Does Spatial Attribute between 2D and 3D Virtual Spaces make Different User Immersion of Audio-Visual Events?

Sehwan Lee
Seoul Media Institute of Technology
99, Hwagokro-61gil, Gangeo-gu
Seoul, 0750, South Korea
+82-2-6393-3240
nhistory@naver.com

Ju-Hwan Lee*
Seoul Media Institute of Technology
99, Hwagokro-61gil, Gangeo-gu
Seoul, 0750, South Korea
+82-2-6393-3240
jhlee@smit.ac.kr

ABSTRACT
In the real world, people use various sensations to perceive the significant information and judge the critical situations and circumstances from the outside, but in the virtual environment the limited sensory information is only provided. What immersive experience the user experiences in the virtual environment depends on how the 3D visual information and the auditory information are combined and provided to the user. Most of the quality factors that determine the immersion feeling of the auditory virtual environment from the viewpoint of the user are the items which are important in the sound design for the conventional two dimensional images. In this study, we measured and evaluated how the relative loudness of ambience sound affecting the spatial feeling and reverberation of auditory virtual environment quality factors are related to user's immersion feeling when they are adjusted to different forms in two-dimensional and three-dimensional space. The experiment was conducted to determine the amount of ambience sound volume which is most suitable for moving objects and space while observing the same visual information with different two- and three-dimensional displays. As a result, when a three-dimensional event is reproduced in a three-dimensional representation space, an ambience sound with more ambience sound level than a two-dimensional expression space can be provided to the user.

CCS Concepts
• Human-centered computing→Empirical studies in interaction design • Human-centered computing→Virtual reality.

Keywords
Auditory Virtual Environment; Immersion; Ambient Sound; Sound Design.

ICCAE '17, February 18-21, 2017, Sydney, Australia

ACM 978-1-4503-4809-6/17/02…$15.00

DOI: http://dx.doi.org/10.1145/3057039.3057092

1. INTRODUCTION
In our daily life, human use the several sensations to perceive any meaningful information from the outside world and judge the significant situations and circumstances. As media technology evolves, attempts to deliver narrative and emotion intended to users by using information perceived by the visual and auditory sensations have been extensively expanded in various contents. In recent years, with the expansion of content development in the virtual reality environment, related equipment and technology markets are growing beyond the introduction stage [1, 2].

Since the first talking film in 1927, many attempts have been made about how to combine visual and auditory information on a two-dimensional screen [3]. However, as to what kind of sound design is required in the emerging virtual environment contents. Unlike the images displayed on the existing two-dimensional screen, the completeness of the provided content is determined by the virtual environment provided through the three-dimensional space according to 'how much the user feels in the virtual environment created by the user (presence) to be'. Whether a user experiences immersion or disparateness in a virtual environment depends on how 3D visual and auditory information are appropriately combined and provided to the user [4-6].

The user perspective in the auditory virtual environments of the immersion decisions that quality factor include 1) Location Accuracy, 2) Timbre, 3) Loudness Balance, 4) Auditory Spaciousness, 5) Reverberation, 6) Dynamic Accuracy, and 7) Artefacts. Most of the above elements are important items in sound design for existing 2D image. In addition, there has been a research to classify the importance of these quality factors, but since it was a questionnaire survey, there was a limit to quantitatively grasp the degree of change in auditory information required according to the change of visual information [7].

In this study, when the relative loudness difference of the ambient sound which affects the spatial sensitivity and reverberation of the auditory virtual environment quality factor is adjusted to different patterns in the two-dimensional and three-dimensional expressive virtual spaces of audio-visual events, it is measured how they relate to user immersion [8-10]. When the user measures the volume difference between an object and ambient sound optimized for user immersion in audio-visual events in different expressive spaces of 2D and 3D, the other elements in the auditory virtual environment are also changed as the visual environment. It can be judged that it is affected. This suggests that we have to approach the sound design in 3D virtual space in a way different from the sound design in the conventional 2D screen display [11, 12]. Therefore, there is a need to revise the standards for contents production and sound design in the virtual environment in the future [13].

One might wonder how best to deliver effective and efficient multisensory experience for virtual reality or augmented reality devices, such games, simulations, and interactive virtual contents. In the present study, we specifically investigated whether the perception of the quality of auditory stimuli delivered by the virtual spaces involving a user's visual spatial attribute (i.e., when a person visually experiences with 2D or 3D display settings) could be modulated simply by changing the quality of visual stimuli that a participant watched at the same time.

Such a finding would clearly have potentially important implications for the design of virtual reality environment feedback for display devices in the years to come. On the basis of the findings of the study reported here, we hoped to gain further insight into the nature of any multisensory interactions taking place in the practical situation of an interface operator's use of a virtual reality device. We also suggest a guideline for the future presentation of harmonized sensory information via different sensory modalities in order to increase the likelihood of effective multisensory contents and user interface design.

2. EXPERIMENTS

2.1 Participants

Twenty graduate students and adults (males, n = 11; females, n = 9; mean age = 29 years, SD = 2.1) volunteered for this experiment that required approximately 15 min to complete. All participants reported they are being in good health with no physical handicaps. Their visual acuities were normal or corrected-to-normal. All participants reported normal hearing and normal tactile sensitivity, and they were all right-handed. All of the participants were naïve as to the purpose of the study, and all gave their informed consent prior to taking part in the experiment.

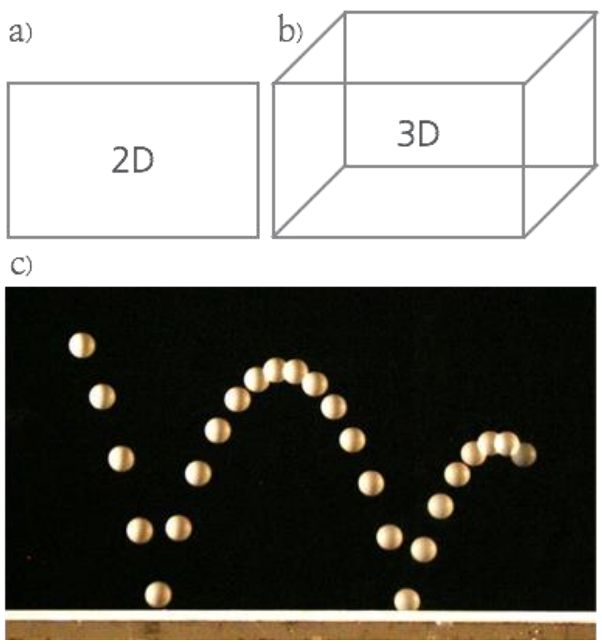

Figure 1. Examples of 2D(a) and 3D(b) expressional virtual spaces of the audio-visual events and the bouncing ball event(c) in the spaces.

Figure 1. Examples of 2D(a) and 3D(b) expressional virtual spaces of the audio-visual events and the bouncing ball event(c) in the spaces.

2.2 Experimental Settings

Experiments were conducted to determine the amount of ambience sound volume that is most appropriate for moving objects in a certain virtual space while observing the same visual information in different 2D and 3D scenes (Figure 1). In order to measure the relevance according to spatial changes of the visual medium experienced by the experiment participants, the same event was played differently on the 2D screen and the 3D virtual space, and the object sound was set to the same auditory image volume.

The experiment takes the participants to listen to the moving objects in audio-visual events such as the 15-second bouncing ball, so that the participants can use the faders of the ambience soundtrack to determine the amount of volume that is best suited to each audio-visual event. The two-dimensional display was presented as a 27-inch monitor, and the 3D display was presented through virtual reality devices (HTC VIVE & Oculus Rift)(Fig. 2). The virtual space adopted in this experiment was the inside of a modern house fully furnished and included a complete environment system from the Unity Asset Store (link: https://www.assetstore.unity3d.com/kr/#!/content/73674, ArchVizPro Interior Vol.4; see Figure 2). Each Participant performed a total of 4 trials, and recorded the change of the sound of the determined fader knob position. The trial order was count-balanced.

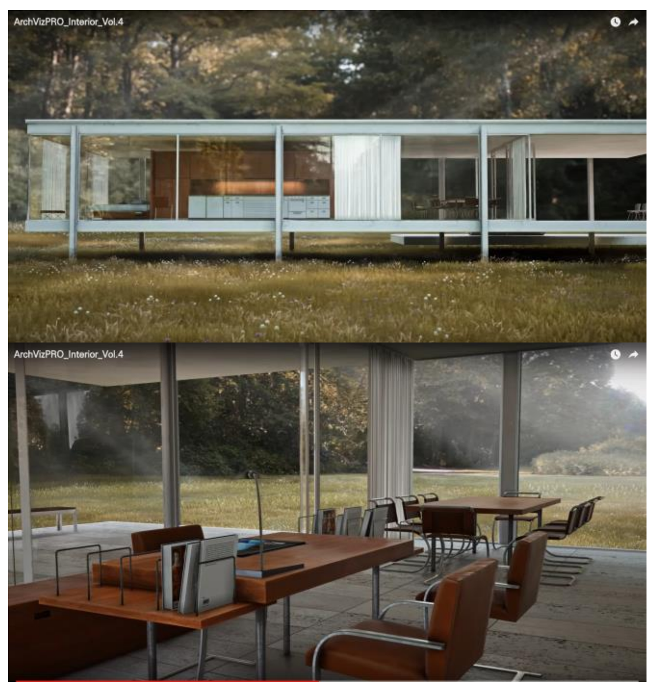

Figure 2. Screen-shots of the virtual space adopted in this experiment with 2D and 3D displays.

2.3 Apparatus and Stimuli

In the experiment, two auditory stimuli were a bouncing ball sound (target) and ambient sound (ambience) equivalent to the visual environment as Figure 2. While the target was a rigid plastic ping-pong ball on the wooden table, the ambient sound was a mixed sound including birds and cricket chirping, wind, and

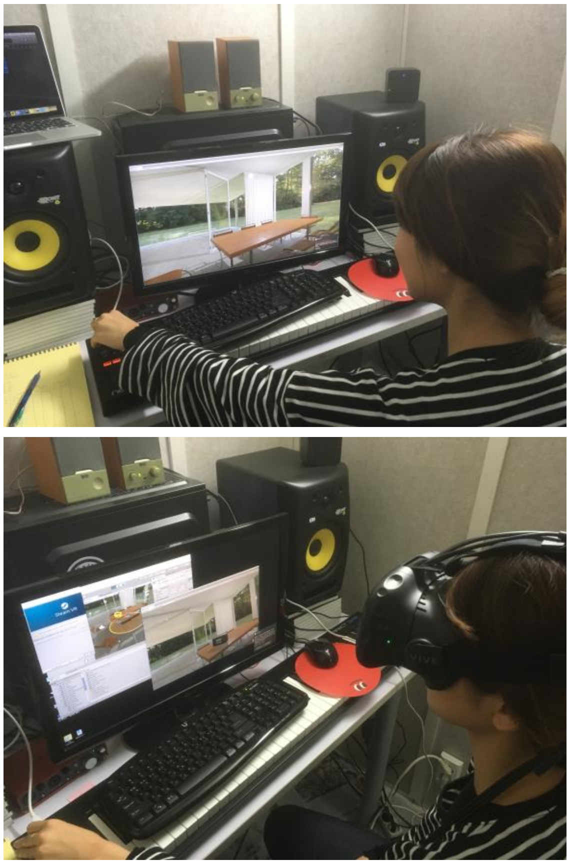

Figure 3. Audio-visual events' experience settings in 2D(above) and 3D(below) virtual spaces.

a typical sound scene of the rural landscape. The level of target sound was -32dB LUFS and ambient sound was -41dB LUFS.

2.4 Procedure

The experiment takes the participants to listen to the moving objects in audio-visual events such as the 15-second bouncing ball and 30-second radio speech clip, so that the participants can use the faders of the ambience soundtrack to determine the amount of volume that is best suited to each audio-visual event. The two-dimensional event was presented as a 27-inch monitor, and the 3D event was presented through a virtual reality device (HTC VIVE & Oculus Rift)(see Figure 3). Participants performed a total of 4 trials for each condition repeated 2 times, and recorded the change of the sound of the determined fader knob position. The trial order was count-balanced.

3. RESULTS AND ANALYSIS

Experimental results are compared with the *t*-test analysis of the amount of volume differences between object and ambient sounds optimized for the user's immersion feeling caused by the visual information of the two-dimensional screen display and the visual information of the three-dimensional virtual reality space.

As can be seen in Figure 4, there is a statistically significant difference between the volume changes of the ambience sound measured as the expressive virtual space changes, and the volume difference of the ambience sound measured in the three-dimensional space is larger than the measurement area in the two-dimensional representation of the audio-visual events ($t(19)$ = 5.492 $p<.001$). This is because the spatial quality of the visual information given to the user in the three-dimensional virtual space may be more than that given in the two-dimensional virtual space, thus supporting the hypothesis that more ambience sound is required in the presented condition.

In the case of an audio-visual event maximizing the immersion feeling of a 3D virtual space, it is considered that the sound design in the 3D space is different from the sound design in the two dimensional screen. In 3D virtual space, ambience sound should be louder than 2D space to make the user immersive as the spatial attribute.

4. CONCLUSIONS

In this study, we tried to evaluate the difference between the object and ambience sound volumes which can experience deeper immersion when the same three-dimensional audio-visual event is presented in different expressive spaces by 2D monitor and 3D virtual reality device. We also analyzed whether this effect leads to a larger ambience sound volume in either 2D or 3D depending on the expressive space. Actually, the user needs the volume of different ambience sound as the dimension of the expressive space changes. In addition, when presenting the same 3D audio-visual event in a three-dimensional virtual space rather than presenting it in a two-dimensional virtual space, a larger ambience sound volume appeared to naturally fit into 3D events and to provide a sense of immersion. These results suggest that a criterion different from the two-dimensional expressive space is needed for the volume balance between the object sound and the ambience sound in that the currently produced virtual reality content aims at the user's immersion feeling it suggests.

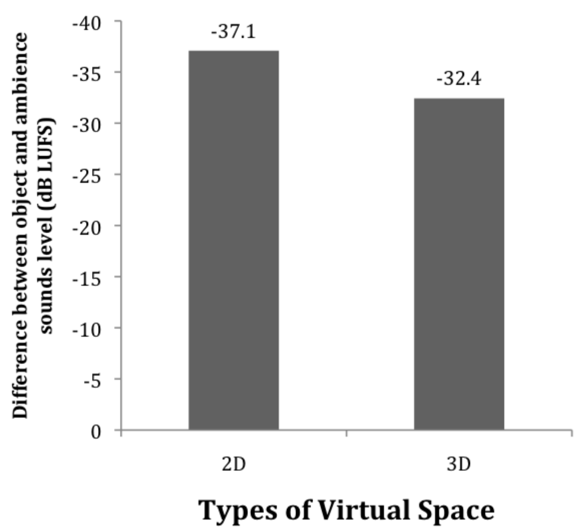

Figure 4. Results of the sound level differences (dB LUFS) between object and ambient sounds of the audio-visual events according to the 2D and 3D virtual spaces; the results mean that ambient sound in 3D virtual space needs to be increased near to the object sound volume than 2D space.

According to the results of the experiment conducted in this study, when three-dimensional audio-visual events are presented in the three-dimensional virtual space, it is found that ambience sound can be provided to the user even with more ambience sound than the two-dimensional expressive space. So far, there has been little research on what kind of sound design change is needed between the two-dimensional and the three-dimensional expressive virtual space. Since the amount and quality of the visual information given to the user in the three-dimensional virtual space such as the virtual reality are greatly increased unlike the two-dimensional expressive space, we can provide guidelines for creating content suitable for a virtual environment by considering complementary relationships between the visual and auditory information.

To sum up, based on the findings reported here, we can suggest the following guidelines for virtual reality designers: 1) Matching the quality of stimuli presented to visual and auditory modalities should help to avoid interfering with the users' immersive perception, while at the same time providing auditory stimulation as a harmonized and complementary means of information transfer; 2) Because the user's immersion and presence feelings that occur when the spatial quality of visual and auditory stimulation are well-matched (or congruent) increase as the attention and understanding of a situation (or task) increases, the addition of congruent sound feedback will likely enhance the users' perception in the virtual space experience, therefore the virtual reality content designer can use auditory stimulation in order to manipulate the perceived quality of audio-visual events.

5. ACKNOWLEDGMENTS

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A5A2A01025875). Corresponding author(*) is Prof. Ju-Hwan Lee.

6. REFERENCES

[1] De Kort, Y.A., Ijsselsteijn, W.A., Kooijman, J., & Schuurmans, Y. 2003. Virtual laboratories: Comparability of real and virtual environments for environmental psychology. *Presence: Teleoperators and virtual environments*, 12(4), 360-373.

[2] Drettakis, G., Roussou, M., Reche, A., & Tsingos, N. 2007. Design and evaluation of a real-world virtual environment for architecture and urban planning. *Presence: Teleoperators and Virtual Environments*, 16(3), 318-332

[3] Lee, D-.H. 2013. Narrative Function of Sound Design in Films. *The Journal of the Korea Contents Association* 13, 12, 626-637.

[4] Altinsoy, M.E. 2012. The quality of auditory-tactile virtual environments. *Journal of the Audio Engineering Society*, 60(1/2), 38-46.

[5] Dinh, H. Q., Walker, N., Hodges, L. F., Song, C., & Kobayashi, A. 1999. Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In Virtual Reality, 1999. In *Proceedings., IEEE* (222-228). IEEE.

[6] Nechvatal, J. 2010. *Immersive ideals / critical distances: study of the affinity between artistic ideologies in virtual Reality and previous immersive idioms*. Lambert Academic Publishing AG & Co KG, 48-60.

[7] Silzle, A. 2007. Generation of quality taxonomies for auditory virtual environments by means of systematic expert survey. *FORTSCHRITTE DER AKUSTIK*, 33(2), 869.

[8] Slater, M., Linakis, V., Usoh, M., Kooper, R., & Street, G. 1996. Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess. In *ACM virtual reality software and technology (VRST)* (Vol. 163, p. 72). New York, NY: ACM Press.

[9] Lampton, D.R., Kolasinski, E.M., Knerr, B.W., Bliss, J.P., Bailey, J.H., & Witmer, B.G. 1994. Side effects and aftereffects of immersion in virtual environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 38, No. 18, 1154-1157). Sage CA: Los Angeles, CA: SAGE Publications.

[10] Slater, M., & Usoh, M. 1993. Presence in immersive virtual environments. In *Virtual Reality Annual International Symposium, 1993*., 1993 IEEE (90-96). IEEE.

[11] Begault, D.R., & Trejo, L.J. 2000. *3-D sound for virtual reality and multimedia*. Technical Report. NASA Ames Research Center; Moffett Field, CA United States.

[12] Gabbard, J.L., Hix, D., & Swan, J.E. 1999. User-centered design and evaluation of virtual environments. *IEEE computer Graphics and Applications*, 19(6), 51-59.

[13] Weibel, P. 2003. Expanded cinema, vide o and virtual environments. *Future Cinema: The Cinematic Imaginary after Film*. Cambridge: MIT Press. Sandra Álvaro.