PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Simplification of reflection orders in virtual soundscapes through a subjective evaluation

Chunyang XU[1]; Jian KANG[2]

[1, 2] Institute for Environmental Design, The Bartlett, University College London, United Kingdom

ABSTRACT

Soundscape research stresses the interaction between human beings and sound environments. Virtual reality (VR) technologies provide a vivid approach to boost the virtual sound environment experience for people. However, the real-time VR audio simulation cannot process the complex acoustic conditions (e.g., boundary conditions) due to the high computational cost up to now. This study therefore aims to investigate the applicability of less reflection orders for reproduced sounds in virtual sound environment. To test the effect of reflection orders perceived by people, a series of subject evaluations was carried out based on audio-visual VR experience with different scale public squares reproduced in Unity to provide visual scenarios for the participants. The participants heard typical sounds (e.g., birdsong, clapping and fountain) in these squares, and gave their perception evaluation to these sounds. The results show that these reproduced sounds and typical building spaces can employ less reflection orders which render similar realism and immersion for participants in virtual sound environment.

Keywords: Soundscape, Virtual Reality, Sound Reflection

1. INTRODUCTION

The pioneering work to understand of how people perceive sound reflections that arrive within a short time after the direct sound was explored in the 1850s (1). In room and auditorium acoustics, sound reflection significantly affects how sounds are perceived by human beings. Clarity, reverberance (2) and speech transmission quality (3) are all involved with sound reflection in acoustic design. Both positive and negative effects of sound reflections were found on speech intelligibility in room acoustics (4). For outdoor sound environment, configurations of urban spaces, sound contexts, and propagation conditions are more complex. The multiple and moving sources can be additional variables in outdoor sound environment research. The numerical simulation and simplifications were examined in urban squares (5) and urban streets (6–9), and these studies parametrically explored sound field and sound propagation in different kinds of urban spaces.

The current soundscape research has been stressing more on how people perceive the sound environment. Sound reflection is one of the significant components affecting sound propagation in urban sound environments. To reproduce an ideal sound environment, the reflection order of sounds should be infinite. The VR technology provides a vivid approach to render virtual built environment synthesized with sounds. However, real-time auralization is becoming time-consuming with a high reflection order. Meanwhile, the human-environment interaction based on VR experience requires high synchronization for public participation. It is necessary to auralize sound environment efficiently and accurately with less computational cost. The previous research explored the simplifications with the audio-only condition (10). The interactive visual-aural stimuli should be paid more attention on sound environment evaluation when the current VR technologies have been boosted with higher visual resolution and more degrees of freedom.

This work therefore aims to investigate the relationship between subjective responses including immersion, realism and reverberance and different reflection orders under virtual sound environment. The effects of interactive visual stimulus, the area of squares, and the relationship between reverberance and reflection orders will be examined with multiple sounds and squares.

[1] chunyang.xu.18@ucl.ac.uk
[2] j.kang@ucl.ac.uk

2. METHODOLOGIES

2.1 Scene Selection and Visualization

To investigate the visual-audio coupling under VR experience in urban spaces, four squares were selected in London: the campus square behind Wilkins Building in University College London, Paternoster Square south to St. Paul's Cathedral, Cabot Square in Canary Wharf, and Granary Square next to Caravan King's Cross. The square behind the Wilkins Building in University College London was chosen, where the rectangular square was fully surrounded by building facades. Paternoster square is an urban square next to St Paul’s Cathedral re-developed in the 2000s, and it is enclosed by buildings with multiple facades and textures. Cabot square is located in Canary Wharf where it is one of the significant commercial estates in London. This square is surrounded by the buildings over 60 m. Granary square is a large open square in King’s Cross with the functions of relaxation, commerce and office.

Originally, there were fountains in Cabot Square and Granary Square. To investigate to the sounds of fountains in all squares, extra fountains were virtually placed at the UCL campus square and Paternoster Square. These sites crossed a wide geographical range in London with different functions, areas and built environments. In order to eliminate the undesired scattering and absorption, limited vegetation exists in these squares, and it was not reproduced during visualization. The visualization of four squares was through the modeling software (SketchUp Pro 2018). These four models consist of 2,733, 4,317, 2,379 and 4,143 edges. The photographs and reproduced scenes of four squares are shown in Table 1. The areas of these four squares are 900 m^2, 1700 m^2, 6000 m^2, 7000 m^2 in order.

Table 1 – Real sites and reproduced scenes

Sites	Photographs	Reproduced Scenes
UCL Campus		
Paternoster Square		
Granary Square		
Catbot Square		

2.2 Sound Selection and Auralization

To investigate the acoustic behaviors of different sounds in these squares, several typical sounds were chosen including birdsong, clapping and fountain. These three types of sounds have different acoustic characterizations and soundscape contexts. The sample rate of the three sounds is 44100 Hz, and the depth is 24 bit. These three sounds represent different kinds of time-frequency characterizations. Birdsong has the high frequency contents of specific pitches. Birdsong is usually considered as a natural sound with a positive effect on urban sound environment. Clapping is discrete in time domain analogue to an impulse sound, and it is also a sound generated by human activities. Fountain sound is continuous in time domain, and fountain is a common landscape component in urban public spaces generating water sound.

Odeon (9.2 Auditorium) was used to characterize the acoustic performances of these public squares. Meanwhile, the process of auralization was also based on the impulse response generated from Odeon. The boundary absorption and diffusion conditions were assigned with different parameters matched the real sites. The built-in spatial audio plugin in Unity implements the real-time tracing in an assumed box to achieve real-time auralization with the participant's spatial displacement. The hybrid method in Odeon could deal with the complex boundary conditions and generate the point-to-point impulse responses with acceptable accuracy.

The previous study on simplification through subjective tests validated the reflection orders of 5, 20, and 50 with the aural stimuli only (10). In this work, four reflection orders, i.e., 1, 5, 20 and 1000, were chosen, and these four different orders were first examined in Odeon. The distance between the sound source and the receiver in the four squares was set to 8 m, and it was a reasonable distance when a participant stood in these squares to observe the events.

2.3 Reproduction and VR Synthesis

The game engine (Unity) was used to synthesize visualization and auralization. The lighting condition was set to a reasonable solar zenith angle and illuminance according to their geometrical locations. To synchronize the virtual visual-audio environment, a particle system of water splash was attached to the fountain, and the animation of clapping was given to the characters on the square. Birds were not rendered in the VR design.

The first-order ambisonic impulse responses were convolved with three dry sounds. The filtered sounds were decoded and spatialized in Unity according to the headset motion and location. The VR environment was streamed through SteamVR. The high performance graphics card (GeForce GTX 1070) and Central Processing Unit (Intel Core i7-8700K) were used to guarantee the rendering quality. HTC VIVE was used to provide VR experience. The VR headset was connected with a headphone. The participants could turn around their heads, but they cannot have the spatial displacement due to the first-order ambisonics.

2.4 Subjective Evaluation

The subjective test was divided into two stages. In stage I, the participants were informed to hear these sounds in the UCL campus square without the VR headset. Since all participants were familiar with the UCL campus, they could have the corresponding spatial impression on this square. In stage II, the participants heard these sounds with the VR headset under four reproduced scenes.

Three perceived indicators were selected including reverberance, immersion and realism. Reverberance was categorized into a technically perceived indicator in the previous study (11). For different functions of interior spaces (e.g., a lecture room and a concert hall), acoustic performances are required to render totally different reverberance (12). This discrepancy has been studied in room acoustics for a long time. Reverberance is more difficult to be perceived under continuous sounds, compared with sounds of stopped reverberance. Immersion and realism were considered as reproduced indicators (13). The three adjectives 'reverberant', 'immersive' and 'real' were rephrased into three indicators: reverberance, immersion and realism. The Likert scale was utilized to describe the rating scale with the description from 'not at all', 'slightly', 'moderately', 'very' to 'extremely'. The questions used in the both stages are listed below:

(1) How reverberant is this sound environment?
(2) How immersive is this sound environment?
(3) How real is this sound?

For each type of sounds in a scene, a reference sound was given to the participants first. The reference sound was convolved with the impulse response of the reflection order of 1000. Then, each sound lasted 10 seconds, and these sounds with different reflection orders were played one by one. The sequence of these sounds was randomized in each reproduced scene. In addition, 3D Graphical User Interface (GUI) synthesized in VR designs was shown in front of the participants when they were doing the evaluation. Whilst it was possible to wear the VR headset, the participants could give their subjective perception through a five-button controller. Four scenes were conducted with the same procedure. Figure 1 shows the GUI in the reproduced scene and the subjective test by a participant.

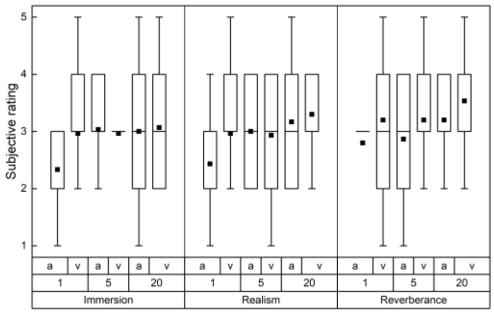
Thirty participants with normal hearing and vision took part in the subjective evaluation. The participants were not informed by which reflection orders were applied in each sound during the evaluation. They all had the basic understanding of perceived indicators used in the formal evaluation.

Figure 1 – GUI in Unity 3D and subjective test through the VR headset

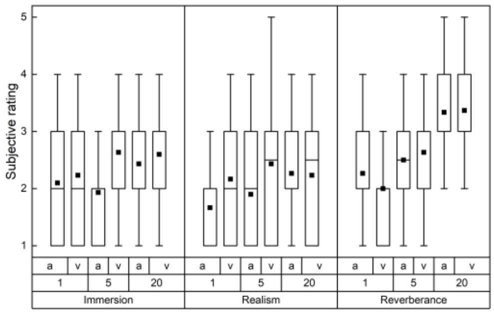
3. RESULTS

3.1 Effect of VR experience

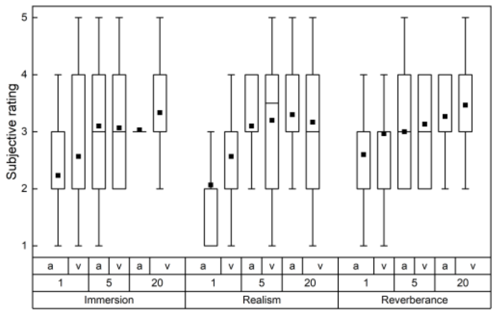
The two-stage subjective test was carried in the UCL campus square. Figure 2 shows the comparison between audio-only and VR experience conditions under different reflection orders. Three types of sounds present the different subjective ratings. For birdsong, the rating difference between the two approaches is significant under the reflection order of 1 for all three perceived indicators. The rating of clapping is lower than the other two sounds in immersion and realism. Meanwhile, the variation of reverberance is the most distinct. When the reflection order is 1, the rating difference between two evaluation approaches was found to be significant. Compared with the other two sounds, the clapping sound presents relatively low ratings on immersion and realism for both approaches, and the variation of reverberance is the most distinct. The clapping sound is discrete in time domain, and the stopped reverberance is easily perceived by the participants. The two approaches were found to have a significant difference on immersion and realism under the reflection order of 1.



(a)



(b)



(c)

Figure 2 – Comparison between audio-only [a] and VR experience [v] conditions under different reflection orders (1, 5 and 20). (a) birdsong, (b) clapping, (c) fountain.

When the reflection order was set to 1, VR experience during the subjective test can increase the immersion and realism of sounds. All participants were required to evaluate sound impression during the subjective test rather than VR impression. Thus, the rating gain on immersion and realism of sounds was generated from the interactive environment.

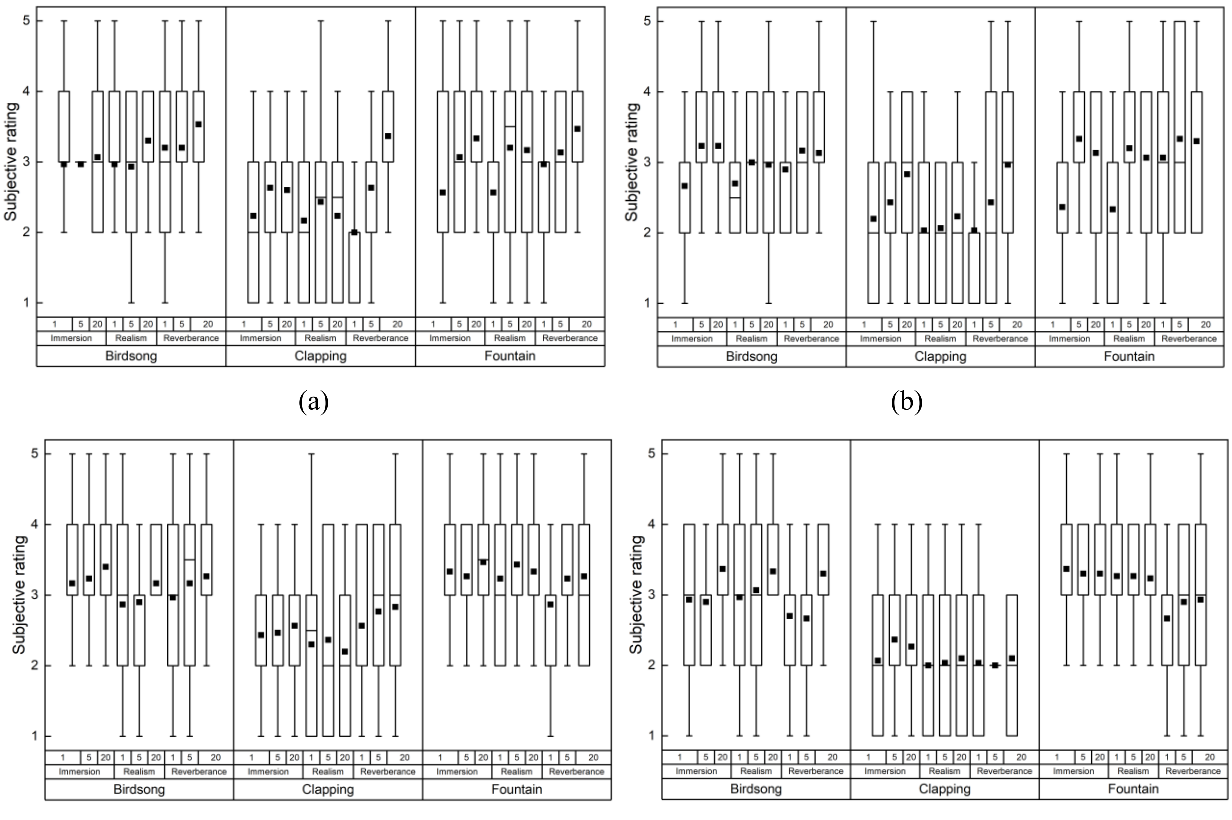
An unusual point occurs in clapping sound. When the reflection order is only 1, the mean rating of VR experience is lower than the rating of audio-only condition. However, the mean rating difference between two approaches is not statistically significant under a reflection order of 1. Compared with other two sounds, clapping is the only discrete sound in time domain, and for each clapping, the signal analogue to an impulse response includes all acoustic information of a space. Clapping sound with stopped reverberance therefore presents a rating range wider than the other two sounds.

The comparison results between audio-only and VR experience conditions reveal that both evaluation approaches present the similar results on three indicators when the reflection order reaches 20. The VR experience can improve subjective ratings of immersion and realism when the reflection order is set to 1.

3.2 Effect of sounds and urban spaces

In stage II of the subjective test, the results in the four squares are shown in Figure 3. The four squares present different results on three types of sounds and three indicators involved with different reflection orders.

In UCL campus square, the mean ratings of the three sounds present a significant increase when the reflection order is changed from 5 to 20 for reverberance. In Paternoster square, except clapping sound, a reflection order of 5 is acceptable to render birdsong and fountain with similar results compared with the reflection order of 20. For Granary square, the mean ratings of immersion and realism for the three sounds are approximate within a narrow range. In Cabot square, the birdsong was found to have a significant variation when the reflection order is changed from 5 to 20. The clapping and fountain sounds show no notable tendency on different reflection orders for three perceived indicators. It is feasible to use only a reflection order of 1 to render these two sounds in Cabot square.



(a) (b)

(c) (d)

Figure 3 – Subjective ratings under different reflection orders (1, 5 and 20) among four squares. (a) UCL campus square, (b) Paternoster square, (c) Granary square, (d) Cabot square.

For outdoor sound environment, the large square area (e.g., Cabot square and Granary square) reduces the rating difference among multiple reflection orders. The sound is more likely to be dissipated in a large space, and the subjective result is also in accordance with the sound propagation law. In addition, reverberance of clapping sound significantly depends on the acoustic behaviors of urban spaces. It is feasible to reduce the reflection order when the space is large enough. For a small public square, a reflection order of 5 is acceptable to reproduce fountain sound.

4. CONCLUSIONS

This project examined subjective responses of immersion, realism and reverberance on multiple sounds with different reflection orders through a VR subjective test. The results of the subjective test revealed that:

(1) Compared with the audio-only results, the VR experience was found to increase realism and immersion of sounds when the reflection order was set to a relatively low range (e.g., 1 and 5).

(2) When the public square was large enough (e.g., >5000 m^2), the reflection order was found to have a weak impact on immersion, realism and reverberance under VR experience for various sounds including birdsong, clapping and fountain. When the square was enclosed within a relatively small area (e.g., <2000 m^2), a reflection order of 5 or a higher order was required, and the VR experience still cannot fill the gap of realism and immersion generated by the reflection order from 1 to 20.

(3) The sound of stopped reverberance (e.g., clapping) significantly depended on the reflection order, and it needed to be carefully auralized during audio processing and reproduction with the consideration of spatial characteristics and acoustic behaviors.

Overall, it is acceptable to employ less reflection order during auralization to render sounds under VR experience with similar realism and immersion. The results on perceived indicators could make contribution to fast auralization with less reflection order and reasonable accuracy through VR experience.

ACKNOWLEDGEMENTS

The authors are grateful to Huan Tong and Tingting Yang for their help in modeling built environment. The authors also thank Andrew Mitchell and Dr Francesco Aletta for valuable comments. The project was supported by the European Research Council (ERC) Advanced Grant (No. 740696).

REFERENCES

1. Shankland RS. Architectural acoustics in America to 1930. J Acoust Soc Am. 1977;61(2):250–4.
2. Griesinger D. Room impression, reverberance, and warmth in rooms and halls. In: Audio Engineering Society Convention 93. 1992.
3. Warzybok A, Rennies J, Brand T, Doclo S, Kollmeier B. Effects of spatial and temporal integration of a single early reflection on speech intelligibility. J Acoust Soc Am. 2013;133(1):269–82.
4. Bradley JS, Sato H, Picard M. On the importance of early reflections for speech in rooms. J Acoust Soc Am. 2003;113(6):3233–44.
5. Kang J. Numerical modeling of the sound fields in urban squares. J Acoust Soc Am. 2005;117(6):3695–706.
6. Kang J. Numerical modelling of the sound fields in urban streets with diffusely reflecting boundaries. J Sound Vib. 2002;258(5):793–813.
7. Myncke H, Cops A. Reverberation in Town Streets. Acta Acust united with Acust. 1978;40(2):115–9.
8. Kang J. Sound Propagation in Interconnected Urban Streets: A Parametric Study. Environ Plan B Plan Des [Internet]. 2001 Apr 30 [cited 2018 Nov 16];28(2):281–94. Available from: http://journals.sagepub.com/doi/10.1068/b2680
9. Kang J. Urban sound environment. CRC Press; 2006.
10. Meng Y, Kang J, Huerta R, Zhang M. Soundscape animation of urban open spaces in an interactive virtual environment. In: Proceedings of the 12th International Congress of Sound and Vibration, Lisbon, Portugal. 2005.
11. Xu C, Kang J. Soundscape evaluation: binaural or monaural? J Acoust Soc Am. 2019;145(5). [Accepted]
12. Kuttruff H. Room acoustics. Crc Press; 2016.
13. Hong JY, Lam B, Ong Z-T, Ooi K, Gan W-S, Kang J, et al. Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications. Build Environ. 2019;149:1–14.