Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications

Joo Young Hong [1], Bhan Lam [1], Zhen-Ting Ong, [1] Kenneth Ooi [1], Woon-Seng Gan [1], Jian Kang [2], Jing Feng [3] and Sze-Tiong Tan [3]

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

[2] UCL Institute for Environmental Design and Engineering, The Bartlett, University College London, Central House, 14 Upper Woburn Place, London WC1H 0NN, United Kingdom

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

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

50 Nanyang Ave, S2-B4a-03, Digital Signal Processing Lab,

School of Electrical & Electronic Engineering,

Nanyang Technological University

Singapore 639798, Phone: +65 98310715

ABSTRACT

With the advent of virtual reality (VR) technology, spatial audio has been increasingly adopted to evaluate the acoustic environment in soundscape research. It is therefore imperative to assess the quality of commonly used spatial audio reproduction methods to determine their ecological validity. Through subjective evaluations with 30 participants, the same participant evaluated four outdoor in situ locations vis-à-vis its corresponding audio-visual recording in VR on a separate day. A total of three spatial audio reproduction methods were assessed in VR, and they were all down-mixed from the first-order ambisonics (FOA) recordings to headphone-based FOA-static binaural, FOA-tracked binaural; and FOA 2-dimensional (2D) octagonal speaker array. The participants evaluated the acoustic environment in terms of the overall soundscape quality and perceived spatial qualities at each location. Regarding overall soundscape quality, there were no significant differences in evaluating the sound-source dominance and affective soundscape qualities between in situ and all VR methods. However, significant differences were found in the perceived spatial qualities between three reproduction methods and in situ. Among the source-related spatial attributes, the perceived distance of the dominating sounds was farther in the virtual than in the in situ evaluations. In the localization of sound sources, both the FOA-tracked binaural and the FOA-2D speaker array exhibited higher spatial acoustic fidelity than FOA-static binaural. Regarding the environment-related spatial quality attributes, the 2D speaker array reproduction was perceived as more immersive and realistic than other reproduction methods. Overall, the FOA-tracked binaural appears to exhibit sufficient fidelity for cinematic VR evaluation of soundscapes.

Keywords: Soundscape; Acoustic reproduction; Cinematic virtual reality; Binaural; Multichannel speaker; Ambisonics

1. Introduction

Throughout the soundscape design process, a heavy focus is placed on assessing human auditory perception over a physical acoustic environment [1,2]. Soundscapes can be assessed in situ and/or virtually (reproduced or synthesized) [3]. An in situ soundscape study can be conducted as an on-site survey, interview or soundwalk [4], in any combination, guaranteeing high ecological validity, because it reflects the real-world scenario. However, there are uncontrolled and extraneous factors, such as meteorological conditions or unwanted public interactions, that can adversely bias the results [5].

In contrast, soundscape evaluation in a reproduced or synthesized acoustic environment allows precise control of extraneous and independent variables. This allows researchers to study the cause-and-effect relationship between dependent and independent variables, but at the likely expense of ecological validity, owing to the artificiality of the controlled virtual acoustic environment. Hence, the quality of acoustic recordings and reproduction techniques plays a vital role in achieving sufficiently high ecological validity under laboratory conditions throughout the soundscape design process [2,6].

2. Background

2.1. Binaural audio in soundscape studies

Binaural and ambisonics are two common recording techniques in soundscape studies [4]. Binaural recording with a calibrated artificial head is the *de facto* choice for most soundscape evaluation studies, owing to its recognized international standardization. Calibrated condenser microphones in these binaural recording devices provide excellent timbre quality, an important factor in achieving realism and immersiveness. However, these binaural tracks are usually recorded and rendered in a static position (head-locked) over headphones with non-individualized head-related transfer functions (HRTF). This often results in front-back confusions and in-head localizations, which may affect the spatial impression of the reproduced sound field [7–11]. HRTFs characterize how the sound is heard at the eardrums and is thus unique to every individual. Whereas the importance of individualized HRTFs over head-tracking is still unclear [12], the inclusion of head-tracking to non-individualized HRTFs have been found to enhance externalization for frontal and rear sources [9,13,14]. However, there is currently no straightforward method to capture head-tracked binaural with calibrated artificial heads for head-tracked reproduction (i.e. in virtual reality (VR)).

The usability of head-locked/static binaural is thus technically restricted to a single viewing direction when paired with a dynamic visual media, reducing its immersiveness. If the visual media is omnidirectional and dynamic (i.e. in VR), the spatial perception of sound sources will be greatly affected owing to visual dominance [13,15,16].

2.2. Ambisonics in soundscape studies

Ambisonics, a method of recording and reproducing a sound field in full-sphere surround [17–19], is the leading recording technique for interactive spatial audio reproduction [2]. The ambisonic format provides flexibility in conversion to a multitude of audio formats that can be reproduced over headphones or multichannel speaker systems. Owing to consumer availability of ambisonic microphones and decoders, first-order ambisonic (FOA) formats are usually down-mixed to headphone-based head-tracked binaural, or 'FOA-tracked binaural', for clear distinction from traditional binaural rendering from artificial head recordings. FOA-tracked binaural is usually integrated with consumer VR head-mounted devices (HMD), enabling a 3-degrees-of-freedom (3DoF) audio-visual experience.

For increased spatial fidelity, ambisonic recordings can be reproduced using elaborate multichannel speaker systems in acoustically treated rooms. These treated rooms are usually quietened to reduce reverberations that could adversely affect the listening experience. Ambisonic speaker reproduction is a rather costly setup that also lacks the convenience of portability, important for quick iterations.

2.3. VR in soundscape studies

Because past studies have revealed that audio-visual interaction affects both soundscape and landscape perception [20–23], a multisensory design approach in architecture and urban planning is important for a holistic perception of the environment. In tandem with the consumerisation of VR technologies, there is a rapid adoption of VR in assessing existing or future designs of the urban environment, which consider different perceptual modalities in an integrative design process [24–28]. In addition to displaying landscapes cinematically [29], VR techniques can be used in instances, such as computer-generated 3-dimensional (3D) audio-visual environments that have yet to be built [24,30]. Thus, they allow for the evaluation of soundscape design processes in earlier design stages. This reiterates the significance of acoustic reproduction spatial quality for an accurate perceptual auralisation [31] of the 3D acoustic environment.

Recently, researchers have begun evaluating soundscapes in VR through ambisonic recordings down-mixed to head-tracked binaural [24] and multichannel speaker setups [27], shifting away from static artificial head binaural audio recordings [30]. To achieve high ecological validity in evaluating soundscapes of existing locations, cinematic VR systems that reproduce real-world audio-visual scenes through omnidirectional stereo videos and headphones presented on an HMD, have been increasingly adopted [26]. It is worth noting that past studies mainly focused on audio-visual qualities, such as pleasantness of soundscape or aesthetic landscape quality, and less on the spatial acoustic quality of different spatial audio reproduction methods, such as between headphone-based binaural and speaker arrays [2,26].

2.4. Choosing the appropriate reproduction media

Several emerging studies have investigated the differences of spatial sound quality ratings using different visual and audio reproduction techniques in VR [16,32–34]. However, the results are limited to the evaluation of static musical performers in an indoor music listening environment, and not the overall impression of an outdoor space, the primary focus of soundscape research.

As highlighted by Guastavino et al., audio recording techniques and reproduction media should be chosen with respect to perceptual experiences and the source material [35,36]. Because the primary focus of soundscape studies is on the outdoor environment, and there is an impending need to auralise soundscape interventions for evaluation [37], such as in investigating the spatial orientation of masking sounds [38], the suitability of widely available acoustic reproduction methods should be assessed based on these requirements.

Whereas an FOA down-mixed 2D hexagonal array of loudspeakers (i.e. FOA-2D) have been shown to be the preferred method for recreating outdoor environments, it only holds true when compared to 1-dimensional and 3D speaker arrays [35]. Moreover, past studies that compared binaural dummy head recordings to ambisonic speaker reproductions have displayed similar localization accuracy [39], and focused on user preference on musical tracks [40]. Despite preference of binaural artificial ear/head recordings (with [41] and without head-tracking [42]) over FOA-tracked binaural, it is still worthwhile to investigate FOA-tracked binaural because of recent improvements in ambisonic microphone quality and in the context of soundscape appraisals. It is thus timely to assess the perceptual differences between different FOA down-mixed media for VR HMDs in evaluating outdoor sound scenes: static binaural (i.e. FOA-static binaural), as a 'fairer' representation of the binaural artificial head using microphones with the same timbral qualities; FOA-tracked binaural and FOA-2D speaker array.

2.5. Research questions

This study seeks to answer the following questions. Is there a difference in the perceived overall soundscape quality across the three FOA reproduction methods? Does the perceived spatial quality of the dominant sound sources and the overall acoustic environment differ across the three FOA reproduction methods? What role can each method play in soundscape research, especially pertaining to soundscape design, if there are significant differences in the overall soundscape and the perceived spatial qualities across the acoustic reproduction methods? These questions will be answered by subjectively assessing the difference between in situ and virtually-reproduced audio-visual scenes in a VR HMD using headphone- (FOA-static and FOA-tracked binaural) and speaker-based (FOA-2D) spatial audio reproduction methods, which are described in detail in the next section. The outcomes of experiments pertaining to the first and second research questions will be analysed in Sections 3 and 4, respectively. Finally, the third research question will be discussed in Section 5, together with the limitations of this study.

3. Method

3.1. Participants

To determine the sample size needed for the within-subject design in this study, a priori-statistical power analysis was conducted with: an expected effect size $f = 0.40$, an $\alpha = 0.05$, a power $(1 - \beta) = 0.80$, and number of measurement = 4, using G*Power 3.1[43]. The expected effect size was determined based on the partial eta squared (η_p^2) of 0.14, considered a medium effect size [44]. A previous preliminary test showed η_p^2 values ranging from 0.1 to 0.3 [37]. The result of the power analysis suggested that a sample of 28 participants was needed.

In total, 30 participants (18 males and 12 females) were recruited for this experimental study. The age distribution of the participants ranged from 19 to 29 yrs. ($\mu_{\text{age}} = 22.7$, $\sigma_{age} = 2.6$). In previous studies, in situ soundscape evaluations with small groups of people were recommended to minimize the interaction effects among participants [45,46]. A large group of participants might cause potential detrimental effects on soundscape evaluation and acoustic recordings (e.g. human sounds generated by a large group, large masses of people altering the directivity of sounds, and attenuating sounds significantly). Therefore, the 30 participants were divided into three groups with, at most, 13 people in any group.

3.2. Materials

The questionnaire used in the soundscape evaluation aligns with the first two research questions and was thus separated into two parts for clarity. The first part pertained to overall soundscape quality, and the second part focused on the perceived spatial quality of the soundscape. The questionnaire in its entirety can be found in the appendix. The subjective results of parts one and two are presented in Sections 3 and 4, respectively.

3.2.1. Questionnaire part one: overall soundscape quality

The overall soundscape quality, with respect to the identification of perceived dominant sound sources and perceived affective quality of soundscapes, were assessed based on the Swedish Soundscape Quality Protocol [47].

Because sound sources play an important role in soundscape assessment, design, and implementation, the dominance of a pre-determined list of sound sources in each location was assessed on a 5-point scale (1: do not hear at all; 2: hear a little; 3: hear moderately; 4: hear a lot; 5: dominates completely). Similar to previous studies [47–49], the types of sound sources were classified into six categories: traffic noise, sounds from humans, water sounds, bird sounds, wind sounds and ventilation noises from HVAC systems. An open-ended 'other sounds' option was included to capture sounds of interest to the participant.

Of the numerous soundscape descriptors, the perceived affective quality model has been shown to provide the most comprehensive information of the soundscape [3]. Hence, the perceived affective quality model was employed here as a representative measure of overall soundscape quality. The perceived affective quality model utilizes two orthogonal descriptors in the form of the following four paired adjectives: 'Unpleasant – Pleasant', 'Uneventful – Eventful', 'Chaotic – Calm' and 'Boring – Lively'. These semantic differential attributes were rated on a 7-point bipolar scale. For the rest of this paper, the paired adjectives are represented with a single (right-hand) term for brevity (i.e. 'Pleasant' instead of 'Unpleasant – Pleasant').

3.2.2. Questionnaire part two: perceived spatial quality

To explore the ecological validity of the reproduced acoustic environments, the perceptual spatial quality of the reproduced acoustic environments was evaluated between the FOA reproduction methods and in situ. Because the scene-based approach to spatial quality evaluation [50] was based on the same concept of auditory scene analysis (identification of sound sources in a complex auditory scene), it was adopted in two parts: source-related (individual or ensemble) spatial attributes, followed by an assessment of the overall acoustic environment via environment-related spatial attributes and an overall quality metric.

The source-related spatial attributes were further classified into micro and macro attributes, where micro attributes describe the features of the individual elements (i.e. dominant sound source) and macro attributes describe a cognitive group of sources (i.e. combination of the identified sound sources). Specifically, the micro attributes were evaluated on a 7-point scale of direction (1: non-directional, 7: directional), width (1: narrow, 2: wide) and distance (1: near, 7: far) [35,51–53]. These source-related micro spatial attributes describe the 3D characteristics of the dominant sound sources in the acoustic environment, which are important factors for tweaking sound sources in the soundscape design process. To analyse multiple sound sources as an ensemble, a single source-related macro attribute was used to evaluate the perceived spatial quality. Participants were asked to rate how distinctly they could perceive the directions and locations of the sound sources around them at their present location using a 7-point scale condensed into a single attribute of distinctiveness (1: Instinct, 7: Distinct).

The environment-related spatial attributes describe the scene and are independent of source-related attributes. Environment-related attributes can be evaluated in terms of environmental dimensions (width and depth) and immersion attributes (presence and envelopment) [50].

Because envelopment usually describes reverberant indoor environments, the presence attribute proposed by Rumsey was employed to determine the immersiveness across the reproduction methods on a 7-point scale (1: not immersed at all, 7: fully immersed).

The dimensional qualities of the environment were evaluated by the extent to which the environment was perceived outside the head in an attempt to evaluate the environment width and to quantify in-head localisation, a known phenomenon that affects headphone-based spatial audio playback [10,54,55]. Hence, the participants were required to grade the externalization effect of the sound environment (1: inside head, 7: outside head).

Spatial quality can be also evaluated in the context of similarity to a reference (in situ) experienced previously by the participants. The attribute, realism, was defined as the degree of being realistic relative to the real-world scene. The participants were, thus, explicitly asked to grade the realism of the overall acoustic environment with respect to the in situ evaluation that they experienced previously (1: not realistic at all, 7: extremely realistic). Lastly, the reproduction fidelity was evaluated with a single quality measure of the overall listening experience (1: very bad, 7: very good). This single quality measure is analogous to the Basic Audio Quality metric commonly used to evaluate 3D sound playback media, and it captures both the timbral and spatial audio qualities [56].

3.3. Stimuli

3.3.1. Audio-visual recording

During the in situ soundscape evaluations, the audio-visual environment at all four locations was captured for the virtual soundscape evaluations. A spherical panoramic camera (Garmin VIRB 360 Action Camera, USA) was used to record a high-quality omnidirectional video of each location at 4K 30-FPS resolution with a bit-rate of 80 Mbps. Simultaneously, the acoustic environments at the locations were recorded using a low-noise ambisonic microphone (Sennheiser AMBEO VR 3D Microphone, Germany) via a recorder (Zoom F8 Multi-Track

Field Recorder, Japan), in A-format FOA [2]. Both the spherical camera and the ambisonic microphone were placed on a tripod at a height of 1.6 m from the ground. Additionally, a calibrated class 1 microphone (G.R.A.S. Type 40-PH CCP Microphone, Denmark) paired with a 24-bit analogue-to-digital converter (NI 9234) system recorded the A-weighed sound pressure levels ($L_{A_{eq}}$) at each location, which is the reference level for the playback of the sound recordings in laboratory conditions. The in situ soundscape evaluation and the audio-visual recording was time-synchronized with a clapper to ensure synchronicity of the in situ and virtual evaluations.

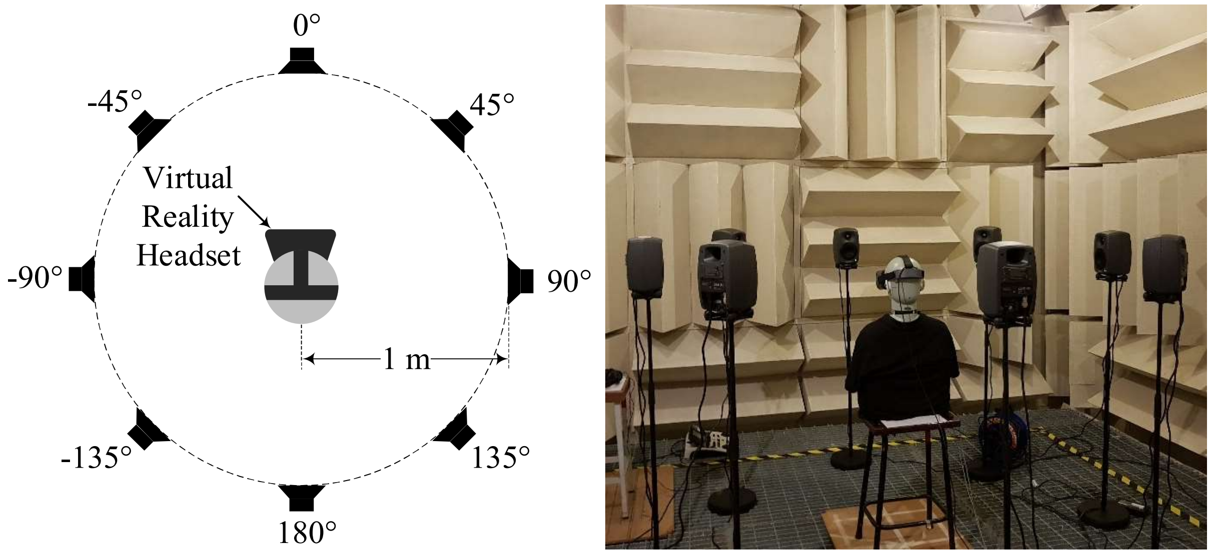
The recorded videos were post-processed (Adobe Premiere Pro CC 2017) into spherical projections for playback in a VR HMD, because acoustic evaluations through a VR HMD were found to be perceptually similar to the 180° CAVE-light system [16]. The recorded acoustic environment of the four locations was reproduced with three methods: FOA-static binaural; FOA-tracked binaural and an FOA-2D octagonal speaker array.

3.3.2. Reproduction methods

In the headphone-based reproduction methods, the A-format tracks were converted to B-format FOA, and were then down-mixed to create the binaural tracks using the KEMAR small pinnae HRTF.

Spatial audio based on head-tracked binaural sound was rendered with the Reaper (Reaper version 5.4, USA) Digital Audio Workstation (DAW) along with the Facebook Spatial Workstation plugin for Reaper [57]. The 3DoF head-tracking steers the direction of the binaural sound in synchronicity with the head movement for a more accurate perception of directional cues in the acoustic environment.

Lastly, the B-format FOA tracks were decoded to the FOA-2D octagonal array for an octagonal speaker setup using the Ambisonic Toolkit (ATK) plugin for the Reaper DAW [58].



(a) (b)

Fig. 1. The FOA-2D octagonal array loudspeaker array configuration: (a) top view of the loudspeaker configuration and (b) photograph of the system in an anechoic chamber at Nanyang Technological University.

All loudspeakers were placed 1-m away from the centre position of the octagon where the participant was seated, at a height of 1.3 m, as shown in Fig. 1.

3.3.3. Audio-visual synchronisation and calibration

The ambisonic audio and video recordings were synchronized with a clapper at the start of each recording session. Because the participants took approximately 1 min on average to complete the questionnaire at each location, 1-min excerpts of the audio-visual stimuli were used for the laboratory experiment.

The A-weighted equivalent continuous sound level ($L_{A_{eq},1-min}$) of the 1-min sound excerpt for each location for all three days were calibrated in an anechoic chamber by using a head and torso simulator (Brüel & Kjær 4128-C, Denmark) according to the measured $L_{A_{eq},1-min}$ in the in situ soundscape evaluations. The acoustic stimuli were equalized by inverse filtering with the headphone transfer function (HPTF) to neutralise changes to the frequency characteristics of stimuli owing to the headphones. The loudspeakers (Genelec 8320A Smart Active Monitor, Finland) were calibrated to a flat frequency response with the Genelec Loudspeaker Manager 2.0 (GLM) software at the sitting position (centre of the octagonal speaker array).

3.4. Experimental design

To compare and validate the soundscape reproduction methods, soundscape evaluations were conducted both in situ and in a virtual environment under laboratory conditions. A within-subjects design with repeated measures was employed. The independent variables were the types of acoustic environment, including the in situ environment and the three FOA reproduction methods [59]. The same participants took part in both the physical in situ soundscape evaluation and the corresponding virtualized version in the laboratory to minimize the individual differences during the comparison.

3.5. Procedure

3.5.1. In situ soundscape evaluation

Four locations in the campus of Nanyang Technological University, Singapore, were selected for the soundscape evaluation, with each location exhibiting different soundscape characteristics, as shown in Fig. 2. Location A is an open area with high human traffic flow near a canteen, a supermarket and a convenience store. Location B is a tranquil area beside a small lake surrounded by high-slopes and greenery. Location C is an open area with a fountain between a museum and a minor road with low traffic volume. Location D is a park exposed to heavy traffic noise, because it is flanked by a major expressway with the traffic lane mainly used by heavy vehicles nearest to the park. All four locations were within walking distance to minimize the potential psychological and physical distress of the participants during the in situ experiments, which could otherwise influence the soundscape assessment [45,46].

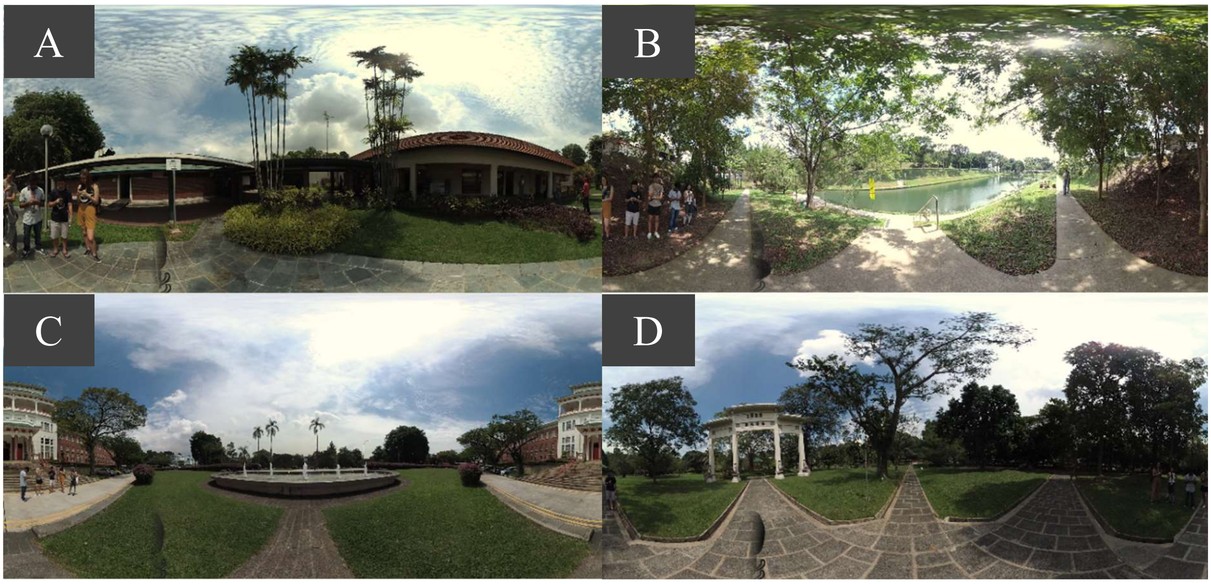


Fig. 2. Equirectangular panoramic stills from the spherical videos of selected evaluation locations for the experiments.

The in situ soundscape evaluations were conducted on three different days at the same time of day between 12:30 and 13:30 PM with similar weather conditions. The number of participants for three days was 5, 12 and 13, respectively. The mean and standard deviations (in parenthesis) of temperature and relative humidity across the three days were 30 °C (2.0 °C) and 81% (6.6%), respectively.

It is worth noting that, while the participants were evaluating the soundscape at location B on day 2, there was a momentary sun shower, whereas there was no rain on the other days. The presence of rain influenced the perception of dominant sources to an extent that will be discussed in the results section. The mean and standard deviations (in parenthesis) of the measured sound pressure levels at location A to D across three days were 67.9 (1.2) dB, 51.8 (1.1) dB, 70.4 (0.5) dB and 71.3 (0.6) dB, respectively. The standard deviations of the measured sound pressure levels (SPL) at each location across three days were all approximately 1 dB, indicating that sound environment at each location across the three days was stable and consistent.

All participants had normal hearing as evaluated with an audiometre (Interacoustics AD629, Denmark) before the experiment. In compliance with ethical procedures, the participants were provided with written information about this study, and written consent was obtained from all participants. Formal ethical approval to carry out this study was granted by the institutional review board of the Nanyang Technological University, Singapore (IRB-2017-07-025).

The participants were at least 2-m away from the audio-visual recording devices to avoid being located in the frontal view of the visual recordings. The participants could turn their heads, but they were instructed to stand still while facing the same direction as the frontal view of the audio-visual recordings to reduce the discrepancies and errors caused by different viewpoints of each participant between the in situ and virtual soundscape evaluations.

3.5.2. Virtual soundscape evaluation

After completing the in situ evaluations, the same participants evaluated the same audio-visual environments via the cinematic VR system in an anechoic chamber in the subsequent week. The cinematic VR experience is the realistic projection of the recorded audio-visual scene of the in situ soundscape evaluation and will henceforth be referred to as a ‘virtual soundscape evaluation’ for ease of reference. By having the same participants evaluate the VR scenes with a priori experiences, a fair comparison can be made between the in situ soundscapes and their virtualized versions.

The recorded visual scenes were presented through a consumer VR HMD (Pimax 4K VR, China) with the highest per-eye resolution at the time of this study, and the corresponding FOA down-mixed binaural tracks were played through headphones (Beyerdynamic Custom One Pro, Germany) connected to a sound card (Creative SoundBlaster E5, Singapore).

The decoded FOA-2D octagonal array tracks were presented through the octagonal speaker setup via a low-latency multichannel soundcard (MOTU UltraLite-mk4, USA). In total, three sets (3 days) of the 12 audio-visual stimuli (4 locations × 3 FOA reproductions) were generated for the virtual soundscape evaluations. Each participant took part in virtual soundscape evaluation with 12 audio-visual stimuli (4 locations × 3 FOA reproductions × 1 day) corresponding to the day of the in situ evaluation. The audio-visual stimuli were presented to the participants in a random order to eliminate memory bias from prior judgments. After experiencing each audio-visual stimulus, participants removed their VR HMD to complete the questionnaire. All the virtual soundscape evaluations lasted approximately 40 min.

3.6. Data analyses

Because the same participants took part in both the virtual and in situ soundscape evaluations, repeated measures (RM) analysis of variance (ANOVA) tests were conducted to investigate the within-subjects effects in the subjective responses between the three FOA reproductions and the in situ soundscape evaluations. A series of two-way RM ANOVAs were performed for the perceived dominance of sound sources (six types) and each source-related spatial attribute (i.e. direction, width, distance and distinctiveness), assuming independence between attributes.

When the set of attributes measured different aspects of a cohesive theme, such as environment-related spatial quality (e.g. immersion, realism, externalization and overall listening quality) [50], or when multiple attributes were theoretically correlated, such as the perceived affective quality of soundscape (e.g. pleasant, eventful, lively and calm) [3], RM multivariate ANOVA (MANOVA) was conducted to investigate the within-subject effects. Normality assumptions regarding the residuals of dependent variables for each level of independent variable were examined with Shapiro–Wilk's test. The results showed that the datasets violated the normality assumption. However, it is known that ANOVA yields robust and accurate p-values, even when the normality assumption is violated [60,61]. Thus, the RM ANOVAs in this study should be considered robust against the normality assumption.

Mauchly's test of sphericity was conducted to examine whether a dataset met the assumption of sphericity. If the assumption of sphericity was violated, then the Greenhouse–Geisser correction was applied to correct the degrees of freedom of the F-distribution. In all the RM ANOVA tests, post hoc comparisons were conducted with Bonferroni correction. Partial eta squared (η_p^2) values were reported as an effect size measure. Additionally, post hoc power ($1 - \beta$) was computed to assess whether a statistical test had a fair chance of rejecting an incorrect null hypothesis (H_0). All statistical analyses were performed using the statistical software package, SPSS (version 23.0, IBM, USA).

4. Results

4.1. Effect on the evaluation of overall soundscape quality across reproduction methods

Soundscape descriptors, subjective measures of how people perceive the acoustic environment, were used to determine the degree to which the different FOA reproduction methods (calibrated to the same in situ sound levels) affected the evaluation of soundscapes. As detailed in Section 3.2.1, the overall soundscape quality was evaluated based on the identified dominant sound sources and the perceived affective quality of each location.

4.1.1. Dominance of perceived sound sources

To identify discrepancies in the perceived dominant sound sources in each location, mean rating scores across the three FOA reproduction techniques were compared with those in situ, as shown in Fig. 3. The subjective responses of the dominance of sound sources were grouped into six sub-datasets corresponding to the six sound source types. Subsequently, the two-way RM ANOVA tests were conducted for each sound source type to examine the statistical significance in the perception of dominating sound sources amongst the locations and the soundscape evaluation sessions (in situ and all three FOA reproductions).

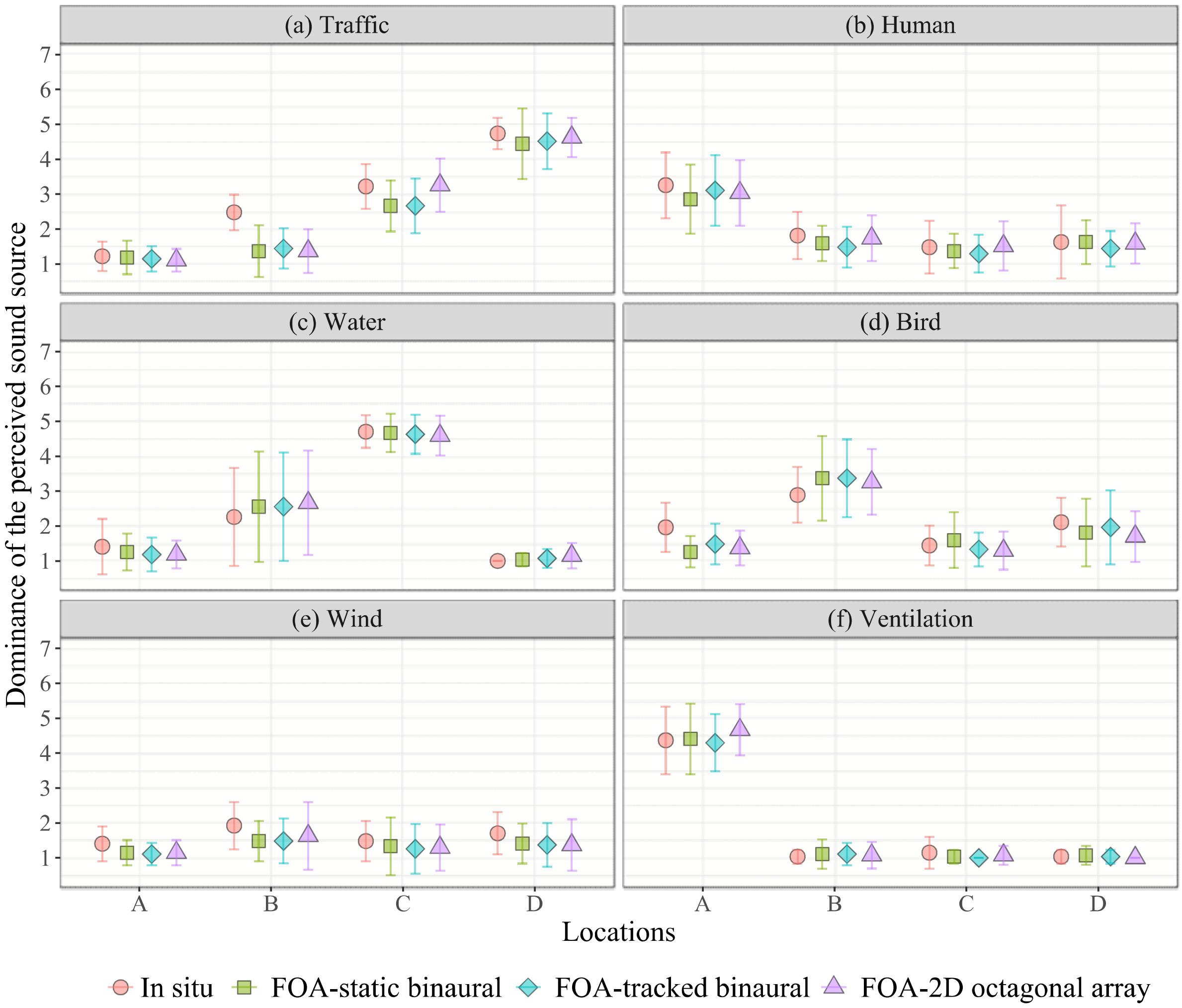


Fig. 3. Mean rating scores of the dominance of sound source types: (a) traffic, (b) human, (c) water, (d) bird, (e) wind and (f) ventilation, as a function of locations A to D, across the (1) in situ ( ), (2) FOA-static binaural ( ), (2) FOA-tracked binaural ( ) and (3) FOA-2D octagonal array ( ) conditions. The error bars indicate standard deviation.

As shown in Table 1, the two-way RM ANOVA results showed that the main effects of location on the dominance of perceived sound sources were significant across all six types of sound sources at a 0.05 significance level. This is because the dominant perceived sound sources significantly differed across the locations, as shown in Fig. 3. The post hoc tests were conducted to examine the mean differences in the dominance of perceived sound sources among the four locations across all six types of sound sources. The post hoc test results showed

Table 1. Summary of the RM ANOVA for the six types of sound sources with the reproduction methods and the locations.

Sound	Factors	df_1	df_2	F	p	η_p^2	$1-\beta$
Bird	Reproduction [g)]	2.38	61.77	1.85	0.16	0.07	0.41
	Location [g)]	2.19	56.86	54.36	< .001	0.68	1.00
	Interaction [g)]	5.68	147.73	4.55	< .001	0.15	0.98
Human	Reproduction [g)]	2.18	56.65	2.82	0.06	0.10	0.56
	Location [g)]	2.19	56.99	75.48	< .001	0.74	1.00
	Interaction [g)]	3.96	102.86	0.76	0.55	0.03	0.24
Traffic	Reproduction	3.00	78.00	17.78	< .001	0.41	1.00
	Location	3.00	78.00	410.95	< .001	0.94	1.00
	Interaction [g)]	5.36	139.33	8.07	< .001	0.24	1.00
Ventilation	Reproduction	3.00	78.00	1.42	0.24	0.05	0.36
	Location	1.23	32.00	408.70	< .001	0.94	1.00
	Interaction [g)]	3.49	90.85	1.98	0.11	0.07	0.53
Water	Reproduction	3.00	78.00	0.44	0.73	0.02	0.13
	Location [g)]	1.34	34.74	131.53	< .001	0.83	1.00
	Interaction [g)]	4.17	108.34	2.45	0.05	0.09	0.70
Wind	Reproduction [g)]	1.85	48.04	9.87	< .001	0.28	0.97
	Location [g)]	2.32	60.24	3.81	0.02	0.13	0.72
	Interaction [g)]	5.52	143.43	0.49	0.80	0.02	0.19

[g)] Assumption of sphericity was violated, and Greenhouse–Geisser correction was applied.

that the ventilation and human sounds were more dominantly perceived at location A than at the other locations ($p < 0.001$). Birdsongs were identified more prominently at location B than the other locations ($p < 0.001$). Water sounds, primarily from the fountain and most likely from the sun shower, were dominantly perceived at locations B and C, respectively ($p < 0.001$). Additionally, traffic noise at location D was the most dominant sound ($p < 0.001$) compared to the other locations.

Significant main effects of reproduction methods were only found in traffic [$F(3,78) = 17.78, \eta_p^2 = 0.41, p < 0.001, 1-\beta = 1.00$] and wind sounds [$F(1.85,48.04) = 9.87, \eta_p^2 = 0.28, p < 0.001, 1-\beta = 0.97$]. Hence, for traffic and wind sounds, post hoc multiple comparisons were conducted to examine the mean differences in the dominance amongst all the FOA reproductions and in situ. The post hoc tests revealed that both traffic ($p < 0.001$)

and wind sounds ($p < 0.05$) were more prominently perceived in situ than in all three FOA reproduction assessments.

There were significant interaction effects (reproduction method × location) for birdsongs [$F(5.68,102.86) = 4.55$, $\eta_p^2 = 0.15$, $p < 0.001$, $1 - \beta = 0.98$] and traffic noise [$F(5.36,139.33) = 8.07, \eta_p^2 = 0.24, p < 0.001, 1 - \beta = 1.00$]. Because the main effect is only meaningful when the interaction effect is not significant, the simple effects should be examined. Thus, the simple effects of the FOA reproduction methods and in situ, and locations for birdsongs and traffic noise were investigated, where RM ANOVA tests were performed for each within-subject variable (i.e. reproduction methods and locations). Because the family-wise error rate of 0.05 will be inflated approximately four times (four levels), the level of significance was set to 0.0125 (0.05/4) when finding the simple effects and for testing of the significance of the F value for each RM ANOVA.

The simple effects of the location for birdsongs and traffic sounds were significant across the FOA reproduction methods and in situ ($p < 0.001$), as shown in Table 2. Post hoc tests revealed that the birdsong was more significantly dominant at location B than at the other locations across the FOA reproductions and in situ ($p < 0.0125$). For traffic noise, there were significant differences in dominance amongst the locations ($p < 0.0125$), except between locations A and B across the FOA reproduction methods and in situ.

Table 2. Simple effects of locations for bird and traffic sounds

Sound	Reproduction	df_1	df_2	F	p	η_p^2	$1-\beta$
Bird	In-situ	3.00	78.00	21.58	< .001	0.45	1.00
	FOA-static binaural	3.00	78.00	31.61	< .001	0.55	1.00
	FOA-tracked binaural [g)]	2.34	60.82	31.78	< .001	0.55	1.00
	FOA-2D octagonal array	3.00	78.00	49.12	< .001	0.65	1.00
Traffic	In-situ	3.00	78.00	249.18	< .001	0.91	1.00
	FOA-static binaural	2.35	61.00	130.38	< .001	0.83	1.00
	FOA-tracked binaural	3.00	78.00	181.82	< .001	0.87	1.00
	FOA-2D octagonal array	3.00	78.00	265.86	< .001	0.91	1.00

[g)] Assumption of sphericity was violated and Greenhouse–Geisser correction was applied.

The simple effects test results for the FOA reproduction methods and in situ are summarized in Table 3. The mean rating scores for birdsong were significantly different amongst the FOA reproduction and in situ evaluations at locations A $[F(2.13,55.32) = 9.31, \eta_p^2 = 0.26, p < 0.001, 1 - \beta = 1.00]$ and B $[F(3,78) = 4.02, \eta_p^2 = 0.23, p = 0.01, 1 - \beta = 1.00]$. For traffic noise, statistically significant differences were found between the FOA reproduction and in situ evaluations at locations B $[F(3,78) = 36.7, \eta_p^2 = 0.59, p < 0.001, 1 - \beta = 1.00]$ and C $[F(3,78) = 8.64, \eta_p^2 = 0.25, p < 0.001, 1 - \beta = 1.00]$. The post hoc tests revealed that the mean rating scores of both birdsong and traffic noises in situ were significantly higher than those in the FOA reproductions.

There were three participants who selected the dominating sounds inconsistently across the FOA reproduction evaluations, and even in situ. Thus, subjective responses of these three participants were classified as outliers and were removed. Hence, a total of 27 subjective responses were used for statistical analyses in the sections that follow.

4.1.2. Perceived affective quality of soundscape

To investigate the perceived affective quality of soundscape between the three FOA reproductions and in situ evaluations, mean rating scores of the four attributes, (a) pleasant, (b)

Table 3. Simple effects of reproduction methods for bird and traffic sounds

Sound	Location	df_1	df_2	F	p	η_p^2	$1-\beta$
Bird	A [g)]	2.13	55.32	9.31	< .001	0.26	0.98
	B	3.00	78.00	4.02	0.01	0.13	0.82
	C [g)]	1.69	44.04	1.53	0.23	0.06	0.29
	D	3.00	78.00	1.85	0.14	0.07	0.46
Traffic	A [g)]	2.40	62.43	0.66	0.55	0.02	0.17
	B	3.00	78.00	36.71	< .001	0.59	1.00
	C	3.00	78.00	8.64	< .001	0.25	0.99
	D [g)]	1.86	48.47	1.18	0.31	0.04	0.24

[g)] Assumption of sphericity was violated and Greenhouse–Geisser correction was applied.

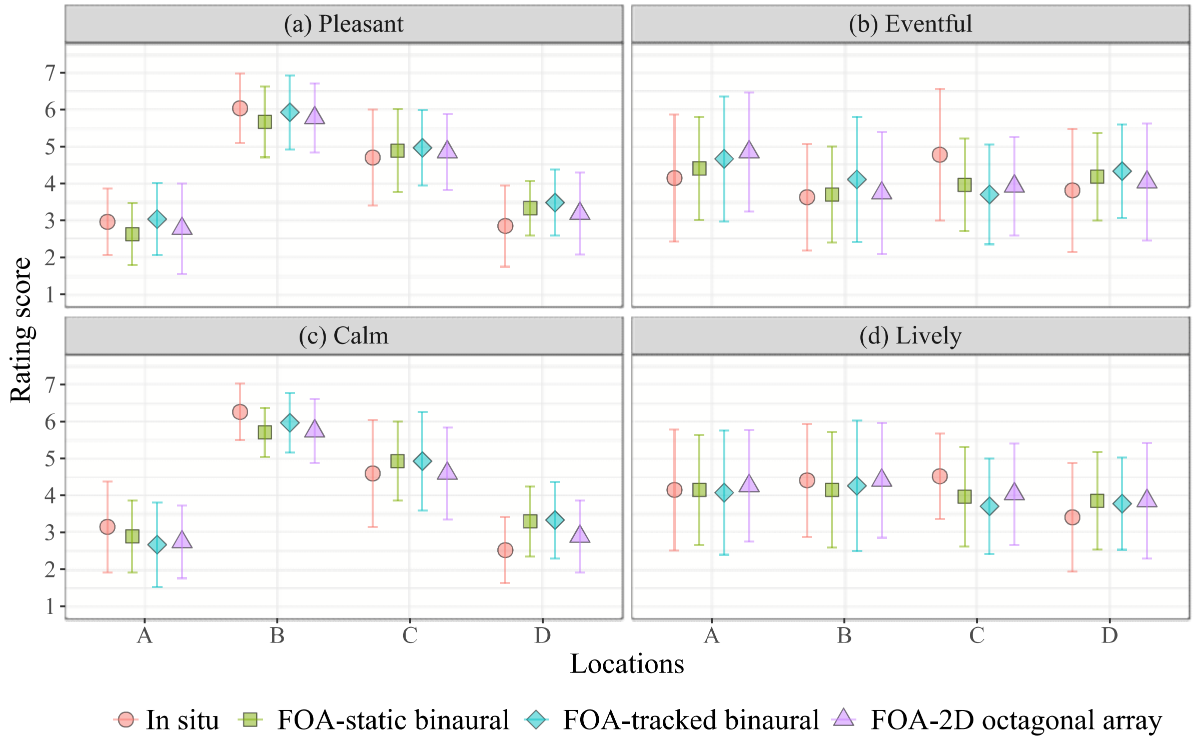


Fig. 4. Mean rating scores of the perceived affective quality attributes: (a) pleasant, (b) eventful, (c) calm and (d) lively, as a function of locations A to D, across the (1) in situ ( ), (2) FOA-static binaural ( ), (2) FOA-tracked binaural ( ) and (3) FOA-2D octagonal array ( ) conditions. The error bars indicate standard deviation.

eventful, (c) calm and (d) lively, are plotted in Fig.4 (a to d), respectively, as a function of the locations A to D.

As shown in Table 4, a two-way RM MANOVA test was conducted to investigate the effects of the reproduction method, the location and their interactions with attributes pertaining to the affective quality of soundscape. The main effects of the reproduction method in each of the four attributes were not significant. This implies that there were no significant differences in affective quality of soundscape between the in situ and virtual soundscape evaluations. Regarding the main effects of location, significant differences were found in 'Chaotic–Calm' and 'Unpleasant–Pleasant', whereas there were no significant differences in 'Uneventful–

Table 4 Summary of the RM MANOVA for the affective quality of soundscape with the reproduction methods and locations

Factors	Affective quality of soundscape	df_1	df_2	F	p	η_p^2	$1-\beta$
Reproduction	Calm	3.00	78.00	1.80	0.15	0.06	0.45
	Eventful	3.00	78.00	0.25	0.86	0.01	0.10
	Lively	3.00	78.00	0.79	0.50	0.03	0.21
	Pleasant	3.00	78.00	1.40	0.25	0.05	0.36
Location	Calm	2.13	55.47	93.11	< .001	0.78	1.00
	Eventful [g)]	1.94	50.43	2.14	0.13	0.08	0.41
	Lively [g)]	1.77	46.04	1.18	0.31	0.04	0.23
	Pleasant	3.00	78.00	117.25	< .001	0.82	1.00
Interaction	Calm	9.00	234.00	4.47	< .001	0.15	1.00
	Eventful [g)]	5.72	148.68	3.44	< .001	0.12	0.93
	Lively [g)]	5.54	143.91	1.53	0.18	0.06	0.55
	Pleasant [g)]	5.59	145.35	1.40	0.22	0.05	0.51

[g)] Assumption of sphericity was violated and Greenhouse–Geisser correction was applied.

Eventful' and 'Boring–Lively'. Post hoc tests for comparison of mean values across the four locations revealed that the mean rating score of 'Chaotic–Calm and 'Unpleasant–Pleasant' at locations A and D were significantly lower than those at locations B and C ($p < 0.001$). In other words, locations A and D were characterized as unpleasant and noisy, whereas soundscapes at locations B and C were described as pleasant and calm.

Because the effects of interaction (reproduction method × location) were significant in 'Chaotic–Calm' and 'Uneventful–Eventful', the simple effects of reproduction and location were computed. The significance level for the simple effect tests was set to 0.0125 (0.05/4). To test the simple effects of the reproduction method, separate one-way RM ANOVA tests in terms of 'Chaotic–Calm' and 'Uneventful–Eventful' were performed for each location as shown in Table 5. It was found that the mean rating score for 'Chaotic–Calm' at locations D only showed statistically significant differences between in situ and virtual soundscape evaluations [$F(3,78) = 8.56$, $\eta_p^2 = 0.25$, $p < 0.001$, $1-\beta = 0.99$]. The post hoc tests revealed that the participants perceived the soundscape at location D as louder in the in situ

Table 5. Simple effects of the reproduction methods for 'Calm' and 'Eventful'

	Location	df_1	df_2	F	p	η_p^2	$1-\beta$
Calm	A	3.00	78.00	2.04	0.11	0.07	0.50
	B	3.00	78.00	3.77	0.01	0.13	0.79
	C [g)]	2.37	61.72	1.54	0.22	0.06	0.34
	D	3.00	78.00	8.56	< .001	0.25	0.99
Eventful	A	3.00	78.00	3.03	0.03	0.10	0.69
	B	3.00	78.00	0.97	0.41	0.04	0.25
	C [g)]	2.24	58.32	3.80	0.02	0.13	0.70
	D	3.00	78.00	1.33	0.27	0.05	0.34

[g)] Assumption of sphericity was violated and Greenhouse–Geisser correction was applied.

environment (mean = 2.52, SD = 0.89) than those in FOA-static binaural (mean = 3.30, SD = 0.95) and FOA-tracked binaural (mean = 3.33, SD = 1.04) reproductions ($p < 0.001$), whereas there was no significant difference between in situ environment and FOA-2D octagonal array reproduction ($p = 0.40$).

Simple effects of the location were investigated by conducting separate one-way RM ANOVA tests in terms of 'Chaotic–Calm' and 'Uneventful–Eventful' for each reproduction method, as shown in Table 6. The effects of locations were significant for 'Chaotic–Calm' across the in situ and all virtual reproduction methods, whereas those for 'Uneventful–Eventful' were not significant at the significance level of 0.0125. Post hoc tests showed that the mean

Table 6. Simple effects of the locations for 'Calm' and 'Eventful'

	Reproduction	df_1	df_2	F	p	η_p^2	$1-\beta$
Calm	In-situ	3.00	78.00	65.41	< .001	0.72	1.00
	FOA-static binaural	3.00	78.00	60.51	< .001	0.70	1.00
	FOA-tracked binaural	3.00	78.00	55.86	< .001	0.68	1.00
	FOA-2D octagonal array	3.00	78.00	56.31	< .001	0.68	1.00
Eventful	In-situ	3.00	78.00	3.21	0.03	0.11	0.72
	FOA-static binaural	3.00	78.00	1.93	0.13	0.07	0.48
	FOA-tracked binaural [g)]	1.92	49.83	2.09	0.14	0.07	0.40
	FOA-2D octagonal array	3.00	78.00	3.17	0.03	0.11	0.71

[g)] Assumption of sphericity was violated and Greenhouse–Geisser correction was applied.

rating score of 'Chaotic–Calm' at locations B and C were significantly higher than those at locations A and D across the reproduction methods ($p < 0.001$).

4.2. Effect on perceived spatial quality across reproduction methods

The spatial quality of the reproduction method was investigated in terms of source-related attributes and environment-related attributes independently, as detailed in Section 3.2.2.

4.2.1. Source-related spatial attributes

A total of four source-related attributes were investigated, where three attributes (i.e. direction, distance and width) were minor attributes pertaining only to dominant sources, and the last

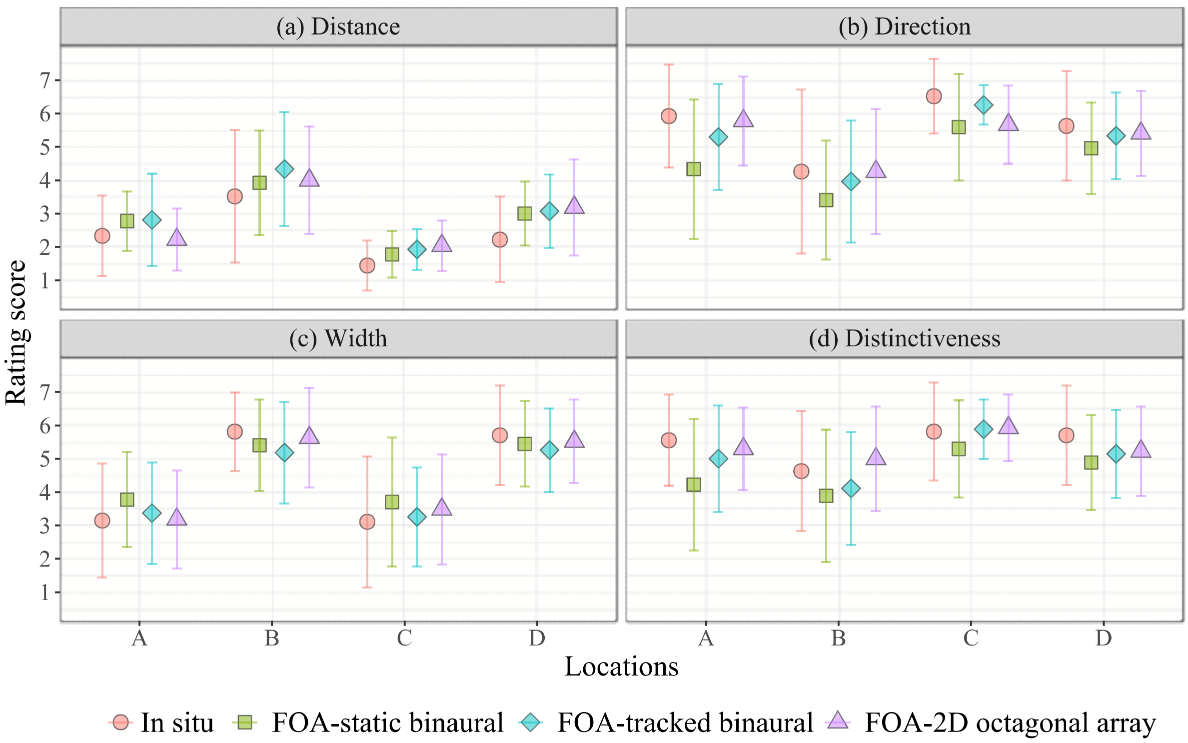


Fig. 5. Mean rating scores of the source-related spatial attributes: (a) distance, (b) direction, (c) width of the dominating sound and (d) distinctiveness of the sound sources, as a function of the locations A to D, across the (1) in situ ( ), (2) FOA-static binaural ( ), (2) FOA-tracked binaural ( ) and (3) FOA-2D octagonal array ( ) conditions. The error bars display standard deviations.

attribute (i.e. distinctiveness) was a major attribute that considers all the audible sound sources (ensemble) in the whole environment. The mean rating scores of the four source-related attributes, (a)direction, (b) distance, (c) width of the dominating sound and (d) distinctiveness of the sound sources, are plotted in Fig. 5(a to d), respectively, as a function of the locations A to D. Because the spatial characteristics of the dominant sound sources varied across the locations, a two-way RM ANOVA test was performed for each source-related spatial attribute to examine the statistical differences with respect to the reproduction methods and locations, as summarized in Table 7. The RM ANOVA results revealed that the main effects of the reproduction methods and the locations on four spatial attributes were statistically significant ($p < 0.001$). This indicates that source-related spatial quality differed across the reproduction methods and locations with no interactions found between the spatial attributes and the locations.

Table 7. Summary of the RM ANOVA for the four source-related spatial attributes with the FOA reproduction methods and the locations.

Source-related attributes	Independent factors	df_1	df_2	F	p	η_p^2	$1-\beta$
Direction	Reproduction	3	78	6.51	< .001	0.20	0.96
	Location	3	78	19.51	< .001	0.43	1.00
	Interaction [g)]	5.64	146.72	1.54	0.17	0.06	0.56
Distance	Reproduction	3	78	9.30	< .001	0.26	1.00
	Location [g)]	1.68	43.79	29.50	< .001	0.53	1.00
	Interaction	9	234.00	1.61	0.11	0.06	0.74
Width	Reproduction [g)]	2.16	56.12	1.35	0.27	0.05	0.29
	Location	3	78	44.49	< .001	0.63	1.00
	Interaction	9	234.00	1.30	0.24	0.05	0.63
Distinctiveness	Reproduction	3	78	7.13	< .001	0.22	0.98
	Location [g)]	2.14	55.75	10.76	< .001	0.29	0.99
	Interaction	9	234.00	1.20	0.30	0.04	0.58

[g)] Assumption of sphericity was violated and Greenhouse–Geisser correction was applied.

To examine the mean differences of each source-related spatial attribute across the reproduction methods and locations, post hoc tests were conducted. As shown in Fig. 5(a), the distance of the dominating sound varied across the four locations. The participants perceived the dominant sound at location B, which was tranquil, further than the other locations ($p < 0.05$), whereas the dominant fountain sound at location C was perceived as nearer than the dominant sounds at the other locations ($p < 0.001$). Regarding the reproduction methods, the participants perceived the distance of the dominating sounds further in the virtual than in the in situ soundscape evaluations ($p < 0.05$). Whereas there were statistical differences, the absolute differences were small.

The dominant sound sources at location A, C and D had strong directivities, whereas the acoustic environment at location B was less directive and ambient, as shown in Fig. 5(b). Post hoc tests revealed that there were no significant differences in perception of direction for the dominating sound at each location among the in situ, FOA-tracked binaural and FOA-2D octagonal array reproductions. Only the FOA-static binaural reproduction was significantly different from the in situ cases ($p < 0.05$).

Regarding the sound source width, the participants perceived that the dominating sounds were coming from a wide area in locations B and D, whereas the width of the main sound sources at locations A and C were relatively narrow, as shown in Fig 5(c). These observations are consistent with the acoustical characteristics of the dominant sources, for instance: the exhaust could be considered a point source in location A; birdsongs were present all around in location B; the fountain sound could also be considered as a point source in location C and the major expressway behaved like a line source in location D. There were no significant differences in the perception of sound source widths between the in situ and virtual soundscape evaluations in all FOA reproduction methods.

For the perception of distinctiveness of sound sources in the acoustic environment, the mean rating scores are shown in Fig. 5(d). The participants distinctively perceived the directions and locations of the sound sources in all the locations, except location B. There were no significant differences in the spatial distinctiveness of sound sources at each location between the in situ and the virtual soundscape evaluations with the FOA-tracked binaural and the FOA-2D octagonal array reproduction. However, it is important to note that the distinctiveness of sound sources for FOA-static binaural was significantly lower than those in situ and of the FOA-2D octagonal array reproduction ($p<0.05$).

4.2.2. Environment-related spatial quality attributes

The perceived environment-related spatial quality across three FOA reproduction methods was implicitly compared with the in situ experience based on memory (except for 'realism', which was explicitly compared). The subjective ratings of the four environment-related spatial attributes, (a) immersion, (b) realism, (c) externalization and (d) overall listening of experience, are presented in Figs. 6(a–d), respectively, across the three reproduction methods.

Two-way RM MANOVA tests were performed to examine the effects of reproduction method, location and interaction (reproduction method × location) in the set of attributes (i.e. immersion, realism, externalization and overall listening experience) describing environment-related spatial qualities.

As shown in Table 8, F-values for testing main effects of reproduction method were significant for all four attributes ($p < 0.001$). Post hoc tests were performed to find the statistical differences in perceived environmental-related spatial attributes among the three reproduction methods. The results showed that the rating scores of FOA-static binaural and FOA-tracked binaural reproductions regarding all four attributes (i.e. realism, externalization of sounds and overall listening experience) were significantly lower than those of FOA-2D octagonal array reproductions ($p < 0.05$). There were no significant differences between FOA-

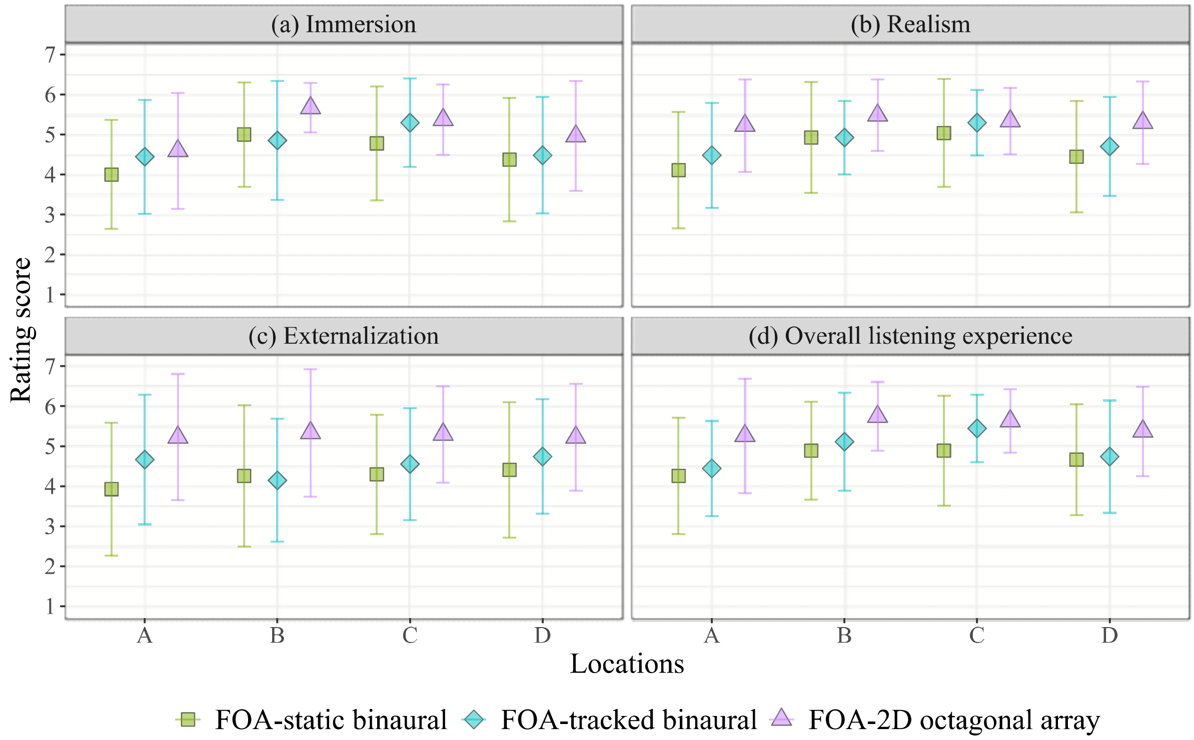


Fig. 6. Mean rating scores for the environment-related spatial quality attributes: (a) immersion, (b) realism, (c) externalization of sounds and (d) overall listening experience, as a function of the locations A to D, with three FOA reproduction methods (1) FOA-static binaural (■), (2) FOA-tracked binaural (◆) and (3) FOA-2D octagonal array (▲) reproductions. The error bars indicate standard deviation.

static binaural and FOA-tracked binaural reproductions for all environment-related spatial attributes.

As listed in Table 8, regarding the main effects of location, the differences in immersion [$F(3,78) = 8.62$, $\eta_p^2 = 0.25$, $p < 0.001$, $1 - \beta = 0.99$], realism [$F(3,78) = 7.62$, $\eta_p^2 = 0.23$, $p < 0.001$, $1 - \beta = 0.98$], overall listening experience [$F(2.16,56.28) = 6.15$, $\eta_p^2 = 0.19$, $p < 0.001$, $1 - \beta = 0.89$], among the three reproduction methods, were significant, except for externalization [$F(2.03,52.80) = 0.45$, $\eta_p^2 = 0.25$, $p = 0.64$, $1 - \beta = 0.12$].

The post hoc tests for immersion, realism and overall listening experience showed that location C (open area with a water fountain) had significantly higher scores of immersion,

Table 8 Summary of the RM MANOVA tests for environment-related spatial attributes with the FOA reproduction methods and the locations

Factors	Environment-related attributes	df_1	df_2	F	p	η_p^2	1-β
Reproduction	Immersion	2.00	52.00	6.72	< .001	0.21	0.90
	Realism	2.00	52.00	9.29	< .001	0.26	0.97
	Externalization	2.00	52.00	11.27	< .001	0.30	0.99
	Overall listening quality	2.00	52.00	12.95	< .001	0.33	1.00
Location	Immersion	3.00	78.00	8.62	< .001	0.25	0.99
	Realism	3.00	78.00	7.62	< .001	0.23	0.98
	Externalization [g)]	2.03	52.80	0.45	0.64	0.02	0.12
	Overall listening quality [g)]	2.16	56.28	6.15	< .001	0.19	0.89
Interaction	Immersion [g)]	4.32	112.41	1.00	0.41	0.04	0.32
	Realism [g)]	4.58	119.18	1.54	0.19	0.06	0.50
	Externalization [g)]	4.17	108.45	1.25	0.29	0.05	0.39
	Overall listening quality [g)]	4.15	107.92	0.94	0.44	0.04	0.30

[g)] Assumption of sphericity was violated and Greenhouse–Geisser correction was applied.

realism and overall listening experience than locations A (crowded area with high ventilation noise) and D (part exposed to heavy traffic). No significant difference in the three spatial attributes were found between locations C and B (tranquil area beside a small lake). Lastly, there were no significant differences between locations A and D regarding the three spatial attributes. This implies that participants might rate higher environment-related spatial quality in locations with good soundscape quality than locations with poor soundscape quality. The interactions between reproduction and location were not significant for all four environment-related attributes, as shown in Table 8.

5. Discussion

5.1. Limitations

The present study has some inherent limitations. One limitation is the age distribution of the participants, who were mainly in their 20s. This may not be representative of other age groups. Additionally, the in situ soundscape evaluations were conducted across three different days with three different groups of participants. Because a within-subjects design with repeated measures was employed, the evaluation of the audio-visual recordings was confined to the corresponding in situ soundscape evaluation experienced by the participant. This evaluation protocol may have introduced a bias, even though the acoustic and meteorological conditions were consistent and stable across the three days. Moreover, there is a potential order bias in the in situ evaluations. Thus, a future study should be conducted by employing additional subjects, alternate ordering and multiple comparisons with different recordings (i.e. different from the participants' in situ evaluation) to enhance the repeatability and bias independency in this study.

Whereas the virtual soundscapes were presented in random order across all locations and reproduction techniques to ameliorate a potential memory bias, there is still an issue that participants may have tried to be consistent with their prior judgements, which they might have remembered, thereby underestimating the actual differences between the three FOA reproduction methods and in situ.

Furthermore, there was a slight difference in the evaluation procedure between the in situ and virtual soundscape evaluations. During the in situ experiments, the participants evaluated the acoustic environment approximately within 1 min, whereas in the virtual soundscape evaluations, the participants filled in the questionnaire after experiencing the reproduced audio-visual stimulus. Hence, the participants assessed the soundscape based on their recall of the experience, possibly resulting in discrepancies when comparing between in situ and virtual soundscape evaluations.

5.2. Differences in perceived overall soundscape quality

For a clearer analysis, the subjective assessment results across the three FOA reproduction techniques and in situ are summarised in Table 9. Regarding the overall soundscape quality (Section 4.1) and source-related spatial attributes (Section 4.2), we focused on the relative comparisons based on the statistical analyses (i.e. RM MANOVAs and post hoc tests) between the in situ and the three FOA reproduction methods.

Table 9. Summary of the statistical significance of the overall soundscape quality and source-related attributes when compared with the in situ case are represented either by a circle (○: not significant) or a triangle (▲: significant).

Subjective attributes	Acoustic reproduction methods		
	FOA-static binaural	FOA-tracked binaural	FOA-2D octagonal array
Overall soundscape quality			
• Dominance of sound sources	○	○	○
• Affective quality of soundscape	○	○	○
Source-related spatial attributes			
• Distance	▲	▲	▲
• Directivity	▲	○	○
• Width	○	○	○
• Distinctiveness	▲	○	○

The overall soundscape quality was assessed based on the perceived dominant sound sources and the perceived affective quality in each location. Overall, there were no significant differences in perceived dominant sound sources and affective quality of soundscape between three virtual reproduction methods and in situ evaluations, as shown in Table 9. These findings are in agreement with previous studies showing that VR HMDs could be a reliable tool for soundscape assessment as an alternative to on-site surveys [25–27].

However, there were some differences in the perceived dominance of intermittent sounds at the locations, such as birdsong, wind and traffic, from the minor road. These findings suggest that the 1-min recordings might be insufficient for representing the experienced in situ acoustic environments, especially for intermittent sounds. For instance, locations A and B were adjacent to minor roads where vehicles were intermittently passing-by. However, the recordings used in the laboratory tests did not contain many traffic sounds.

5.3. Differences in perceived spatial quality

Unlike the results of the overall soundscape quality evaluations, significant differences were found in the perceived spatial quality among the three FOA reproduction methods. Based on the statistical analyses of the subjective experiments, the headphone-based FOA-tracked binaural tracks had similar source-related spatial qualities (i.e. direction, distance, width and distinctiveness) with the FOA-2D octagonal array speaker system, whereas FOA-static binaural reproduction exhibited lower scores. In assessing environment-related spatial attributes (i.e. realism, externalization of sounds and overall listening experience), the FOA-2D octagonal array system was preferred over both FOA-static and FOA-tracked binaural reproduction.

The significant differences in the perception of dominating source distance could be attributed to the differences between the in situ evaluation position and the recording device position or the occlusion (i.e. bird, traffic) or omnidirectionality (i.e. wind) of dominating sources. It is worth noting that the overestimation of the perceived distance, as depicted in Fig, 5(a), contradicts the classic underestimation of source distance for far sources in virtual acoustics [62–65]. Additionally, this overestimation also contradicts the underestimation of perceived visual distances in VR HMDs [66,67] and of sound source distances in audio-visual modalities [68,69]. However, the aforementioned findings may not directly apply to cinematic VR systems [70] in the judgement of distance in outdoor scenarios with modern VR HMDs. Therefore, further investigation is required.

Regarding directivity and distinctiveness, the participants evaluated that dominant sound sources were less directive and distinctive with FOA-static binaural, as shown in Table. 9. Whereas head movements were found to improve externalization in non-head-tracked binaural playback [13] and the localization of sound sources in the environment, the effects of head movements appear to be insufficient, as revealed in this study. This result suggests that the improvement in externalization from using head-tracked binaural [54,55] with non-individualised HRTF [13] (i.e. FOA-tracked binaural) was statistically similar to the in situ perception of the directivity and distinctiveness of dominating sounds.

Hence, there is evidence supporting the use of FOA-tracked binaural for assessing outdoor acoustic environments with heavy emphasis on source-related spatial impressions. However, this observation should not be generalized to cases where localization and externalization of sound sources, especially with elevation, are the subjects of investigation. Although the omission of elevated sources in this study could be attributed to the poor elevation projection of the FOA reproduction methods used, the 3D speaker array nor higher-order ambisonics (HOA) were investigated as this study was motivated by the consumer availability of FOA recording and playback devices.

5.4. Role of virtual acoustic reproduction methods in soundscape applications

The findings of this study reaffirm previous recommendations that the level of perceptual accuracy of the acoustic reproduction techniques should correspond to the research objectives, even with the introduction of VR techniques [2,36]. For instance, when assessing a location’s overall soundscape quality [47], the FOA-static binaural or dummy-head-recorded binaural reproduction is sufficient. However, when the spatial aspects of soundscape elements are explored, FOA-tracked binaural or FOA-2D speaker arrays should be utilised to achieve sufficient realism and localization accuracy.

Cinematic VR or computer-generated VR techniques have enabled the stakeholders of the soundscape design process, such as urban planners and architects, to realize their designs in the virtual 3D space. Hence, the spatial fidelity of the acoustic elements is important for an accurate aural perception of the design iterations. For instance, when investigating the degree to which the spatial positioning of added sounds (e.g. water fountains, virtual loudspeakers playing birdsongs) energetically or informationally 'mask' [71,72] noise in the environment, sufficient timbre quality and accurate localization of sound sources in the 3D space is required [38]. Currently, the FOA techniques investigated in this study seem sufficient in the perception of source directivities (i.e. FOA-tracked binaural, FOA-2D speaker array) but are unable to provide sufficient fidelity in terms of distance perception. Perception of distance could be affected by both the reproduction medium and source material [35]. Guastavino and Katz found that FOA-2D speaker array produced an auditory scene that appeared nearer than 1D and 3D speaker setups. More importantly, a large variation in the perceived distance of a traffic scene was observed across 1D, 2D and 3D speaker setups, which warrants further investigation.

When non-expert listeners are recruited to evaluate the soundscape designs in VR during the participatory design approach [73,74], spatial audio reproduction methods with sufficient ecological validity (i.e. FOA-tracked binaural, FOA-2D speaker array) is crucial in ensuring a perceptually accurate assessment [24].

The versatility of ambisonics recording and reproduction [2,35] prompts further investigation into its reliability for soundscape design that involves virtual augmentation of sound sources in the real (augmented reality) or virtual acoustic environments (VR). Recommendations for enhancing FOA reproduction methods, especially for headphone-based methods, owing to their portability, will be discussed in Section 5.5.

5.5. Recommendations for future work

The study of FOA-based reproduction methods was motivated by the increasing consumer availability of FOA microphones and decoders. However, the limited spatial resolution of FOA could be the main limiting factor in the localization accuracy of sound sources. The spatial fidelity can be increased by adopting HOA microphones of at least third and up to the fifth order [39]. It is also worthwhile to investigate more cost-efficient parametric methods such as

DirAC [75–77] and HARPEX [78,79], which have been found to improve the spatial fidelity of FOA signals. Recently, DirAC has also been extended to enhance HOA signals [76,77].

The FOA-2D octagonal array reproduction was limited to the horizontal plane, projecting the sound scene predominantly in the horizontal plane. Hence, the FOA-2D octagonal array could be extended to a periphonic (dual-circular array) speaker system paired with higher-order ambisonics and/or parametric methods for enhanced elevation reproduction [19].

Because there were significant differences observed when evaluating spatial attributes using non-individualized HRTFs with expert listeners [80,81], further investigation into the degree to which non-expert listeners were affected when the binaural tracks were rendered with individualized HRTFs [82,83] and/or headphones with frontal emitters [7] could shed light on studies involving non-expert participants (i.e. soundscape). Despite the limitations, this study provides important information on the viability of FOA reproduction methods in terms of their application in soundscape research and design.

6. Conclusions

The FOA reproduction methods investigated here built upon previous findings to include emerging technologies and methodologies for soundscape assessment. Specifically, the headphone-based FOA-static and FOA-tracked binaural reproductions and the FOA-2D octagonal speaker array setups were evaluated in terms of overall soundscape quality and perceived spatial quality via virtual soundscape evaluations in comparison (implicitly and explicitly) to in situ experiences.

In the evaluation of overall soundscape quality, no significant differences were found in identifying the dominance of sound sources and the perceived affective quality of soundscape in the comparison between virtual and in situ evaluations, whereas significant differences were found in the perceived spatial quality across the virtual reproduction methods.

Regarding source-related spatial quality, the FOA-tracked binaural playback and the FOA-2D octagonal speaker array could reproduce sufficient spatial aural fidelity in VR for soundscape assessment. However, the environment-related spatial reproduction quality of the FOA-tracked binaural was lower than the FOA-2D speaker array, limiting its utility in instances requiring spatial localization accuracy (e.g. the perception of distance, direction and elevation) and high timbre quality.

In the context of VR HMD in soundscape assessment and design, the FOA-tracked binaural reproduction appeared to be a viable alternative to FOA-2D octagonal array speaker system. Considering the benefits in the portability of headphones, the possible improvements to the FOA-based binaural methods (e.g. HOA microphones, parametric decoders) to meet the requirements of soundscape design evaluation should be pursued.

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.: L2NICCFP2-2015-5. The first author is currently supported by the 2017 Lee Kuan Yew Postdoctoral Fellowship. The authors are also grateful to Nanyang Technological University School of Mechanical & Aerospace Engineering for providing access to their anechoic chamber.

References

[1] ISO 12913-1, Acoustics — Soundscape — Part 1: Definition and conceptual framework, International Organization for Standardization, Geneva, Switzerland, 2014.

[2] J.Y. Hong, J. He, B. Lam, R. Gupta, W.-S. Gan, Spatial Audio for Soundscape Design: Recording and Reproduction, Appl. Sci. 7 (2017) 1–21. doi:10.3390/app7060627.

[3] F. Aletta, J. Kang, Ö. Axelsson, Soundscape descriptors and a conceptual framework for developing predictive soundscape models, Landsc. Urban Plan. 149 (2016) 65–74. doi:10.1016/j.landurbplan.2016.02.001.

[4] ISO 12913-2, Acoustics — Soundscape — Part 2: Data collection and reporting requirements, Geneva, Switzerland, 2018.

[5] 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. doi:10.1016/j.buildenv.2016.08.011.

[6] F. Aletta, J. Xiao, What are the Current Priorities and Challenges for (Urban) Soundscape Research?, Challenges. 9 (2018) 16. doi:10.3390/challe9010016.

[7] K. Sunder, 3D audio reproduction using frontal projection headphones, PhD Thesis, Nanyang Technological University, Singapore, 2016.

[8] D. Hammershoi, H. Moller, Binaural technique - Basic methods for recording, synthesis, and reproduction, Commun. Acoust. (2005) 223–254. doi:10.1007/3-540-27437-5_9.

[9] P. Stitt, E. Hendrickx, J.-C. Messonnier, B.F.G. Katz, The Role of Head Tracking in Binaural Rendering, 29th Tonmeistertagung - VDT Int. Conv. (2016) 1–5.

[10] D.R. Begault, 3-D Sound for Virtual Reality and Multimedia, Academic Press Professional, Inc., San Diego, CA, USA, 2000. doi:10.2307/3680997.

[11] W.M. Hartmann, How We Localize Sound, Phys. Today. 52 (1999) 24–29. doi:10.1063/1.882727.

[12] F. Rumsey, Binaural Audio and Virtual Acoustics, J. Audio Eng. Soc. 65 (2017) 524–528.

[13] E. Hendrickx, P. Stitt, J.-C. Messonnier, J.-M. Lyzwa, B.F. Katz, C. de Boishéraud, Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis, J. Acoust. Soc. Am. 141 (2017) 2011–2023. doi:10.1121/1.4978612.

[14] E. Hendrickx, P. Stitt, J.C. Messonnier, J.M. Lyzwa, B.F.G. Katz, C. De Boisheraud, Improvement of externalization by listener and Source movement using a “binauralized” microphone array, J. Audio Eng. Soc. 65 (2017) 589–599. doi:10.17743/jaes.2017.0018.

[15] N. Kitagawa, S. Ichihara, Hearing visual motion in depth, Nature. 416 (2002) 172–174. doi:10.1038/416172a.

[16] D. Thery, D. Poirier-Quinot, B.N.J. Postma, B.F.G. Katz, Impact of the Visual Rendering System on Subjective Auralization Assessment in VR, in: J. Barbic, M. D’Cruz, M.E. Latoschik, M. Slater, P. Bourdot (Eds.), Virtual Real. Augment. Real., Springer International Publishing, Cham, 2017: pp. 105–118.

[17] M.A. Gerzon, Ambisonics. Part two: Studio techniques, Stud. Sound. 17 (1975) 24–26.

[18] P. Fellgett, Ambisonics. Part one: General system description, Stud. Sound. 17 (1975) 20–22.

[19] M.A. Gerzon, Periphony: With-Height Sound Reproduction, J. Audio Eng. Soc. 21 (1973) 2–10.

[20] J.Y. Hong, J.Y. Jeon, The effects of audio-visual factors on perceptions of environmental noise barrier performance, Landsc. Urban Plan. 125 (2014) 28–37. doi:10.1016/j.landurbplan.2014.02.001.

[21] 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.

doi:10.1016/j.landurbplan.2015.05.004.

[22] S. Viollon, C. Lavandier, C. Drake, Influence of visual settings on sound ratings in an urban environment, Appl. Acoust. 63 (2002) 493–511.

[23] L. Galbrun, F.M. a Calarco, Audio-visual interaction and perceptual assessment of water features used over road traffic noise., J. Acoust. Soc. Am. 136 (2014) 2609. doi:10.1121/1.4897313.

[24] G.M. Echevarria Sanchez, T. Van Renterghem, K. Sun, B. De Coensel, D. Botteldooren, Using Virtual Reality for assessing the role of noise in the audio-visual design of an urban public space, Landsc. Urban Plan. 167 (2017) 98–107. doi:10.1016/j.landurbplan.2017.05.018.

[25] 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. doi:10.1016/j.scs.2016.06.022.

[26] V. Puyana-Romero, L.S. Lopez-Segura, L. Maffei, R. Hernández-Molina, M. Masullo, Interactive Soundscapes: 360°-Video Based Immersive Virtual Reality in a Tool for the Participatory Acoustic Environment Evaluation of Urban Areas, Acta Acust. United with Acust. 103 (2017) 574–588. doi:10.3813/AAA.919086.

[27] L. Maffei, M. Masullo, A. Pascale, G. Ruggiero, V. Puyana Romero, On the validity of immersive virtual reality as tool for multisensory evaluation of urban spaces, Energy Procedia. 78 (2015) 471–476. doi:10.1016/j.egypro.2015.11.703.

[28] F. Ruotolo, L. Maffei, M. Di Gabriele, T. Iachini, M. Masullo, G. Ruggiero, V.P. Senese, Immersive virtual reality and environmental noise assessment: An innovative audio-visual approach, Environ. Impact Assess. Rev. 41 (2013) 10–20. doi:10.1016/j.eiar.2013.01.007.

[29] J. Thatte, J.-B. Boin, H. Lakshman, B. Girod, Depth augmented stereo panorama for cinematic virtual reality with head-motion parallax, in: 2016 IEEE Int. Conf. Multimed. Expo, IEEE, 2016: pp. 1–6. doi:10.1109/ICME.2016.7552858.

[30] P. Thomas, W. Wei, T. Van Renterghem, D. Botteldooren, Measurement-based auralization methodology for the assessment of noise mitigation measures, J. Sound Vib. 379 (2016) 232–244. doi:10.1016/j.jsv.2016.05.039.

[31] M. Vorländer, Auralization, Springer-Verlag Berlin Heidelberg, Berlin, Germany, 2008. doi:10.1007/978-3-540-48830-9.

[32] C. Millns, H. Lee, An Investigation into Spatial Attributes of 360° Microphone Techniques for Virtual Reality, in: Audio Eng. Soc. Conv. 144, Audio Engineering Society, Milan, Italy, 2018: pp. 1–9.

[33] B.N.J. Postma, B.F.G. Katz, The Influence of Visual Distance on the Room-acoustic Experience of Auralizations, J. Acoust. Soc. Am. 142 (2017) 3035–3046. doi:10.1121/1.5009554.

[34] M. Olko, D. Dembeck, Y.-H. Wu, A.F. Genovese, A. Roginska, Identification of Perceived Sound Quality Attributes of 360º Audiovisual Recordings in VR Using a Free Verbalization Method, in: Audio Eng. Soc. Conv. 143, Audio Engineering Society, New York, NY, USA, 2017: pp. 1–10.

[35] C. Guastavino, B.F.G. Katz, Perceptual evaluation of multi-dimensional spatial audio reproduction, J. Acoust. Soc. Am. 116 (2004) 1105–1115. doi:10.1121/1.1763973.

[36] 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.

[37] K. Ooi, J. Hong, B. Lam, Z.T. Ong, W. Gan, Validation of binaural recordings with head tracking for use in soundscape evaluation, in: INTER-NOISE NOISE-CON Congr. Conf. Proc., Institute of Noise Control Engineering, Hong Kong SAR, China, 2017: pp. 63–71.

[38] Z.T. Ong, J.Y. Hong, B. Lam, K. Ooi, R. Gupta, Effect of masker orientation on masking efficacy in soundscape applications, in: INTER-NOISE NOISE-CON Congr. Conf. Proc., Institute of Noise Control Engineering, Hong Kong SAR, China, 2017: pp. 55–61.

[39] L. Thresh, C. Armstrong, G. Kearney, A Direct Comparison of Localisation Performance When Using First, Third and Fifth Order Ambisonics For Real Loudspeaker And Virtual Loudspeaker Rendering, in: Audio Eng. Soc. Conv. 143, Audio Engineering Society, New York, NY, USA, 2017: pp. 1–9.

[40] J. Francombe, T. Brookes, R. Mason, J. Woodcock, Evaluation of Spatial Audio Reproduction Methods (Part 2): Analysis of Listener Preference, J. Audio Eng. Soc. 65 (2017) 212–225. doi:10.17743/jaes.2016.0071.

[41] B.F.G. Katz, A. Tarault, P. Bourdot, J. Vézien, The use of 3D-audio in a multi-modal teleoperation platform for remote driving/supervision, in: Audio Eng. Soc. 30th Int. Conf. Intell. Audio Environ., Saariselkä, Finland, 2007: pp. 1–9.

[42] L. Picinali, A. Afonso, M. Denis, B.F.G. Katz, Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge, Int. J. Hum. Comput. Stud. 72 (2014) 393–407. doi:10.1016/j.ijhcs.2013.12.008.

[43] F. Faul, E. Erdfelder, A.-G. Lang, A. Buchner, G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences, Behav. Res. Methods. 39 (2007) 175–191. doi:10.3758/BF03193146.

[44] J. Cohen, Statistical power analysis for the behavioral sciences, 2nd ed., Routledge, 1988.

[45] J.Y. Jeon, J.Y. Hong, P.J. Lee, Soundwalk approach to identify urban soundscapes individually, J. Acoust. Soc. Am. 134 (2013) 803–812. doi:10.1121/1.4807801.

[46] C. Semidor, Listening to a city with the soundwalk method, Acta Acust. United with Acust. 92 (2006) 959–964.

[47] Ö. Axelsson, M.E. Nilsson, B. Berglund, A principal components model of soundscape perception, J. Acoust. Soc. Am. 128 (2010) 2836–2846. doi:10.1007/978-1-4419-0561-1_48.

[48] J.Y. Jeon, J.Y. Hong, C. Lavandier, J. Lafon, Ö. Axelsson, M. Hurtig, A cross-national comparison in assessment of urban park soundscapes in France, Korea, and Sweden through laboratory experiments, Appl. Acoust. 133 (2018) 107–117. doi:10.1016/j.apacoust.2017.12.016.

[49] 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. doi:10.1016/j.buildenv.2017.10.021.

[50] F. Rumsey, Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm, J. Audio Eng. Soc. 50 (2002) 651–666.

[51] S. Le Bagousse, C. Colomes, M. Paquier, State of the Art on Subjective Assessment of Spatial Sound Quality, in: Audio Eng. Soc. Conf. 38th Int. Conf. Sound Qual. Eval., Audio Engineering Society, Piteå, Sweden, 2010: pp. 1–8.

[52] G.M. Echevarria Sanchez, T. van Renterghem, K. Sun, B. De Coensel, D. Botteldooren, The relative importance of visual and sound design in the rehabilitation of a bridge connecting a highly populated area and a park, in: INTER-NOISE NOISE-CON Congr. Conf. Proc., Institute of Noise Control Engineering, Hamburg, Germany, 2016: pp. 6810–6816.

[53] J. Kang, M. Zhang, Semantic differential analysis of the soundscape in urban open public spaces, Build. Environ. 45 (2010) 150–157. doi:10.1016/j.buildenv.2009.05.014.

[54] A. Roginska, P. Geluso, eds., Immersive Sound, Routledge, New York ; London : Routledge, 2017., 2017. doi:10.4324/9781315707525.

[55] S.A. Gelfand, Hearing: An Introduction to Psychological and Physiological Acoustics, 6th ed., CRC Press, Boca Raton, FL, USA, 2017. doi:10.1136/jnnp.45.12.1175-b.

[56] M. Schoeffler, A. Silzle, J. Herre, Evaluation of Spatial/3D Audio: Basic Audio Quality Versus Quality of Experience, IEEE J. Sel. Top. Signal Process. 11 (2017) 75–88. doi:10.1109/JSTSP.2016.2639325.

[57] Spatial workstation, Facebook 360: Spatial Workstation User Guide (Release 3.0), 2017. https://facebook360.fb.com/spatial-workstation/.

[58] T. Lossius, J. Anderson, ATK Reaper: The Ambisonic Toolkit as JSFX plugins, in: Proc. Jt. 40th Int. Comput. Music Conf. 11th Sound Music Comput. Conf., Athens, Greece, 2014: pp. 1338–1345.

[59] J.P. Verma, Repeated measures design for empirical researchers, John Wiley & Sons, 2015.

[60] 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. doi:10.1027/1614-2241/a000016.

[61] 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. doi:10.7334/psicothema2016.383.

[62] A.J. Kolarik, B.C.J. Moore, P. Zahorik, S. Cirstea, S. Pardhan, Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss, Attention, Perception, Psychophys. 78 (2016) 373–395. doi:10.3758/s13414-015-1015-1.

[63] G. Kearney, M. Gorzel, H. Rice, F. Boland, Distance Perception in Interactive Virtual Acoustic Environments using First and Higher Order Ambisonic Sound Fields, Acta Acust. United with Acust. 98 (2012) 61–71. doi:10.3813/AAA.918492.

[64] P. Zahorik, D.S. Brungart, A.W. Bronkhorst, Auditory Distance Perception in Humans :

A Summary of Past and Present Research, Acta Acust. United with Acust. 91 (2005) 409–420.

[65] P.W. Anderson, P. Zahorik, Auditory/visual distance estimation: Accuracy and variability, Front. Psychol. 5 (2014) 1–11. doi:10.3389/fpsyg.2014.01097.

[66] V. Interrante, B. Ries, L. Anderson, Distance Perception in Immersive Virtual Environments, Revisited, in: IEEE Virtual Real. Conf. (VR 2006), IEEE, n.d.: pp. 3–10. doi:10.1109/VR.2006.52.

[67] C.J. Ziemer, J.M. Plumert, J.F. Cremer, J.K. Kearney, Estimating distance in real and virtual environments: Does order make a difference?, Atten. Percept. Psychophys. 71 (2009) 1095–106. doi:10.3758/APP.71.5.1096.

[68] M. Rébillat, X. Boutillon, É. Corteel, B.F.G. Katz, Audio, visual, and audio-visual egocentric distance perception in virtual environments, Forum Acusticum. (2011) 1–5.

[69] M. Rébillat, X. Boutillon, É. Corteel, B.F.G. Katz, Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments, ACM Trans. Appl. Percept. 9 (2012) 1–17. doi:10.1145/2355598.2355602.

[70] A. McArthur, M. Sandler, R. Stewart, Perception of Mismatched Auditory Distance—Cinematic VR, in: Audio Eng. Soc. Conf. 2018 AES Int. Conf. Audio Virtual Augment. Real., Audio Engineering Society, Redmond, WA, USA, 2018: pp. 1–10.

[71] I. Pollack, Auditory informational masking, J. Acoust. Soc. Am. 57 (1975) S5–S5. doi:10.1121/1.1995329.

[72] G.J. Kidd, C.R. Mason, V.M. Richards, F.J. Gallun, N.I. Durlach, Informational masking, Audit. Percept. Sound Sources, Springer Handb. Audit. Res. (2008) 143–189. doi:10.1007/978-0-387-71305-2.

[73] C. Li, Y. Liu, M. Haklay, Participatory soundscape sensing, Landsc. Urban Plan. 173 (2018) 64–69. doi:10.1016/J.LANDURBPLAN.2018.02.002.

[74] J. Xiao, L. Lavia, J. Kang, Towards an agile participatory urban soundscape planning framework, J. Environ. Plan. Manag. 61 (2018) 677–698. doi:10.1080/09640568.2017.1331843.

[75] V. Pulkki, Spatial sound reproduction with directional audio coding, J. Audio Eng. Soc. 55 (2007) 503–516.

[76] A. Politis, L. McCormack, V. Pulkki, Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing, in: 2017 IEEE Work. Appl. Signal Process. to Audio Acoust., IEEE, New York, NY, USA, 2017: pp. 379–383. doi:10.1109/WASPAA.2017.8170059.

[77] V. Pulkki, S. Delikaris-Manias, A. Politis, eds., Parametric Time-Frequency Domain Spatial Audio, John Wiley & Sons, Ltd, Chichester, UK, 2017. doi:10.1002/9781119252634.

[78] S. Berge, N. Barrett, High angular resolution planewave expansion, in: 2nd Int. Symp. Ambisonics Spherical Acoust., Paris, France, 2010: pp. 1–6.

[79] S. Berge, N. Barrett, A new method for B-format to binaural transcoding, in: 40th Int. Audio Eng. Soc. Conf. Spat. Audio, Audio Engineering Society, Tokyo, Japan, 2010:

pp. 1–10.

[80] L.S.R. Simon, N. Zacharov, B.F.G. Katz, Perceptual attributes for the comparison of head-related transfer functions, J. Acoust. Soc. Am. 140 (2016) 3623–3632. doi:10.1121/1.4966115.

[81] B.N.J. Postma, B.F.G. Katz, Perceptive and objective evaluation of calibrated room acoustic simulation auralizations, J. Acoust. Soc. Am. 140 (2016) 4326–4337. doi:10.1121/1.4971422.

[82] J. He, R. Ranjan, W.-S. Gan, Fast continuous HRTF acquisition with unconstrained movements of human subjects, in: 2016 IEEE Int. Conf. Acoust. Speech Signal Process., IEEE, Shanghai, China, 2016: pp. 321–325. doi:10.1109/ICASSP.2016.7471689.

[83] R. Ranjan, J. He, W.-S. Gan, Fast Continuous Acquisition of HRTF for Human Subjects with Unconstrained Random Head Movements in Azimuth and Elevation, in: 2016 AES Int. Conf. Headphone Technol., Aalborg, Denmark, 2016. doi:10.17743/aesconf.2016.978-1-942220-09-1.