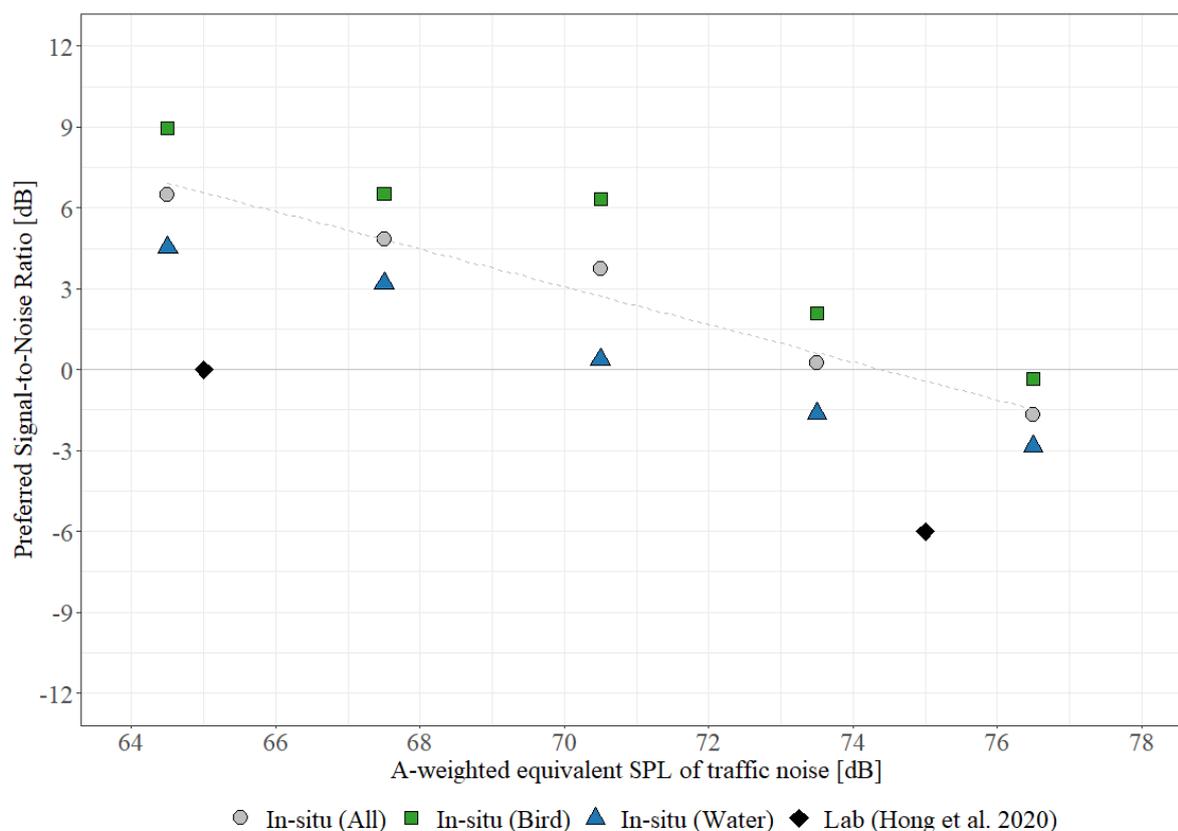


Figure 10. Comparison of mean preferred SNRs obtained from the present in-situ experiment and previous laboratory experiment (Ref [17]).

## **4.2 *Effect of visual objects of the sound source on soundscape and its implications***

Influences of visual objects on the effects of natural sound on PLN, OSQ, and preferred SNRs were examined in this study. The results show that there were no significant differences in the subjective responses between a representative real sound source (i.e. bird/fountain hologram corresponding to the natural sound) and a speaker hologram.

These findings are worth discussing in the context of previous studies suggesting that congruency between acoustic and visual components significantly affects soundscape assessment [52–54]. Most of the studies have revealed that audio-visual incongruency results in degradation of soundscape appraisal, while the perceived acoustic quality is improved when the audio-visual components are highly matched. In this context, it was expected that the perceived soundscape quality of natural sounds with a speaker hologram might be relatively lower than those with relatively real sound sources due to low audio-visual congruency between the natural sound and speaker hologram.

The findings of this study, however, showed that augmentation of natural sounds via a loudspeaker could yield the same effect on reducing the perceived loudness of traffic noise and enhancing soundscape quality. This result could be attributed to the effect of visibility of sound source on the soundscape. Hong et al. [33] found that the pleasantness and appropriateness of natural sounds are affected by the visibility of sound sources. Particularly, they showed that

perceptions of water sounds were significantly affected by the visibility of water features; when water features were not visible in the environments, the pleasantness and appropriateness decreased. In this study, although the natural sounds and speaker hologram are not highly matched in terms of audio-visual congruency, the participants recognized the visual source (i.e., speaker) where the natural sounds were generated from. The sound source visibility might lead to the same effects of natural sounds on soundscape even when combined with the speaker hologram. In other words, congruency may hinge more on the fact that the *source* of the sound is visible, rather than what the visual *representation* of the source is. This finding supports the installation of loudspeaker systems for introducing natural sounds as an effective soundscape design strategy to enhance soundscape in outdoor environments. Conversely, another possible explanation for the results might be related to the low quality of the holograms in terms of realism. Owing to the limitations of the current MR technology, the holograms (i.e, bird, fountain, and speaker) might not have been perceived as realistic, which might have led to no differences being observed in the soundscape attributes across the holograms.

### **4.3 *Limitations and future study***

Despite the findings of this study, some limitations need to be addressed. Although we controlled the audio-visual environments in the in-situ experiment, there were some uncontrolled and extraneous factors, such as the traffic flow and

meteorological conditions. Since participants took part in the experiment on different days, road traffic, and meteorological conditions slightly varied across the participants, which might affect the results. Nonetheless, our analysis regarding the SNRs is immune to this issue because the SNRs were taken with respect to the road traffic level at the point when each participant was doing the experiment.

Regarding the evaluation method, the participants were asked to select one single preferred level of natural sound at each experimental condition. However, there might be a range of preferred natural sound levels that results in the same soundscape quality. Thus, various subjective evaluation methods could be applied to validate the results of the findings in the future. It should be also noted that the relatively small variations in ambient traffic noise levels could be an inherent limitation of this study. Locations A and B were chosen because the background noise levels at each location were stable over time and across participants. However, the mean difference in ambient traffic noise levels between locations A (67.6 dB) and B (73.6 dB) was approximately 6.0 dB. While the experiments conducted in this study have shown that the specific type of ambient noise (traffic noise) can be augmented by the specific types of natural sounds (birdsong and water sound) to significantly reduce the PLN and improve the OSQ, it remains to be seen whether these conclusions are applicable to a greater variety of types and levels of ambient noise, as well as different natural sounds. Future studies could thus consider wider ranges of ambient noise levels and types of natural sound

sources to obtain a more generalized relationship between preferred natural sound levels and ambient noise levels.

Furthermore, the target traffic noise source and audio-visual stimuli were collocated at the same axial direction in space. However, a previous study has revealed that azimuth separation between a target traffic noise and a natural sound could reduce the effect of natural sound on reducing PLN and enhancing OSQ [46]. Therefore, in-situ studies could be carried out to validate the effects of spatial separation between target traffic noise and natural sounds on soundscape assessment in the future.

This study attempts to utilize the state-of-the-art MR displays to evaluate audio-visual augmentation of soundscapes in situ. However, it is worth noting that we have found several inherent technical limitations of the MR HMD while conducting the in-situ experiment. The Hololens provides a limited field of view (FOV) of 34 degrees, resulting in a windowing phenomenon whereby the holograms could only be seen when the participants shift the “window” over the position of the holograms. If the hologram is larger than the FOV, the hologram is clipped around the edges of the “window”. It was also noticed that in bright light, the holograms appeared faded. In addition, the downward-firing speakers used to present the natural sounds offered limited acoustic reproduction quality. Thus, further studies on a compatible headphone system with both natural listening with high acoustic fidelity could be explored in the future [55,56]. These current technological limitations should be considered when employing a MR HMD for in-

situ soundscape research.

Although the results of this study affirm that the soundscape design method by using pleasant natural sounds could enhance perceptions of noisy urban environments, a comparative cost-benefit analysis should be conducted in the future to ensure that the soundscape approach could result in cost-beneficial solutions for sustainable urban sound management. Lastly, a larger community scale study would be required to cross-validate the effects of natural sounds on PLN and OSQ for the perception of the general public or local residents in the future.

## **5. CONCLUSIONS**

In-situ experiments were carried out to examine the effects of natural sound types, background traffic noise levels, and the visual images of sound sources (represented as holograms) on soundscape attributes using a MR HMD. The results showed that introducing the natural sounds for this study significantly reduced the PLN of traffic noise and improved the OSQ in real-world settings, thus validating the existing body of evidence obtained from laboratory experiments. In terms of PLN and OSQ, no statistically significant difference was observed between birdsong and water sounds. Furthermore, there was no significant difference in the subjective evaluations of soundscape due to the holograms of representative real sound sources and the speaker. This suggests that the integration of loudspeaker systems into the urban environment is a viable and

effective soundscape design option.

The preferred SNRs between natural sounds and the traffic noises were explored in varying traffic noise levels in situ. A strong negative correlation was observed between the preferred SNR and ambient traffic noise levels for both birdsong and water sounds. When the traffic noise levels were below 70 dB, higher natural sound levels than the background noise level were considered as desirable (i.e.,  $\text{SNR} > 0$  dB). Meanwhile, at higher traffic noise levels over 70 dB, the participants tended to prefer lower natural sound levels than the ambient traffic noise levels (i.e.,  $\text{SNR} < 0$  dB). Overall, the preferred SNRs observed from the present study in situ were approximately 5 dB higher than those from the lab test that used the same natural sounds.

It was found that the A-weighted equivalent SPL of the traffic noise was the best predictor of this among the acoustic parameters examined. Regression models for predicting the preferred SNR of birdsong and water sounds corresponding to the ambient traffic noises were developed. The models could explain 25% and 42% of the variance in preferred SNRs for the birdsong and water sound, respectively. These models could be useful in the development of an adaptive natural sound generation system that interactively responds to ambient acoustic environments.

It is noted that the work was done with relatively limited, but key conditions of in-situ ambient traffic across a  $L_{Aeq, 3\text{-min}}$  range of 63.3 dB to 78.7 dB, and two representative natural sounds (water stream and birdsongs) which were previously evaluated in a lab study, were adopted. While the key concept of reduced

perceived loudness and improved soundscape quality, as well as the viability of augmentation via loudspeaker has been proven in real-world conditions, which are useful for urban soundscape design, further studies could be carried out with more ambient conditions and natural sound types. For instance, ambient noise from construction sites and along aircraft flight paths, as well as natural sounds from insects and rustling sounds.

### ***ACKNOWLEDGEMENT***

This research/project is supported by the National Research Foundation, Singapore, and Ministry of National Development, Singapore under its Cities of Tomorrow R&D Programme (CoT Award: COT-V4-2020-1). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the view of National Research Foundation, Singapore and Ministry of National Development, Singapore.

### ***DATA AVAILABILITY***

Replication data is available at <https://doi.org/10.21979/N9/KE0901>, an institutional open access research data repository for Nanyang Technological University (NTU) based on the open-source web application, Dataverse [57].

## ***6. REFERENCES***

- [1] B. Berglund, T. Lindvall, D.H. Schwela, Guidelines for community noise, 1999.

- [2] 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.
- [3] Eurostat, Sustainable development in the European Union - Monitoring report on progress towards the SDGs in an EU context, 4th ed., Publications Office of the European Union, Luxembourg, 2020. <https://doi.org/10.2785/555257>.
- [4] J. Kang, F. Aletta, T.T. Gjestland, L.A. Brown, D. Botteldooren, B. Schulte-Fortkamp, P. Lercher, I. Van Kamp, K. Genuit, A.E. Fiebig, J.L. Bento Coelho, L. Maffei, L. Lavia, Ten questions on the soundscapes of the built environment, *Build. Environ.* 108 (2016) 284–294. <https://doi.org/10.1016/j.buildenv.2016.08.011>.
- [5] J. KANG, From dBA to soundscape indices: Managing our sound environment, *Front. Eng. Manag.* 4 (2017) 184. <https://doi.org/10.15302/j-fem-2017026>.
- [6] A.L. Brown, A. Muhar, An approach to the acoustic design of outdoor space, *J. Environ. Plan. Manag.* 47 (2004) 827–842. <https://doi.org/10.1080/0964056042000284857>.
- [7] A. Lex Brown, A review of progress in soundscapes and an approach to soundscape planning, *Int. J. Acoust. Vib.* 17 (2012) 73–81. <https://doi.org/10.20855/ijav.2012.17.2302>.
- [8] F. Aletta, J. Xiao, What are the Current Priorities and Challenges for (Urban) Soundscape Research?, *Challenges*. 9 (2018) 16. <https://doi.org/10.3390/challe9010016>.
- [9] Ö. Axelsson, M.E. Nilsson, B. Hellström, P. Lundén, A field experiment on the impact of sounds from a jet-and-basin fountain on soundscape quality in an urban park, *Landsc. Urban Plan.* 123 (2014) 46–60. <https://doi.org/10.1016/j.landurbplan.2013.12.005>.
- [10] M. Rådsten Ekman, P. Lundén, M.E. Nilsson, Similarity and pleasantness assessments of water-fountain sounds recorded in urban public spaces, *J. Acoust. Soc. Am.* 138 (2015) 3043–3052. <https://doi.org/10.1121/1.4934956>.
- [11] J.Y. Jeon, P.J. Lee, J. You, J. Kang, Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds, *J. Acoust. Soc. Am.* 127 (2010) 1357–1366. <https://doi.org/10.1121/1.3298437>.
- [12] G.R. Watts, R.J. Pheasant, K. V. Horoshenkov, L. Ragonesi, Measurement and subjective assessment of water generated sounds, *Acta Acust. United with Acust.* 95 (2009) 1032–1039. <https://doi.org/10.3813/AAA.918235>.

- [13] J.Y. Jeon, P.J. Lee, J. You, J. Kang, Acoustical characteristics of water sounds for soundscape enhancement in urban open spaces, *J. Acoust. Soc. Am.* 131 (2012) 2101–2109. <https://doi.org/10.1121/1.3681938>.
- [14] J. You, P.J. Lee, J.Y. Jeon, Evaluating water sounds to improve the soundscape of urban areas affected by traffic noise, *Noise Control Eng. J.* 58 (2010) 477–483.
- [15] L. Galbrun, T.T. Ali, Acoustical and perceptual assessment of water sounds and their use over road traffic noise, *J. Acoust. Soc. Am.* 133 (2013) 227–237. <https://doi.org/10.1121/1.4770242>.
- [16] E. Ratcliffe, B. Gatersleben, P.T. Sowden, Associations with bird sounds: How do they relate to perceived restorative potential?, *J. Environ. Psychol.* 47 (2016) 136–144. <https://doi.org/10.1016/j.jenvp.2016.05.009>.
- [17] J.Y. Hong, Z.T. Ong, B. Lam, K. Ooi, W.S. Gan, J. Kang, J. Feng, S.T. Tan, Effects of adding natural sounds to urban noises on the perceived loudness of noise and soundscape quality, *Sci. Total Environ.* 711 (2020) 134571. <https://doi.org/10.1016/j.scitotenv.2019.134571>.
- [18] Y. Hao, J. Kang, J.D. Krijnders, Integrated effects of urban morphology on birdsong loudness and visibility of green areas, *Landsc. Urban Plan.* 137 (2015) 149–162. <https://doi.org/10.1016/j.landurbplan.2015.01.006>.
- [19] B. De Coensel, S. Vanwetswinkel, D. Botteldooren, Effects of natural sounds on the perception of road traffic noise., *J. Acoust. Soc. Am.* 129 (2011) EL148–L153. <https://doi.org/10.1121/1.3567073>.
- [20] J.Y. Hong, J.Y. Jeon, Designing sound and visual components for enhancement of urban soundscapes., *J. Acoust. Soc. Am.* 134 (2013) 2026–36. <https://doi.org/10.1121/1.4817924>.
- [21] J.Y. Jeon, P.J. Lee, J.Y. Hong, D. Cabrera, Non-auditory factors affecting urban soundscape evaluation., *J. Acoust. Soc. Am.* 130 (2011) 3761–3770. <https://doi.org/10.1121/1.3652902>.
- [22] L. Maffei, M. Masullo, A. Pascale, G. Ruggiero, V.P. Romero, Immersive virtual reality in community planning: Acoustic and visual congruence of simulated vs real world, *Sustain. Cities Soc.* 27 (2016) 338–345. <https://doi.org/10.1016/j.scs.2016.06.022>.
- [23] J. Hong, J. He, B. Lam, R. Gupta, W.-S. Gan, Spatial Audio for Soundscape Design: Recording and Reproduction, *Appl. Sci.* 7 (2017) 627. <https://doi.org/10.3390/app7060627>.

- [24] F. Aletta, J. Kang, Ö. Axelsson, Soundscape descriptors and a conceptual framework for developing predictive soundscape models, *Landsc. Urban Plan.* 149 (2016) 65–74. <https://doi.org/10.1016/j.landurbplan.2016.02.001>.
- [25] C. Guastavino, B.F.G. Katz, D.J. Polack, Jean-Dominique; Levitin, D. Dubois, Ecological validity of soundscape reproduction, *Acta Acust. United with Acust.* 91 (2005) 333–341.
- [26] J.Y. Hong, B. Lam, Z.T. Ong, K. Ooi, W.S. Gan, J. Kang, J. Feng, S.T. Tan, Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications, *Build. Environ.* 149 (2019) 1–14. <https://doi.org/10.1016/j.buildenv.2018.12.004>.
- [27] T. Van Renterghem, K. Vanhecke, K. Filipan, K. Sun, T. De Pessemier, Landscape and Urban Planning Interactive soundscape augmentation by natural sounds in a noise polluted, *Landsc. Urban Plan.* 194 (2020) 103705. <https://doi.org/10.1016/j.landurbplan.2019.103705>.
- [28] G. Cerwén, Urban soundscapes: a quasi-experiment in landscape architecture, *Landsc. Res.* 41 (2016) 481–494. <https://doi.org/10.1080/01426397.2015.1117062>.
- [29] J.Y. Hong, J.Y. Jeon, Relationship between spatiotemporal variability of soundscape and urban morphology in a multifunctional urban area: A case study in Seoul, Korea, *Build. Environ.* 126 (2017) 382–395. <https://doi.org/10.1016/j.buildenv.2017.10.021>.
- [30] D. Steele, E. Bild, C. Tarlao, C. Guastavino, Soundtracking the public space: outcomes of the musikiosk soundscape intervention, *Int. J. Environ. Res. Public Health.* 16 (2019) 1–38. <https://doi.org/10.3390/ijerph16101865>.
- [31] H. Li, S.K. Lau, A review of audio-visual interaction on soundscape assessment in urban built environments, *Appl. Acoust.* 166 (2020) 107372. <https://doi.org/10.1016/j.apacoust.2020.107372>.
- [32] H.I. Jo, J.Y. Jeon, Effect of the appropriateness of sound environment on urban soundscape assessment, *Build. Environ.* 179 (2020) 106975. <https://doi.org/10.1016/j.buildenv.2020.106975>.
- [33] J.Y. Hong, B. Lam, Z.T. Ong, R. Gupta, W.S. Gan, Suitability of natural sounds to enhance soundscape quality in urban residential areas, in: *24th Int. Congr. Sound Vib.*, London, UK, 2017: pp. 1–6.
- [34] L. Bjerre, S. Santurette, C. Jeong, Relationship between overall comfort and combined thermal and acoustic conditions in urban recreational spaces, in: *INTER-NOISE NOISE-CON Congr. Conf. Proc.*, Hong Kong SAR, China, 2017: pp. 3613–3620.

- [35] J.Y. Jeon, H.I. Jo, Effects of audio-visual interactions on soundscape and landscape perception and their influence on satisfaction with the urban environment, *Build. Environ.* 169 (2020) 106544. <https://doi.org/10.1016/j.buildenv.2019.106544>.
- [36] Consumer Technology Association (CTA)<sup>TM</sup>, CTA-2069 - Definitions and Characteristics of Augmented and Virtual Reality Technologies, CONSUMER TECHNOLOGY ASSOCIATION, Virginia, USA, 2018.
- [37] R. Gupta, B. Lam, J. Hong, Z. Ong, W. Gan, S.H. Chong, J. Feng, 3D audio AR/VR capture and reproduction setup for auralization of soundscapes, in: 24th Int. Congr. Sound Vib., London, UK, 2017.
- [38] B. Lam, J.Y. Hong, Z.T. Ong, K. Ooi, W.-S. Gan, S. Yeong, I. Lee, S.-T. Tan, Evaluation of preferred levels of natural sounds in situ environment through an augmented reality device, in: Proc. 23rd Int. Congr. Acoust., DEGA, Aachen, Germany, 2019: pp. 2503–2508.
- [39] T.J. Schultz, Synthesis of social surveys on noise annoyance., *J. Acoust. Soc. Am.* 64 (1978) 377–405. <https://doi.org/10.1121/1.382013>.
- [40] European Environmental Agency, Good practice guide on quiet areas, 2014. <https://doi.org/10.2800/12611>.
- [41] International Organization for Standardization, ISO 1996-1:2016 Acoustics — Description , measurement and assessment of environmental noise — Part 1 : Basic quantities and assessment procedures, International Organization for Standardization, Geneva, Switzerland, 2016.
- [42] International Organization for Standardization, Acoustics – Method for calculating loudness – Part 1: Zwicker method, Geneva, Switzerland, 2017.
- [43] DIN 45692, Measurement Technique for the simulation of the auditory sensation of sharpness, Beuth Verlag GmbH, 2009.
- [44] International Organization for Standardization, Acoustics — Soundscape — Part 2: Data collection and reporting requirements, Geneva, Switzerland, 2018.
- [45] ISO 16832, Acoustics — Loudness scaling by means of categories, Geneva, Switzerland, 2006.
- [46] J.Y. Hong, B. Lam, Z.T. Ong, K. Ooi, W.S. Gan, J. Kang, S. Yeong, I. Lee, S.T. Tan, The effects of spatial separations between water sound and traffic noise sources on soundscape assessment, *Build. Environ.* 167 (2020) 106423. <https://doi.org/10.1016/j.buildenv.2019.106423>.
- [47] R. Schatz, S. Egger, K. Masuch, The impact of test duration on user fatigue and reliability of subjective quality ratings, *J. Audio Eng. Soc.* 60 (2012) 63–73.

- [48] E. Schmider, M. Ziegler, E. Danay, L. Beyer, M. Bühner, Is it really robust?: reinvestigating the robustness of ANOVA against violations of the normal distribution assumption, *Methodology*. 6 (2010) 147–151. <https://doi.org/10.1027/1614-2241/a000016>.
- [49] M.J. Blanca, R. Alarcón, J. Arnau, R. Bono, R. Bendayan, Non-normal data: Is ANOVA still a valid option?, *Psicothema*. 29 (2017) 552–557. <https://doi.org/10.7334/psicothema2016.383>.
- [50] K. Filipan, B. De Coensel, P. Aumond, A. Can, C. Lavandier, D. Botteldooren, Auditory sensory saliency as a better predictor of change than sound amplitude in pleasantness assessment of reproduced urban soundscapes, *Build. Environ.* 148 (2019) 730–741. <https://doi.org/10.1016/j.buildenv.2018.10.054>.
- [51] A.S. Sudarsono, Y.W. Lam, W.J. Davies, The effect of sound level on perception of reproduced soundscapes, *Appl. Acoust.* 110 (2016) 53–60. <https://doi.org/10.1016/j.apacoust.2016.03.011>.
- [52] J.Y. Hong, J.Y. Jeon, Influence of urban contexts on soundscape perceptions: A structural equation modeling approach, *Landsc. Urban Plan.* 141 (2015) 78–87. <https://doi.org/10.1016/j.landurbplan.2015.05.004>.
- [53] J.L. Carles, I.L. Barrio, J.V. De Lucio, Sound influence on landscape values, *Landsc. Urban Plan.* 43 (1999) 191–200. [https://doi.org/10.1016/S0169-2046\(98\)00112-1](https://doi.org/10.1016/S0169-2046(98)00112-1).
- [54] S. Viollon, C. Lavandier, C. Drake, Influence of visual settings on sound ratings in an urban environment, *Appl. Acoust.* 63 (2002) 493–511.
- [55] K. Sunder, J. He, E.L. Tan, W.S. Gan, Natural sound rendering for headphones: Integration of signal processing techniques, *IEEE Signal Process. Mag.* 32 (2015) 100–113. <https://doi.org/10.1109/MSP.2014.2372062>.
- [56] R. Ranjan, W. Gan, Natural Listening over Headphones in Augmented Reality Using Adaptive Filtering Techniques, *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 23 (2015) 1988–2002. <https://doi.org/10.1109/TASLP.2015.2460459>.
- [57] J.Y. Hong, B. Lam, Z.T. Ong, K. Ooi, W.-S. Gan, J. Kang, S. Yeong, I. Lee, S.-T. Tan, Replication Data for: A mixed-reality approach to soundscape assessment of outdoor urban environments augmented with natural sounds, (2020). <https://doi.org/doi:10.21979/N9/KE0901>.
- [58] P. Milgram, H. Takemura, A. Utsumi, F. Kishino, Augmented reality: a class of displays on the reality-virtuality continuum, in: H. Das (Ed.), *Telem manipulator Telepresence Technol.*, SPIE, 1995: pp. 282–292. <https://doi.org/10.1117/12.197321>.

- [59] C. Flavián, S. Ibáñez-Sánchez, C. Orús, The impact of virtual, augmented and mixed reality technologies on the customer experience, *J. Bus. Res.* 100 (2019) 547–560. <https://doi.org/10.1016/j.jbusres.2018.10.050>.
- [60] P. Milgram, F. Kishino, Taxonomy of mixed reality visual displays, *IEICE Trans. Inf. Syst.* (1994).

**Appendix A**

Virtual, augmented, and mixed reality displays are often described on the reality-virtuality (RV) continuum [58], which is often misinterpreted, resulting in confusion between AR and MR [59]. Hence, the Consumer Technology Association (CTA) has published the CTA-2069 standard detailing the definitions and characteristics of VR, AR, and MR [36]. The CTA definitions are complementary to those in the RV continuum, as listed in Table A.1.

Table A.1 : Definitions of virtual, augmented, and mixed reality technologies

	Reality-Virtuality Continuum [58,60]	Consumer Technology Association (CTA) [36]
Virtual Reality (VR)	<ul style="list-style-type: none"> <li>Environment in which the participant-observer is totally immersed in a completely synthetic world, which may or may not mimic the properties of a real-world environment, either existing or fictional, but which may also exceed the bounds of physical reality by creating a world in which the physical laws governing gravity, time and material properties no longer hold.</li> </ul>	<ul style="list-style-type: none"> <li>Fully immersive user environment affecting or altering the sensory input(s) (e.g., sight, sound, touch, and smell)</li> <li>Allowing interaction with those sensory inputs by the user's engagement with the virtual world.</li> <li>Typically, but not exclusively, the interaction is via a head-mounted display, use of spatial or other audio, and/or hand controllers (with or without tactile input or feedback).</li> </ul>
Augmented Reality (AR)	<ul style="list-style-type: none"> <li>Any case in which an otherwise real environment is "augmented" by means of virtual (computer graphic) objects</li> </ul>	<ul style="list-style-type: none"> <li>Overlays digitally-created content into the user's real-world environment.</li> <li>AR experiences can range from informational text overlaid on objects or locations to interactive photorealistic virtual objects.</li> <li>AR differs from Mixed Reality in that AR objects (e.g., graphics, sounds) are superimposed on, and not integrated into, the user's environment.</li> </ul>
Mixed Reality (MR)	<ul style="list-style-type: none"> <li>Mixed Reality (MR) environment as one in which real world and virtual world objects are presented together within a single display, that is, anywhere</li> </ul>	<ul style="list-style-type: none"> <li>MR seamlessly blends a user's real-world environment with digitally-created content, where both environments coexist to create a hybrid experience.</li> <li>In MR, the virtual objects behave in all aspects as if they are present in the real</li> </ul>

---

between the extrema of the RV  
continuum

world e.g., they are occluded by physical objects, their lighting is consistent with the actual light sources in the environment, they sound as though they are in the same space as the user.

- As the user interacts with the real and virtual objects, the virtual objects will reflect the changes in the environment as would any real object in the same space.
-