Proceedings of The National Conference
On Undergraduate Research (NCUR) 2020
Montana State University, Bozeman MT
March 26-28, 2020

Examining the Relationship Between the Presence of Audio and Subjective User Experience in the Context of Virtual Reality

John Levee
Department of Creative Arts
Stetson University
421 North Woodland Boulevard
DeLand, Florida 32723 USA

Faculty Advisor: Dr. Nathan Wolek

Abstract

Examining the Relationship Between the Presence of Audio and Subjective User Experience in the Context of Virtual Reality aims to establish correlation between the presence of spatialized audio during a VR experience and improved user experience in terms of increased immersion, realism, enjoyment, and perceived quality of the visual component. Through use of the HTC Vive VR headset, twenty-five participants experienced four variations of a forest-themed environment, two with accompanying audio from field recordings of wind, birds, and other natural sounds. The four variations are as follows:

1. High quality visual render (with audio)
2. High quality visual render (silent)
3. Low quality visual render (with audio)
4. Low quality visual render (silent)

In general, it was found that the presence of audio caused increased feelings of immersion, realism, and enjoyment in the user; however, results were inconclusive regarding whether the presence of audio caused an increase in perceived visual quality.

Keywords: Virtual Reality, Spatialized Audio, Subjective User Experience

1. Hypothesis

It is hypothesized that in this study, the presence of audio will have a profound effect on subjective user experience in the context of virtual reality media. Specifically, the presence of audio will result in increased feelings of immersion, realism, and enjoyment in the subject. Additionally, the presence of audio will result in higher perceived visual quality, meaning that when sound is present, the graphics will seem to be more visually appealing (due to the increased feelings of immersion and realism mentioned previously). Finally, it is possible that visual quality will have a similar effect of perceived audio quality, and that as visual quality increases, perceived audio quality will also be higher.

2. Literature Review

A number of previous studies exist regarding the effects of audio quality on visual quality, but as of the writing of this paper, the author was only able to find such studies dealing with traditional film media. No such studies were found to deal primarily within the realm of virtual reality. These studies have conflicting results as to the degree at which

audio affects perception of visual quality. In her 2015 thesis, Sara Langvik, asked subjects to view audiovisual clips that were identical apart from the quality of the accompanying audio, and the subjects gave subjective ratings of visual and audio quality for these clips. Although the video quality of these clips was never changed, subjects in her study reported that the perceived visual quality of the videos increased as the perceived audio quality increased.[1] A 2011 study by Beerends and De Caluwe, "The influence of video quality on perceived audio quality and vice versa", examined similar questions but had differing results. They found that although both video and audio quality had an effect on one another, increased visual quality had a far greater effect on perceived audio quality than vice versa, and that "the video quality dominates the overall perceived audiovisual quality." [2] These studies differ in their conclusions about the relative importance of visual and audio quality in an audiovisual work, but neither address the topic of virtual reality media.

Other studies involving the role of sound in audiovisual media have also impacted the work described in this paper. A 2018 study found that when compared to videos using stereo sound, students had better comprehension due to increased immersion when watching the same educational films utilizing 3D surround sound.[3] This helped lead to the decision to use spatialized audio diffused through a 5.1 system. A 2015 study found that adjusting certain mixing parameters of the sound design in a film sequence had a "statistically significant effect on plausibility […] to enhance the perception of realism".[4] This lends validity to this paper's hypothesis that the presence of audio can affect subjective user experience, including perceived realism. Finally, a 2015 article by Chattopadhyay found that in Indian films, "ambient sound practice enhances the spatial and atmospheric sensations"[5], yet again lending validity to the claim that the presence of audio, specifically ambient sound such as the field recordings used in this paper, influence the subjective user experience of the audiovisual media. As with the first two studies discussed, all three of these sources dealt primarily with traditional film media and did not include investigation into virtual reality media.

There is, however, one recent study which examines the relationship between visuals and audio quality in the context of virtual reality. In a 2018 study, it was found that "the visual content has a statistically significant effect on the perceived audio quality" in virtual reality media.[6] However, this study does not examine the reverse question, which is whether the audio has any effect of the perceived quality of the visual component.

3. Methodology

1. Create a virtual reality forest environment using Unreal Engine 4, without the addition of audio.
2. The subject experiences each of 4 VR environments (described above) in a random order, pausing between experiences to fill out a survey.
3. Create a five-minute loop of forest-themed ambient noises, spatialized for playback on a 5.1 surround sound speaker array, utilizing SpatGRIS and Logic Pro X.[7]
    1. This audio utilizes field recordings from the BBC Sound Effects library, as used for research and educational purposes in accordance with their licensing agreement.[8]
4. Create an identical duplicate of the VR environment, but with the addition of audio.
5. For both the environment with audio and the environment without, save and export two versions.
    1. One version of each utilizes advanced graphics and particle effects (such as ambient occlusion, motion blur, and bloom), uses the highest quality lighting build, and is packaged uses the highest quality settings.
    2. The other version does not utilize the abovementioned graphics and particle effects, uses the lowest quality lighting build, and is packaged with the lowest quality settings.
        i. This results in four seemingly identical VR environments:
            1. High quality visual render (with audio)
            2. High quality visual render (silent)
            3. Low quality visual render (with audio)
            4. Low quality visual render (silent)
6. The subject experiences each of 4 VR environments (described above) in a random order, pausing between experiences to fill out a survey.
    1. Survey Questions:
        i. Please respond to these statements with a number from 1 to 10, with 1 being "strongly disagree" and 10 being "strongly agree".
            1. I was immersed in the experience.
            2. The experience felt realistic.

3. I enjoyed the experience.
4. The visuals were of high quality.
5. The audio was of high quality.

7. After experiencing all 4 environments, the subject fills out another, brief survey.
    1. Final Survey Questions:
        i. Please rank the 4 VR experiences from best to worst, with 1 being the best and 4 being the worst, in each of the following categories:
            1. Immersion
            2. Realism
            3. Enjoyment
            4. Visual Quality
            5. Audio Quality
8. Calculate averages of subject responses.
9. Analyze averages to draw conclusions regarding patterns in subjective user experiences.

4. Raw Data

Averages When Considered Individually

Environment	Average Rating per Parameter				
	Immersion	**Realism**	**Enjoyment**	**Visual Quality**	**Audio Quality**
1 (High quality visuals, with audio)	8.48	7.44	9	7.56	8.68
2 (High quality visuals, silent)	7.08	6.44	7.8	7.12	1.8
3 (Low quality visuals, with audio)	8.24	7.36	8.4	6.52	8.24
4 (Low quality visuals, silent)	7.2	6.2	7.56	7.32	2.36

Averages When Considered Collectively

Environment	Average Rank per Parameter				
	Immersion	**Realism**	**Enjoyment**	**Visual Quality**	**Audio Quality**
1 (High quality visuals, with audio)	2	2.25	2	2.25	2.2
2 (High quality visuals, silent)	2.6	2.56	2.88	2.29	3
3 (Low quality visuals, with audio)	2.1	2.22	2.11	2.71	2
4 (Low quality visuals, silent)	7.2	2.78	2.8	2.7	3

Figures 1&2. Tables of average rating and rank per parameter for each of 4 environments, considered individually and collectively, respectively.

5. Data Analysis

5.1 In Regards To Immersion

In the category of immersion, results were fairly consistent both immediately after experiencing environments individually and in hindsight after experiencing all four environments.

5.1.1 individually

Individually, Variation 1 was rated as most immersive, followed by Variation 3. Variations 4 and 2 were ranked third and fourth respectively. From this, it can be gathered that the presence of audio resulted in increased immersion, regardless of visual quality. However, increased visual quality only contributed to increased immersion when audio was present. Without the addition of audio, increased visual quality resulted in decreased immersion. From this, no correlation can be observed between visual quality and level of immersion.

5.1.2 collectively

When reflecting upon all four environments, results were similar. Variation 1 was again rated as most immersive, followed by Variation 3. In this situation however, Variation 2 outranked Variation 4. From this, it can be ascertained that the presence of audio resulted in increased immersion, regardless of visual quality. It can also be ascertained that increased visual quality resulted in increased immersion, but that its effects were less profound than those of the audio.

5.2 In Regards To Realism

In the category of realism, results were fairly consistent both immediately after experiencing environments individually and in hindsight after experiencing all four environments.

5.2.1 individually

Individually, Variation 1 was rated as most realistic, followed by Variation 3. Variations 2 and 4 were ranked third and fourth respectively. From this, it can be gathered that the presence of audio resulted in increased realism, regardless of visual quality. It can also be ascertained that increased visual quality resulted in increased realism, but that its effects were less profound than those of the audio.

5.2.2 collectively

When reflecting upon all four environments, results were similar. In this instance however, Variation 3 was rated as most realistic, followed by Variation 1. Again, Variation 2 outranked Variation 4. From this, it can be ascertained that the presence of audio resulted in increased realism, regardless of visual quality. However, increased visual quality only contributed to increased realism when audio was not present. With the addition of audio, increased visual quality resulted in decreased realism. From these conflicting results, no correlation can be observed between visual quality and level of realism.

5.3 In Regards to Enjoyment

In the category of enjoyment, results were fairly consistent both immediately after experiencing environments individually and in hindsight after experiencing all four environments.

5.3.1 individually

Individually, Variation 1 was rated as most enjoyable, followed by Variation 3. Variations 2 and 4 were ranked third and fourth respectively. From this, it can be gathered that the presence of audio resulted in increased enjoyment, regardless of visual quality. It can also be ascertained that increased visual quality resulted in increased enjoyment, but that its effects were less profound than those of the audio.

5.3.2 collectively

When reflecting upon all four environments, results were similar. Variation 1 was again rated as most enjoyable, followed by Variation 3. In this situation however, Variation 4 outranked Variation 2. From this, it can be ascertained that the presence of audio resulted in increased enjoyment, regardless of visual quality. However, increased visual quality only contributed to increased enjoyment when audio was present. Without the addition of audio, increased visual quality resulted in decreased enjoyment. From these conflicting results, no correlation can be observed between visual quality and level of enjoyment.

5.4 In Regards to Perceived Visual Quality

In this category, it was hypothesized that the presence of audio would result in increased perceived visual quality. In other words, it was speculated that being able to hear the audio would cause subjects to perceive the visuals as being of higher quality. This was found to be the case in some instances, but not to a very large extent.

5.4.1 individually

Individually, Variation 1 was perceived as having the highest visual quality. Unexpectedly, Variation 4 was rated second best, followed by Variation 2. Variation 3 was perceived as having the worst visual quality. A number of conclusions can be drawn from these results. When graphics were actually of high quality, it can be seen that the presence of audio resulted in higher perceived visual quality. However, when graphics were actually of low quality, the presence of audio resulted in lower perceived audio quality. Because of this contradiction, no correlation can be confidently stated between the presence of audio and perceived visual quality. Interestingly, subjects also ranked Variation 4 higher than Variation 2. Because of this, there is also no correlation found between the actual and perceived visual quality of the VR environments. It should be noted that the combination of high quality visuals and audio did result in the highest perceived visual quality.

5.4.2 collectively

When all four variations were considered retrospectively, Variation 1 was perceived to have the highest visual quality, followed by Variation 2. Variation 4 was ranked third, and Variation 3 was ranked last. A number of conclusions can be gathered from this data. Firstly, it can be seen that with high quality graphics, the presence of audio resulted in greater perceived visual quality in the subject. In this regard, the hypothesis was correct. However, when the graphics were actually of lower quality, the presence of audio resulted in lesser perceived visual quality, contradicting the hypothesis. Because of these conflicting results, it cannot be confidently stated that the presence of audio has any correlation with perceived visual quality. Unlike in other categories, visual quality had a more profound effect on the data than the presence of audio, as the two high quality renders outranked the two low quality renders.

5.5 In Regards to Perceived Audio Quality

There were conflicting results in this category. Both individually and collectively, the two variations with audio obviously had the greatest perceived audio quality, but their ranks differed. Individually, Variation 1 outranked Variation 3, meaning that higher quality visuals resulted in increased perceived audio quality. In other words, when the graphics looked better, subjects thought the audio sounded better.

However, when looked at collectively, Variation 3 was ranked first, and Variation 1 was ranked second. From this, it can be ascertained that lower quality graphics resulted in greater perceived audio quality. In other words, when the graphics looked better, the subjects thought that the audio sounded worse. Because of these contradicting results, it cannot be confidently concluded that visual quality has any effect of perceived audio quality.

6. Conclusion

Overall, it can be observed that the presence of audio has a profound effect on subjective user experience in the context of virtual reality. It can be confidently concluded that the presence of audio has strong correlation with increased immersion, realism, and enjoyment, regardless of the visual quality of the VR environment. Results are conflicting however, when examining perceived visual quality. This study is unable to find correlation one way or the other between the presence of audio and perceived visual quality, as results were conflicting. It is also unable to conclude that the quality of the visuals has any effect on the perceived audio quality.

Further research should be done to verify these findings, as this sort of study is new in the context of VR, although similar studies in the context of film have been done, with conflicting results. Further, an analysis of these findings in the context of their psychological reasonings would be of great merit. Specifically, this analysis might delve into the

correlation between the dependent measures from this study (immersion, realism, enjoyment, etc.) and the concepts of "presence, engagement, and immersion." In the context of VR, presence has been defined as "the user's sense of being there in the virtual environment", while immersion has been defined as "the illusion that the virtual environment technology replaces the user's sensory stimuli by the virtual sensory stimuli."[9] Additionally, these measures could be used to more broadly describe "UX (user experience)", which is defined as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service."[10] These specific qualities can be measured with questionnaires developed by Witmer and Singer.[11] Perhaps, this particular study could be repeated using language more closely resembling these questionnaires, in order to use the dependent measure of "presence" as is more customary in the field of virtual reality media research.

7. Acknowledgements

The author wishes to express his utmost appreciation to the following people: Dr. Nathan Wolek, Dr. Charles Underriner, Dr. Kimberly Reiter, Madison Creech, Dengke Chen, Rich Grula**,** Ben Noel, Chris Roda, John Rotolo, Hezekiah Olopade, Tony Ganus, Trevor Hamlin, and Lauren Davids. Without the help and support of these individuals, the completion of this study would have proved impossible. Their wisdom and knowledge was invaluable. Additionally, the author would like to thank the Stetson University College of Arts and Sciences: Dean's Fund, as well as the Stetson Undergraduate Research Experience Grant, for funding this research. Finally, the author wishes to thank University of Central Florida: FIEA, for allowing him to use their facilities and virtual reality equipment.

8. References

1. Langvik, Sara. "Influence of audio on perception and comprehension of video sequences:." Master's thesis, KTH Royal Institute of Technology, 2015.
2. Beerends, John G., and F. E. De Caluwe. "The influence of video quality on perceived audio quality and vice versa." *Journal of the Audio Engineering Society* 47, no. 5 (December 2011).
3. Cuadrado, Francisco, Isabel Lopez-Cobo, Tania Mateos, and Beatriz Valverde. "Film Sound, Immersion, and Learning." In *AES E-Library*. Vol. 144. N.p.: Audio Engineering Society, 2018.
4. Hughes, Braham, and Jonathan Wakefield. "An Investigation into Plausibility in the Mixing of Foley Sounds in Film and Television." In *AES E-Library*. Vol. 138. N.p.: Audio Engineering Society, 2015.
5. Chattopadhyay, Budhaditya. "Reconstructing atmospheres: Ambient sound in film and media production." *Communication and the Public* 2, no. 4 (November 13, 2017).
6. Rummukainen, Olli, Jing Wang, Zhitong Li, Thomas Robotham, Zhaoyu Yan, Zhuoran Li, Xiang Xie, Frederik Nagel, and Emanuël A.P. Habets. "Influence of Visual Content on the Perceived Audio Quality in Virtual Reality." In *AES E-Library*. Vol. 145. N.p.: Audio Engineering Society, 2018.
7. Levee, John. "Correction of Spatialization Issues in Acousmatic Music: Remedying Incompatibility Between SpatGRIS and Logic Pro X." *Proceedings of The National Conference of Undergraduate Research (NCUR)*, 2019, 642-49.
8. British Broadcasting Corporation. "BBC Sound Effects." Audio. 2018. Accessed June 16, 2020. http://bbcsfx.acropolis.org.uk/.
9. Tcha-Tokey, K., Loup-Escande, E., Christmann, O., Richir, S. (2016). A Questionnaire to Measure the User Experience in Immersive Virtual Environments. Paper presented at VRIC, Computer Science Conference.
10. Pallot, M., Eynard, R., Poussard, B., Christmann, O. and Richir, S. (2013). Augmented sport: exploring collective user experience. In Proceedings of the Virtual Reality International Conference: Laval Virtual (VRIC '13). ACM, New York, NY, USA, Article 4, 8 pages.
11. Witmer, B. G., & Singer, M. J., 1998. Measuring presence in virtual environments: A presence questionnaire. Presence: Teleoperators and virtual environments, 7(3), 225-240.