Influence of parametrically-decoded First-order ambisonics reproduction in cinematic virtual-reality-based soundscape evaluation

Lam, Bhan; Hong, Jooyoung; Ong, Zhen Ting; Gan, Woon-Seng; Kang, Jian; Feng, Jing; Tan, Sze-Tiong

2019

Lam, B., Hong, J., Ong, Z.-T., Gan, W.-S., Kang, J., Feng, J., & Tan, S.-T. (2019). Influence of parametrically-decoded First-order ambisonics reproduction in cinematic virtual-reality-based soundscape evaluation. 48th International Congress and Exhibition on Noise Control Engineering (INTER-NOISE 2019).

https://hdl.handle.net/10356/105858

MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Influence of Parametrically-decoded First-order Ambisonics Reproduction in Cinematic Virtual-Reality-based Soundscape Evaluation

Lam, Bhan[1]
Hong, Jooyoung[2]
Ong, Zhen-Ting[3]
Gan, Woon-Seng[4]

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

Kang, Jian[5]

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

Feng, Jing[6]
Tan, Sze-Tiong[7]

Building & Research Institute, Housing & Development Board, 738973, Singapore

ABSTRACT

Increasing consumer availability of first-order ambisonics (FOA) microphones have enabled virtual evaluations of soundscapes with head-tracked binaural playback through headphones. Known issues with FOA binaural decoding techniques, such as timbral and spatial artefacts, however, have marred its advantages. Parametric decoding methods, such as directional audio coding (dirAC), high angular resolution plane-wave expansion (HARPEX), and the recently proposed coding and multidirectional parameterization of ambisonic sound scenes (COMPASS) have noted significant improvements in sound quality over time-invariant linear processing of FOA. Whereas subjective assessments of the parametric methods have been traditionally focused on musical media, few studies have focused on the evaluation of outdoor acoustic scenes. Hence, the differences between parametrically and linearly decoded FOA to head-tracked binaural is explored here

---

[1] blam002@e.ntu.edu.sg
[2] jyhong@ntu.edu.sg
[3] ztong@ntu.edu.sg
[4] ewsgan@ntu.edu.sg
[5] j.kang@ucl.ac.uk
[6] feng_jing@hdb.gov.sg
[7] tan_sze_tiong@hdb.gov.sg

through subjective assessments of outdoor acoustic scenes. Participants evaluated the perceived spatial sound quality and the soundscape quality of four reproduced outdoor scenes in cinematic virtual reality through a head-mounted device. The spatial quality was evaluated based on bipolar attributes related to the overall impression and impression of the acoustic environment during head movement. The results show that the evaluation of soundscape quality was independent of the reproduction methods. However, a notable improvement in the perceived spatial quality was observed in the parametric reproductions in terms of directionality and distinctiveness of sound sources.

Keywords: Soundscape, Spatial Audio, Virtual Reality
I-INCE Classification of Subject Number: 63; 74; 76; 79

1. INTRODUCTION

Acoustic-only soundscape evaluations are usually conducted in a laboratory environment to reduce the effects of uncontrolled factors and for better repeatability. Hence, these evaluations are normally based on high-fidelity binaural recordings from calibrated artificial head and torso simulators as recommended by ISO 12931-2 [1]. Due to the complex nature of human perception, these unimodal analyses are limited in scope and ecological validity [2]. Moreover, there is a growing body of literature highlighting the interdependency between audition and vision when judging acoustics in an urban environment [3–9].

For increased ecological validity, there is a notable adoption of computer-generated or cinematic omnidirectional visual scenes that are usually reproduced in a virtual-reality (VR) head-mounted device (HMD). To achieve a high level of audio-visual congruence, the reproduced audio has to be “head-tracked” such that the acoustic scene follows the movements of the human head. However, the static nature of the abovementioned binaural recordings, which often results in front-back confusions and in-head localisations, might affect the overall spatial impressions [10,11] and are unsuitable for such dynamic assessments. At present, there are no straightforward methods to capture high-fidelity binaural recordings for head-tracked reproduction in VR HMDs [12]. However, there are a wide variety of spatial audio recording and reproduction methods that can be adopted for VR-based reproduction [13].

As the audio industry progresses towards reproduction-format-agnostic representations that can be decoded for a wide variety of playback systems [14], ambisonics has emerged as a popular recording format due to its flexibility [13]. Ambisonics can be described as a method to represent the sound field with spherical harmonics [15,16]. Since number of ambisonic channels or orders dictate the perceptual resolution of the acoustic scene, a higher-order setup of at least the third up to the fifth order is preferable to achieve good localization and timbral quality [16–19]. High-order ambisonic (HOA) microphones, however, are not commonplace and require more elaborate setups and higher bandwidths (i.e. 15 channels for third order).

Traditionally, 4-channel first-order ambisonic (FOA) signals are decoded linearly, which possess inherent issues such as timbral coloration in the higher frequencies and poor localization of directional sounds. As a cost-effective solution to alleviate the limitations of FOA without resorting to HOA setups, parametric decoding methods have been developed to enhance the perceptual quality of FOA recordings.

One of these methods, the directional audio coding (DirAC), is based on the extraction of one direction-of-arrival (DoA) and diffuse streams in the time-frequency domain [16,20]. Another method, the high angular resolution plane-wave expansion (HARPEX), decomposes the sound field into two plane waves and estimates their DoA and amplitudes, omitting diffuse components. More recently, another parametric technique termed coding and multidirectional parameterization of ambisonic sound scenes (COMPASS) was proposed [21]. The COMPASS approach builds upon the limitations of DirAC while maintaining computational efficiency, which is important for headphone-based head-tracked binaural playback.

Although the subjective evaluations of the abovementioned parametric methods have yielded convincing results in terms of improvements over linearly decoded FOA [21–23], majority of the sound material evaluated has been based on musical tracks and in indoor environments. Urban sound scenes are generally more complex and diffused with multiple dominating sources. Hence, there is a need to investigate the performance of parametric methods in complex urban sound scenes.

In this paper, the influence of the parametric over linear methods will be subjective assessed based on the perceived overall soundscape quality and the perceived spatial quality. The recorded audio-visual scenes will be projected through VR HMDs with headphones.

2. METHOD

2.1 Participants

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

2.2 Experimental design

Participants will be presented with audio-visual recordings of four outdoor locations through a VR HMD and headphones in a quiet listening room. The audio stimulus was decoded to the headphones with four reproduction methods, which will be described in the next sub-section. The 16 stimuli (4 locations × 4 reproduction methods) were presented in a randomised order and the audio reproduction methods were not identifiable.

After experiencing each audio-visual stimulus, participants were instructed to rate attributes that assesses both the perceived overall soundscape quality and the perceived spatial quality on a questionnaire. Participants were allowed to revisit the stimulus as many times as required throughout the questionnaire.

2.3 Stimuli

The audio-visual environment of four urban outdoor locations were recorded in the NTU campus with a spherical panoramic camera (Garmin VIRB 360 Action Camera, USA) and a low-noise FOA microphone (Sennheiser AMBEO VR 3D Microphone, Germany). The audio-visual capturing system was mounted on a tripod and fixed at approximately 1.6 m off the ground. The same audio-visual recordings of the four locations were used in a prior study [12] and are described in Table 1.

The videos were recorded in 4K resolution and post-processed for white balance and exposure compensations (Adobe Premiere Pro CC 2019). The final video render maintains the 4K resolution but encoded in the H.265 format with a bitrate of 25 Mbps at 30 frames per second.

Table 1. Description of the sound scenes under test

Location	Description
A	Open area near a canteen and walkway with moderately high human traffic
B	Quiet area beside a small lake located in a minor depression surrounded by greenery
C	Open area with a small fountain in between a museum and a minor road
D	Open area in a park flanked by a major expressway

The FOA recordings were converted to the B-format AmbiX with the AMBEO A-B converter. A total of four reproduction methods were tested as described in Table 2. Since a HOA recording was not available, a mono-channel low-quality anchor decoded from the FOA signal was included and denoted as “Mono”. The “FOA-tracked” method is a head-tracked binaural reproduction of the FOA signal based on the Facebook 360 Spatial Workstation plugin for Reaper [24]. The parametric HARPEX method was employed to up-mix the FOA recording to third-order ambisonics but still linearly decoded to head-tracked binaural playback using the same plugin [24] and is denoted as “UFOA-tracked”. The FOA-tracked and UFOA-tracked methods represent the traditional linearly decoded FOA reproduction. Lastly, the parametric COMPASS framework [21] was adopted to generate head-tracked binaural tracks and is denoted as “CFOA-tracked”. The DirAC method was omitted as it is limited to scenarios where there is primarily a single directional source.

One-minute excerpts of the visual scenes were projected to the VR HMD (Pimax 4K VR, China) via the GoPro VR player and the synced audio was presented through headphones (Beyerdynamic Custom One Pro, Germany) via the Reaper digital audio workstation.

Table 2. Reproduction methods under test

Identifier	Description
Mono	Low-quality mono reference decoded from FOA
FOA-tracked	Linear FOA head-tracked binaural reproduction
UFOA-tracked	Linear 3rd order head-tracked binaural reproduction up-mixed from FOA with HARPEX
CFOA-tracked	Parametric FOA head-tracked binaural reproduction using COMPASS

2.4 Materials

The attributes in the questionnaire can be classified into two categories pertaining to the overall soundscape quality and perceived spatial quality. The identifiers and descriptors of all the attributes are summarised in

Table 3. The descriptors in the table represent orthogonal pairs of the rating scale.

The overall soundscape quality is judged based on the perceived dominance of sound sources and the perceived affective quality, which adheres to the Swedish Soundscape Quality Protocol [25,26]. Based on a prior study [12], the participants were instructed to rate the dominance of a pre-determined list of sound sources, as shown in

Table 3, on a 5-point scale. An open-ended “Others” option was included to capture any perceived dominant sources outside the pre-determined list.

Table 3. Identifier and descriptors of the subjective attributes evaluated

	Attributes	Identifier	Descriptor
Overall Soundscape Quality	Dominance of sound sources	Traffic noise	Do not hear at all – Dominates completely
		Human sounds	
		Water sounds	
		Bird sounds	
		Wind sounds	
		Ventilation sounds	
		Others	
	Perceived affective quality	Pleasantness	Unpleasant – Pleasant
		Eventfulness	Uneventful – Eventful
		Calmness	Chaotic – Calm
		Liveliness	Boring – Lively
Spatial Quality	Source-related spatial attributes	Directionality	Non-directional – Directional
		Width	Narrow – Wide
		Distance	Near – Far
		Distinctiveness	Indistinct – Distinct
	Environment-related spatial attributes	Immersiveness	Not immersed at all – Immersed
		Externalisation	Inside head – Outside head
		Realism	Not realistic at all – Extremely realistic
		Timbral quality	Muffled – Clear
		Overall listening experience	Very bad – Very good

Since the perceived affective quality model was previously shown to provide a comprehensive description of the soundscape [27], it was adopted to represent the overall soundscape quality. Participants were instructed to rate the paired adjectives relating to the “pleasantness”, “eventfulness”, “calmness”, and “liveliness” identifiers on a 7-point scale.

The perceived spatial quality was evaluated with a scene-based approach [12,28], which are further categorised into source-related and environment-related spatial attributes. All the spatial quality attributes listed in

Table 3 are rated on a 7-point scale. Whereas the source-related attributes identified as “directionality”, “width” and “distance” describe the spatial characteristics pertaining only to the dominant sound source identified, the attribute “distinctiveness” describes the distinctiveness of the directions of all the audible sound sources in the scene.

In contrast to source-related attributes, environment-related attributes “immersiveness”, “externalisation” and “realism” evaluate the scene as a whole rather than evaluating individual or an ensemble of sources. The “immersiveness” attribute describes the sense of presence in the environment, “externalisation” describes the degree to which the sounds were perceived outside the head, and "realism" describes how realistic the scene sounds as compared to the real environment.

Lastly, the overall sound quality is evaluated with a timbral quality rating describing the sounds were "muffled" or "clear" and with a measure of overall listening experience that is analogous to the Basic Audio Quality metric [29].

2.5 Data Analyses

Two-way repeated measure (RM) analysis of variance (ANOVA) was performed to examine the within-subjects effects of the locations and reproduction methods on the subjective attributes. Particularly, RM multivariate ANOVA (MANOVA) was conducted for environment-related spatial quality and perceived affective quality of soundscape which consisted of the set of attributes measuring cohesive themes. Mauchly's test of sphericity was carried out to examine whether the data set satisfied the assumption of sphericity or not. If the assumption of sphericity was violated, then the Greenhouse–Geisser correction was applied. In all the RM ANOVA tests, post hoc comparisons were performed using Bonferroni correction.

3. RESULTS

3.1 Overall soundscape quality

Regarding the overall soundscape quality, the perceived dominance of sound sources and the perceived affective quality at each location were evaluated with four different acoustic reproduction methods. Mean rating scores of the perceived dominance of the six sound sources at the four locations across the four reproduction methods are plotted in Figure 1.

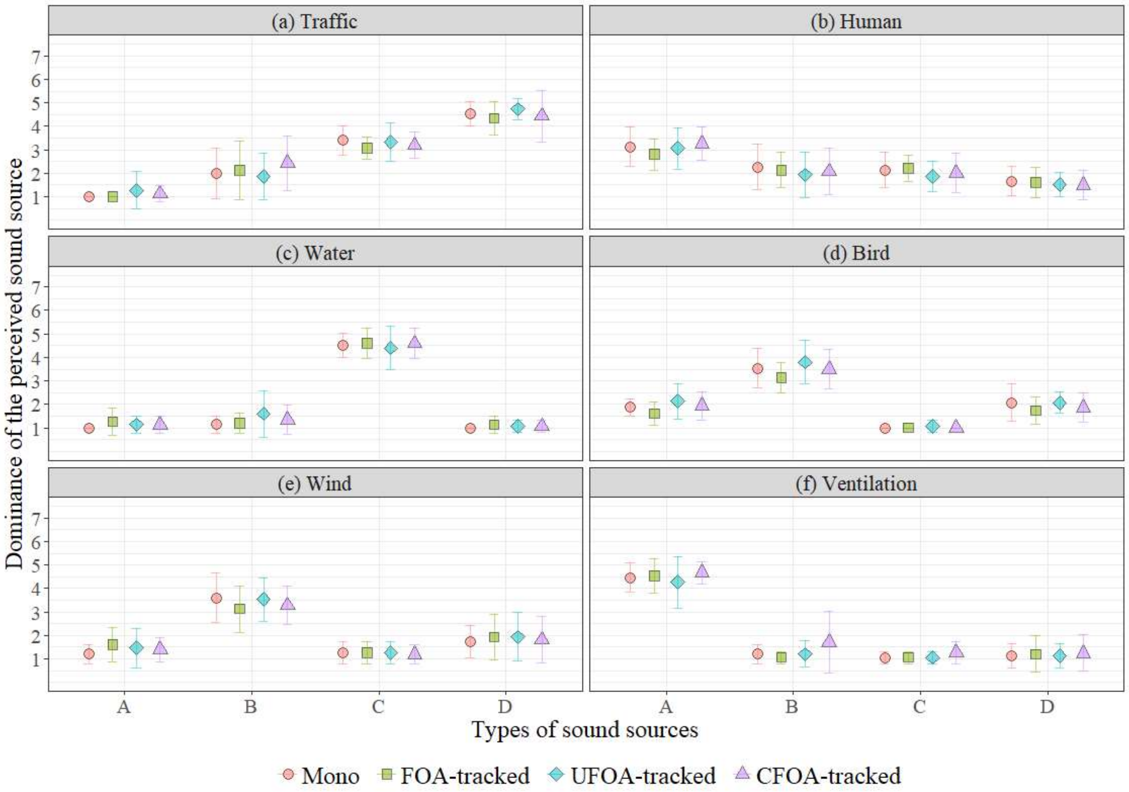

Figure 1. Mean rating scores of the judged dominance of (a) Traffic, (b) human, (c) water, (d) bird, (e) wind, and (f) ventilation sounds, as a function of the locations A to D, across the mono (○), FOA-tracked (■), UFOA-tracked (◆), and CFOA-tracked (△) reproduction methods.

Ventilation noise was the most dominant at location A. Whereas there were no prominent sound source types at location B, birdsongs and wind sound were identified more prominently than other sound source types at location B. Location C was dominated by water sounds from the fountain, followed by traffic sound from a minor road. Traffic noise from an adjacent expressway was the most dominant sound source at location D.

Two-way RM ANOVAs were conducted for each sound source type to determine the effects of locations and reproduction method on perceived dominance of sounds. The results showed that there were significant differences in the dominance of sound sources across the locations at 0.01 significance level, while the effects of reproduction methods were not significant. The interaction effects between the location and reproduction method for all sound source types were not statistically significant.

Mean rating scores of four affective quality attributes (pleasantness, eventfulness, calmness, and liveliness) are plotted as a function of the locations A to D across the four reproduction methods as shown in Figure 2.

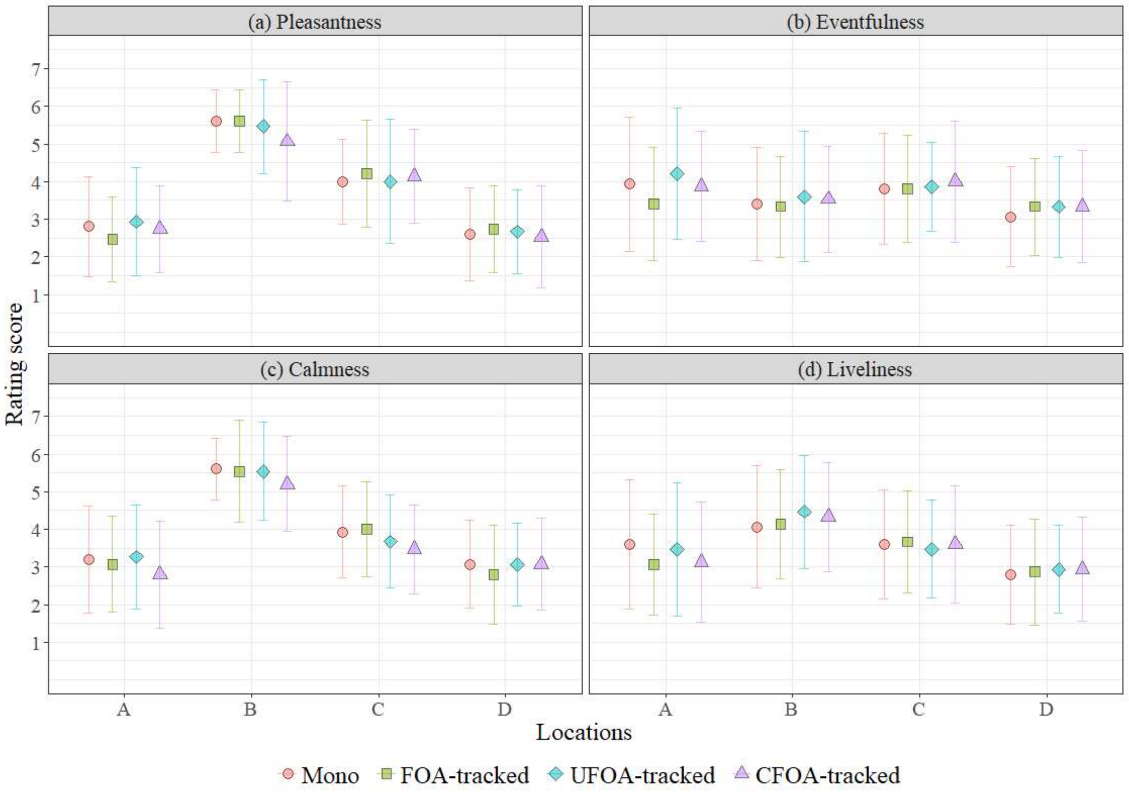

Figure 2. Mean rating scores of the judged affective quality attributes (a) pleasantness, (b) eventfulness, (c) calmness, and (d) liveliness, as a function of the locations A to D, across the mono (●), FOA-tracked (■), UFOA-tracked (◆), and CFOA-tracked (▲) reproduction methods.

Two-way RM MANOVA was also conducted for the perceived affective quality. The results revealed that the mean rating score of the soundscape attributes including pleasantness (F1.7,23.9 = 43.16, p < 0.001), calmness (F3,42 = 32.38, p < 0.001), and liveliness (F3,42 = 6.65, p = 0.001) were significantly difference across the location, whereas there was no significant difference in eventfulness among the locations (F3,42=1.23, p = 0.31). The effects of the reproduction methods and interaction between the location and reproduction method were not statistically significant. These findings support the results of a previous study [12] that the effect of reproduction methods on the evaluation of the sound-source dominance and affective soundscape qualities is rather marginal.

3.2 Perceived spatial quality

Perceived spatial quality of the reproduction method was assessed in terms of both the source-related attributes and environment-related attributes across the locations. Figure 3 shows the mean rating scores of the four source-related attributes as a function of locations across the four reproduction methods.

To examine the statistical mean differences with respect to the reproduction methods and locations, Two-way RM ANOVA tests were performed for each source-related spatial attribute. The results showed that there were significant differences in perception of distance ($F_{1.98,27.76} = 8.61$, $p < 0.001$), distinctiveness ($F_{3,42} = 2.87$, $p = 0.048$), and width ($F_{3,42} = 4.19$, $p = 0.01$) of dominating sound sources across the locations. The main effects of reproduction methods were significant for directionality ($F_{3,42} = 5.64$, $p < 0.01$) and distinctiveness ($F_{3,42} = 5.30$, $p < 0.01$) of sound sources, whereas no significant differences in distance ($F_{3,42} = 0.12$, $p = 0.95$) and width ($F_{3,42}=1.03$, $p = 0.39$) of sound sources were found across the reproduction methods.

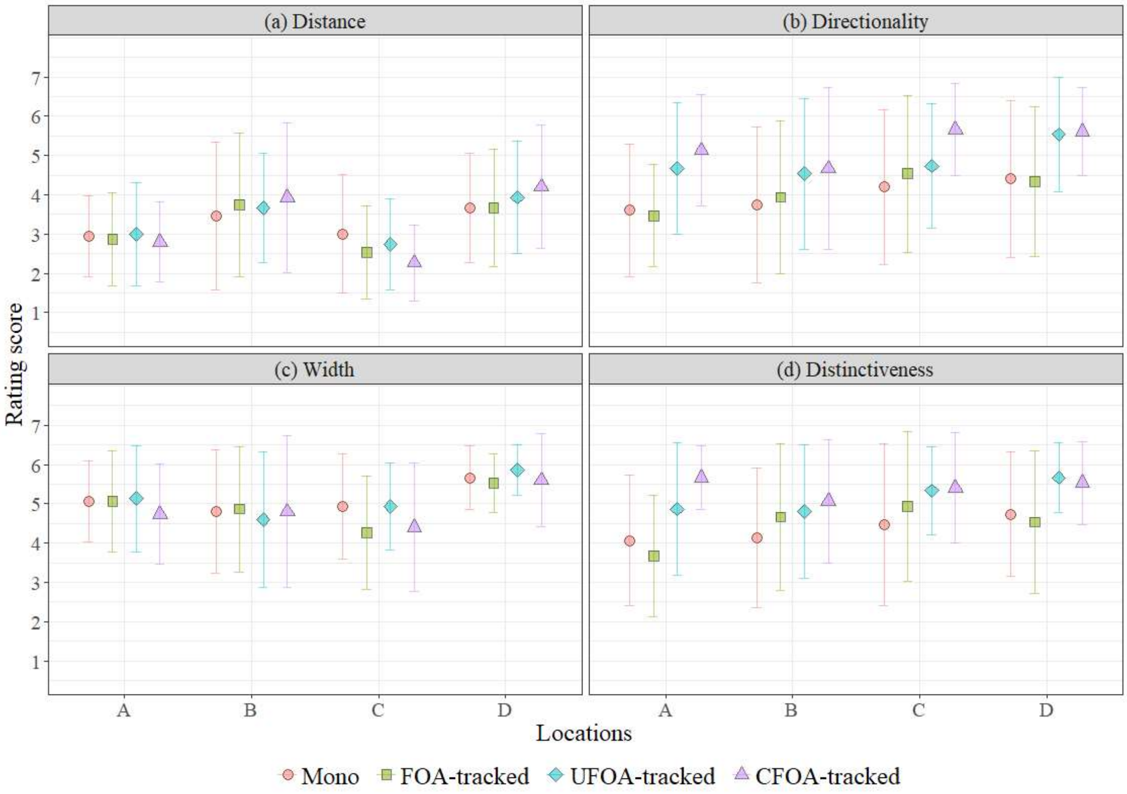

Figure 3. Mean rating scores of source-related spatial attributes (a) distance, (b) direction, (c) width, and (d) distinctiveness, as a function of the locations A to D, across the mono (●), FOA-tracked (■), UFOA-tracked (◆), and CFOA-tracked (▲) reproduction methods.

Specifically, post-hoc test showed that mean rating scores of the CFOA-tracked method was significantly higher than those of the FOA-tracked method regarding both directivity and distinctiveness of sound source ($p < 0.01$).

Figure 4 shows the mean rating scores of the five environment-related spatial attributes as a function of the locations across the four reproduction methods. Two-way RM MANOVA test was conducted to investigate the effects of reproduction method, location and their interaction on the set of attributes, namely, immersion, realism, externalization and timbral quality, overall listening experience.

The results showed that the effects of the reproduction method and interaction were not statistically significant for all the attributes indicating that the four reproduction methods provided similar environment-related spatial quality in this experiment. Meanwhile, the effects of location were significant on immersion ($F_{3,42} = 5.18$, $p = 0.004$), realism ($F_{3,42} = 4.04$, $p = 0.013$), timbral ($F_{3,42} = 4.58$, $p = 0.007$) quality and overall listening quality ($F_{3,42} = 5.36$, $p = 0.003$). Location B, the tranquil area beside a small lake, had significantly higher mean scores of immersion and realism than location C, an open area with a small fountain, at 0.05 significance level. Particularly, the participants rated the overall listening quality of location B is better than those of locations A and C ($p < 0.05$).

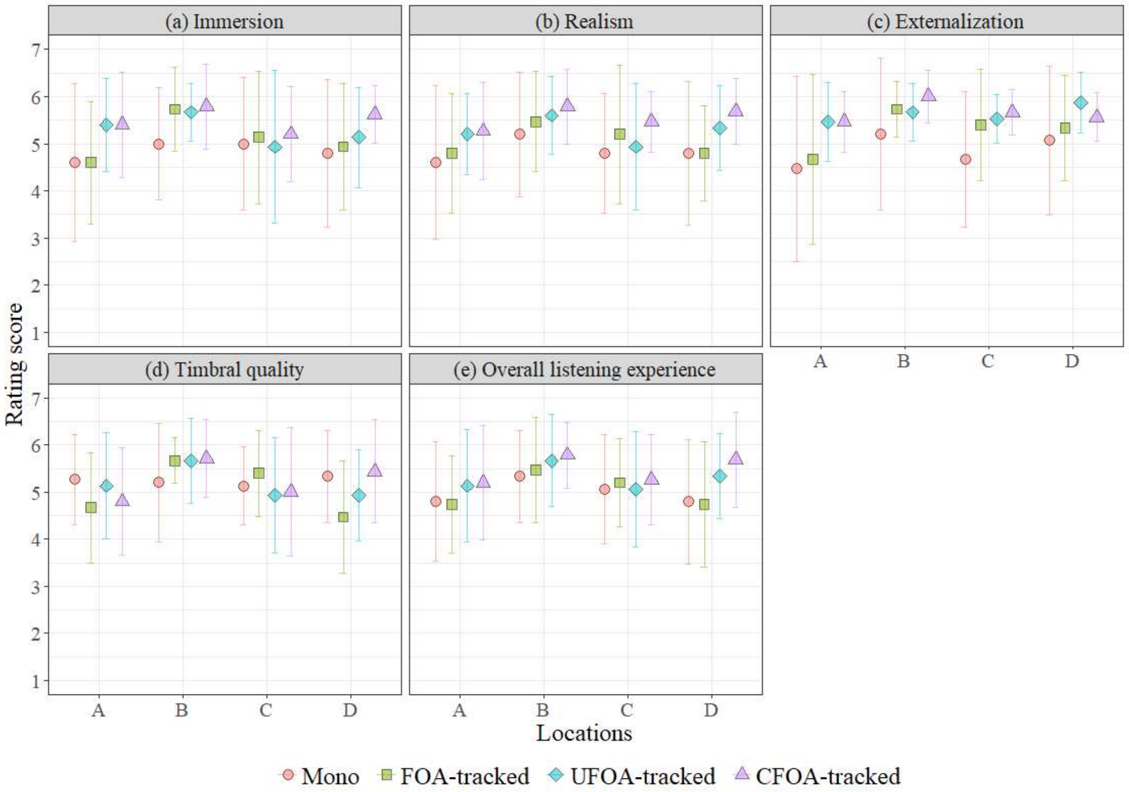

Figure 4. Mean rating scores of the environment-related spatial attributes (a) immersion, (b) realism, (c) externalisation, (d) timbral quality, and (e) overall listening quality, as a function of the locations A to D, and across the mono (●), FOA-tracked (■), UFOA-tracked (◆), and CFOA-tracked (▲) reproduction methods.

4. CONCLUDING REMARKS

The influence of parametric FOA decoding methods on the overall soundscape quality and spatial quality were assessed subjectively in a cinematic VR setup. Four recorded audio-visual urban scenes with varying dominating sounds and ambient characteristics were evaluated. In total, four audio reproduction methods were tested; a low-quality mono-channel anchor downmixed from FOA, linearly decoded FOA, parametrically up-mixed to third order ambisonics with HARPEX but linearly decoded FOA, and the parametrically-decoded FOA using the COMPASS framework. Except for the mono track, identified as "Mono", the rest of the reproduction methods are head-tracked binaural reproductions and were identified respectively as "FOA-tracked", "UFOA-tracked", and "CFOA-tracked".

The evaluation of the overall soundscape quality coincides with prior studies, where the spatial fidelity of the audio reproduction method does not affect the judgement of soundscape descriptors.

Of the source-related spatial attributes evaluated, there were significant differences in directionality and distinctiveness. In particular, CFOA-tracked was rated significantly higher than FOA-tracked for directionality and distinctiveness. This finding reinforces previous findings of improved sense of direction when assessing musical media or speech using the parametrically-decoded FOA. Although there were no significant differences in the judgement of source width and distance, further investigation with a larger sample size is required to determine if the methods indeed have similar accuracies in both source width and distance.

However, there were no significant differences detected in environment-related spatial attributes across all reproduction methods. These attributes are strongly related to the ambient component in the parametric method. Although the ratio of direct sound (i.e. subtle dereveberation) to the ambient component (i.e. increasing reverberation) could be adjusted in each frequency band, a balanced (i.e. 50-50) setting was employed for simplicity. In outdoor scenes, the low-frequency ambient components could be reduced but may in turn generate audible artefacts, especially with FOA tracks. These effects could be alleviated by increasing the degree of linearity in the low-frequencies. Nevertheless, the fine-tuning of these parameters are dependent on the desired reverberation and type of dominant sound sources in the sound scene.

In consideration of the cognitive load and the familiarity with the reproduction media, 15 expert listeners whom have exposure to audio research were recruited for the subjective experiments. However, for greater statistical power and better generality, a follow-up study should employ a larger sample size of naïve subjects. Moreover, other evaluation methods with lower cognitive load, such as the paired-comparison test, could be adopted albeit with greater time-complexity.

5. ACKNOWLEDGEMENTS

This material is based on research/work supported by the Singapore Ministry of National Development and National Research Foundation under L2 NIC Award No.: L2NICCFP2-546 2015-5. Joo Young Hong is currently supported by the 2017 Lee Kuan Yew Postdoctoral Fellowship (Grant no. 50025000).

6. REFERENCES

[1] International Organization for Standardization, Acoustics — Soundscape — Part

2: Data collection and reporting requirements, Geneva, Switzerland, 2018.

[2] 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.

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

[4] O. Rummukainen, J. Wang, Z. Li, T. Robotham, Z. Yan, Z. Li, X. Xie, Influence of visual content on the perceived audio quality in virtual reality, (2018).

[5] 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.

[6] L. Maffei, M. Masullo, F. Aletta, M. Di Gabriele, The influence of visual characteristics of barriers on railway noise perception, Sci. Total Environ. 445–446 (2013) 41–47. doi:10.1016/j.scitotenv.2012.12.025.

[7] 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.

[8] 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.

[9] R. Pheasant, K. Horoshenkov, G. Watts, B. Barrett, The Acoustic and Visual Factors Influencing the Construction of Tranquil Space in Urban and Rural Environments Tranquil Spaces-quiet Places?, J. Acoust. Soc. Am. 123 (2008) 1446–1457. doi:10.1121/1.2831735.

[10] 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.

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

[12] 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. doi:10.1016/j.buildenv.2018.12.004.

[13] 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.

[14] P. Coleman, A. Franck, J. Francombe, Q. Liu, T. de Campos, R.J. Hughes, D. Menzies, M.F.S. Galvez, Y. Tang, J. Woodcock, P.J.B. Jackson, F. Melchior, C. Pike, F.M. Fazi, T.J. Cox, A. Hilton, An Audio-Visual System for Object-Based Audio: From Recording to Listening, IEEE Trans. Multimed. 20 (2018) 1919–1931. doi:10.1109/TMM.2018.2794780.

[15] M. Poletti, UNIFIED DESCRIPTION OF AMBISONICS USING REAL AND COMPLEX SPHERICAL HARMONICS, in: Ambisonics Symp. 2009, Graz, 2009: pp. 1–10.

[16] 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.

[17] E. Bates, S. Dooney, M. Gorzel, H. O'Dwyer, L. Ferguson, F.M. Boland, Comparing Ambisonic Microphones: Part 1, in: Audio Eng. Soc. Conf. Sound F. Control, Audio Engineering Society, Guildford, UK, 2016.

[18] E. Bates, S. Dooney, M. Gorzel, H. O'Dwyer, L. Ferguson, F.M. Boland, Comparing Ambisonic Microphones: Part 2, in: 142th Audio Eng. Soc. Conv., Audio Engineering Society, Berlin, Germany, 2017.

[19] 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.

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

[21] A. Politis, S. Tervo, V. Pulkki, COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes, in: 2018 IEEE Int. Conf. Acoust. Speech Signal Process., IEEE, 2018: pp. 6802–6806. doi:10.1109/ICASSP.2018.8462608.

[22] A. Politis, J. Vilkamo, V. Pulkki, Sector-Based Parametric Sound Field Reproduction in the Spherical Harmonic Domain, IEEE J. Sel. Top. Signal Process. 9 (2015) 852–866. doi:10.1109/JSTSP.2015.2415762.

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

[24] Two Big Ears Ltd., Facebook 360 Spatial Workstation [Computer software], v3.0, 2017.

[25] Ö. Axelsson, How to Measure Soundscape Quality, in: Euronoise 2015, 2015: pp. 1477–1481. http://www.conforg.fr/euronoise2015/proceedings/data/articles/000067.pdf.

[26] 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.

[27] 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.

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

[29] 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.

[30] 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.