

Journal of
• **Virtual Worlds Research**

jvwr.net

ISSN: 1941-8477

Assembled 2020

December 2020

Volume 13 No. 2-3



Volume 13, Number 2-3

Assembled 2020

December 2020

Editor In Chief

Yesha Sivan

CUHK Business School
The Chinese University of Hong Kong, HK

Issue Editors

Angie Cox (Prime)

Trident at American InterContinental University

Ryan Durbin

Trident at American InterContinental University

Vernell Hall

Trident at American InterContinental University

Coordinating Editor

Tzafnat Shpak

Cover image: © i8.ventures



The JVWR is an academic journal. As such, it is dedicated to the open exchange of information. For this reason, JVWR is freely available to individuals and institutions. Copies of this journal or articles in this journal may be distributed for research or educational purposes only free of charge and without permission. However, the JVWR does not grant permission for use of any content in advertisements or advertising supplements or in any manner that would imply an endorsement of any product or service. All uses beyond research or educational purposes require the written permission of the JVWR. Authors who publish in the Journal of Virtual Worlds Research will release their articles under the Creative Commons Attribution No Derivative Works 3.0 United States (cc-by-nd) license. The Journal of Virtual Worlds Research is funded by its sponsors and contributions from readers, with main sponsorship by i8 Ventures.



Volume 13, Number 2-3

Assembled 2020

December 2020

The Use of Ambisonic Audio to Improve Presence, Focus, and Noticing While Viewing 360 Video

Richard E. Ferdig

Karl W. Kosko

Enrico Gandolfi

Kent State University, USA

Abstract

Research has provided evidence of the value of producing multiple representations of content for learners (e.g., verbal, visual, etc.). However, much of the research has acknowledged changes in visual technologies while not recognizing or utilizing related audio innovations. For instance, teacher education students who were once taught through two-dimensional video are now being presented with interactive, three-dimensional content (e.g., simulations or 360 video). Users in old and new formats, however, still typically receive monophonic sound. A limited number of research studies exist that have examined the impact of combining three-dimensional sound to match three-dimensional video in learning environments. The purpose of this study was to respond to this gap by comparing the outcomes of watching 360 video with either monophonic or ambisonic audio. Results provided evidence that ambisonic audio increased perceived presence for those familiar with the content being taught, led to differentiation in what ambisonic viewers noticed compared to monophonic groups, and improved participant focus in watching the 360 video. Implications for the development and implementation into virtual worlds are discussed.

1. Introduction

In the mid 90's, drawing from dual coding theory (see Clark & Paivio, 1991; Paivio, 1990), educational technologists began exploring the value of multimodal learning. Simply stated, multimodal learning entails teaching or learning through "more than one presentation medium" (Mayer & Sims, 1994, p. 390). Such a process could be as simple as understanding outcomes of a learner introduced to visuals alone, to verbal cues alone, or to a combination of verbal and visual material. Research outcomes varied, with significant concerns about cognitive overload; however, most agreed that dual coding of content led to significant gains (e.g., Gellevej, Van Der Meij, De Jong, & Pieters, 2002).

Theoretically, multimodal composition could include any medium. As such, researchers have gone beyond audio and visual to include additional sensory perceptions like haptics (Lacey & Campbell, 2006; Sathian, 2016). These forays notwithstanding, a considerable body of research in fields like education has focused on innovations tied to visual cues or the relationship between visual innovations and traditional audio supplements. For instance, significant work has examined learning through augmented (Gandolfi, Ferdig, & Immel, 2018) or virtual reality (Ferdig, Gandolfi, & Immel, 2018).

Comparatively less work has focused on the evolution and innovation of auditory representations of content. This dearth of research, however, is juxtaposed against new visual innovations that require similarly innovative delivery of audio information. For instance, 360 video allows users to look anywhere within a live or recorded video space. Research is beginning to provide evidence of the usefulness of such tools (Ferdig & Kosko, 2020; Kosko, Ferdig, & Zolfaghari, 2020; Zolfaghari, Austin, Kosko, & Ferdig, 2020) and their impact on perceived presence (Gandolfi, Kosko, & Ferdig, in press) -- a significant component of immersion (Ferdig, Gandolfi, & Immel, 2018).

Presence has been described as the feeling of 'being there' experienced within a mediated environment, with the support of sensorial and cognitive stimuli (Lee, 2004; Mestre, 2005). A high level of presence is a desired outcome because it implies that the user is engaging with the virtual world without perceiving the burden of the technology itself. Therefore, an experience able to promote presence is arguably more immersive, engaging, and meaningful, with relevant learning outcomes and processes at stake (Gandolfi, Kosko, & Ferdig, in press; Lee & Wong, 2014; Webster, 2016). However, most of the 360 video research on immersion and presence relies on video delivered with monophonic audio. Few empirical education studies capitalize on matching the three-dimensional features of 360 video with similar three-dimensional capabilities of ambisonic audio.

This paper presents results from an empirical study on the use of 360 video in a teacher education classroom. The 360 video was delivered to one group using traditional, monophonic audio, while the second group received the same 360 video but with ambisonic audio. The first research question set out to determine whether there were any differences in perceived presence between the groups. The second question explored whether the sound type impacted viewing attributes. Findings are presented as well as implications for the use and development of ambisonic audio for virtual and augmented worlds.

2. Literature Review

Some of the early, important work in the field of educational technology provided support for the importance of dual coding (see Clark & Paivio, 1991) and its specific impact on multimodal learning. Mayer & Sims (1994), for instance, argued such an approach emphasized "the learner's building of mental connections between visual and verbal representations" (p. 400). Such work, in

combination with easier development and dissemination of multimedia materials, began a long and storied relationship between the use of audio and visual information for learning across the lifespan.

In teacher education, for instance, representations of real classrooms with both video and audio stimuli have been used for decades to prepare future teachers (Christ, Arya, & Chiu, 2017). Formats have varied; they have included tools as basic as two-dimensional videos and have been as advanced as the implementation of simulations (Kaufman & Ireland, 2016) or 360 video (Theelen, Van den Beemt, & den Brok, 2019). Although the technologies may vary, the results have been genuinely positive. For instance, the combination of audio and visual representations of teacher practice can lead to improved reflection (Barth-Cohen, Little, & Abrahamson, 2018), listening (Widodo & Rozak, 2016), noticing (Walkoe, Sherin, & Elby, 2020), and analytic skills (Frommelt, Hugener, & Krammer, 2019).

These technologies have continued to advance. Unfortunately, this growth has been one-sided with visual representations continuing to adapt and change. They have included cartoons (Chazan, Herbst, Fleming, & Grosser-Clarkson, 2018), simulations (Kaufman & Ireland, 2016), virtual reality (Billingsley, Smith, Smith, & Meritt, 2019), augmented reality (Sáez-López, Cózar-Gutiérrez, González-Calero, & Gómez Carrasco, 2020), and—most recently—360 video (Ferdig & Kosko, 2020; Theelen et al., 2019). In other words, the visual technologies—and not just their uses in education—have adapted and evolved.

Comparatively, the uses of audio have evolved, but not necessarily investigations of the technological advances for teaching and learning. Rather, research has focused on the implementation of new ways to engage learners with *traditional* audio. For instance, teacher educators have used podcasts (Schreiber & Klose, 2017), audio feedback (Johnson & Cooke, 2016), audio-based stories (Piwinsky, Everett, Fulton, & Ross, 2020), and songs or song-writing (Joseph, Nethsinghe, & Cabedo Mas, 2018).

This is not to suggest there has been a lack of advancements in audio. Indeed, advancement in the technological aspects of audio production and delivery continue to evolve. For instance, those working in audio will point to dramatic changes in moving from single to multi-channel audio systems (Malham & Myatt, 1995). There are also “electronic systems that allow sounds to be positioned artificially in space” (p. 60). One of those is called *ambisonic sound*. According to Malham & Myatt (1995), “the ambisonic surround-sound system is essentially a two-part technological solution to the problems of encoding sound directions (and amplitudes) and reproducing them over practical loudspeaker systems so that listeners can perceive sounds located in three-dimensional space” (p. 62). A user listening to ambisonic sound would hear sound or music beyond simple left and right channels; they would conceivably hear directionality in the audio.

We are also not claiming that audio has yet to be explored within virtual worlds. There are several research studies that have provided evidence of the importance of sound and sound design. For instance, researchers have explored the impact of audio on navigation (Gandolfi & Clements, 2019), the design of virtual worlds with content-based audio retrieval (Janer Mestres, Finney, Roma Trepas, Kersten, & Serra, 2009), and speech animation of virtual characters (Charalambous, Yumak, & van der Stappen, 2019).

Rather, the argument is that we have reached a point where the video delivery for educational outcomes does not necessarily match the audio capabilities available or required for the chosen video format. Said differently, we have virtual worlds that enable three-dimensional engagement; however, most platforms utilize sound systems and features that are presented in a monophonic perspective (e.g., a single channel of sound). Listeners are visually in a three-dimensional space while receiving two-dimensional sound. The challenge, of course, is that a disconnect is introduced between two mental representations (Mayer & Sims, 1994). Users are watching a visual representation that does not equate

to the audio representation of the same phenomenon. It is the cognitive equivalent of hearing one's name called but being unable to pinpoint the voice because the sound is heard in uniform intensity and volume from all directions.

This gap can also be framed through the lenses of presence and embodied cognition. The former concept relies on the ability of a virtual environment to immerse the user by providing a meaningful and realistic setting (Ferdig, Gandolfi, & Immel, 2018) – i.e., the feeling of being there (Lombard & Ditton, 2006). Ambisonic audio would hypothetically improve the fidelity of an immersive video because of the more realistic hearing experience (Malham & Myatt, 1995); as such, the feelings of immersion and presence could theoretically be enriched (Rupp, Odette, Kozachuk, Michaelis, Smither, & McConnell, 2019).

The latter relies on the argument that cognition is embodied rather than purely abstract. As such, involving senses is a key front for stimulating individuals and developing meaningful learning experiences (Amin, Jeppsson, & Haglund 2015; Wilson, 2002). In addition, embodied cognition processes depend on the concrete context of use along with sensorial involvement (Davidson, 2016). Ambisonic audio may play a role in this regard, providing more information about a given situation through a complete auditory embodiment (Chateau, Maffiolo, Pican, & Mersiol, 2005).

To our knowledge, although work has been documented in other fields (particularly music) about the processes surrounding ambisonic sound (e.g., Cheng, Ritz, & Burnett, 2008), very little research has explored the impact of having or not having ambisonic or three-dimensional sound to accompany three-dimensional video in learning environments. Cooper and Taylor (1998) proposed a system using ambisonic sound for virtual environments for the blind. De Vries (2014) also conducted a study but found no differences in perceived immersion between ambisonic and monophonic groups. However, no research was found that matched 360 video with ambisonic audio for learning environments or the impact of demographics within such explorations. Given this dearth of knowledge in understanding the impact of using ambisonic audio with 360 video, this paper set out to address two important questions.

RQ1: Are there any differences in perceived presence between ambisonic and monophonic 360 video, either overall or by demographic factors?

RQ2: Are there any differences in viewing attributes of the 360 video between ambisonic and monophonic listeners, either self-reported or actual?

3. Methods

3.1. Sample

This research study was submitted, reviewed, and approved for human subjects by the authors' institutional research board. Participants were recruited from a college of education in the Midwest United States. Forty-six preservice teachers (PSTs) responded to and participated in the study. The majority of participants were white (89.1%) and female (76.1%). Reported teacher licensure areas were widespread with 19.6% focusing on elementary grades, 6.5% on middle grades, 41.3% on secondary grades, and 32.6% on multi-grade band foci (special education, art education, music education, foreign language, etc.). Of those participants focusing on specific grade-bands (elementary, middle, secondary), 15 (32.6% of the total sample) were preparing to teach mathematics at one of those grade-bands. When asked about their perceived comfort using various technologies (their technological savviness) and prior experiences with VR, the majority of participants considered themselves more technologically savvy than not ($M=6.23$, $SD=1.80$ on a 10 point scale), 73.3% had used a VR headset prior to the study, and 77.8% had watched a 360 video before.

3.2. Procedures

The entire study procedure was completed online and was self-paced. Students began with an initial questionnaire that asked them about basic demographic data and previous experience with 360 video. Demographic data included gender, perceived technological savviness, and their specific education majors (e.g., mathematics or literacy education). Participants then watched a 3-minute tutorial video on how to view and navigate within 360 video.

At this point in the study, participants were randomly assigned to one of two groups. The first group, called the monophonic group, was asked to watch a 360 video about an elementary classroom mathematics lesson. The video focused on the teaching of equivalent fractions. Participants in this group saw the video on their desktop or laptop; however, because it was a 360 video, they were able to use their mouse or finger (e.g., laptop trackpad) to look around the room (see Figure 1). The audio accompanying the video track was monophonic, meaning that they heard the sounds with the same intensity, volume, and directionality, regardless of where they looked.

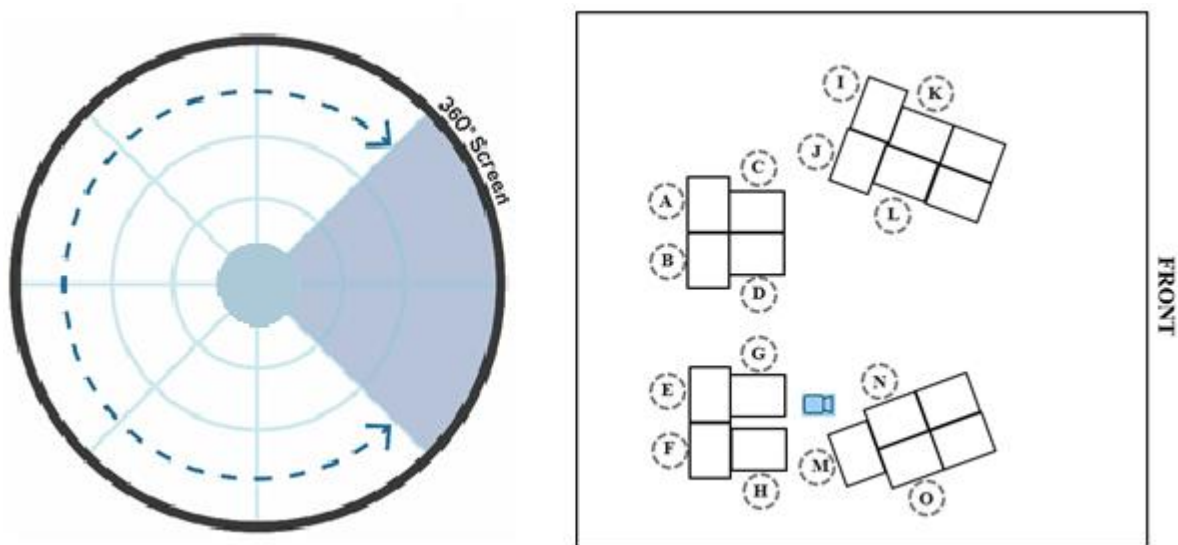


Figure 1: Perceptual capacity of what is viewable (left image) and a classroom map of the recorded lesson with 360 camera location (right image).

The second group, called the ambisonic group, followed the same process. However, the video they watched was accompanied by an ambisonic recording of the same video. Ambisonic audio provided directionality, so that a student talking in front of where the viewer was watching would have sound coming from that direction. A student out of the viewer's perspective would produce sound from the direction of the student.

After the first video, both groups were asked to describe *pivotal moments* in the 360 video; pivotal meant important moments related to teaching and learning that they noticed in the classroom video. They were then asked to watch the video a second time, with a request that they screen record their viewing and share it with the researchers. After the second viewing, participants were asked to select one of the pivotal moments that they noticed in the classroom video and to describe it and why they thought it was important.

The session concluded with a final survey. The survey had three main parts. First, it asked about their perceived presence while viewing the 360 video. This part of the survey was taken from an established and validated instrument called the *Extended Reality Presence Scale*, or XRPS (Gandolfi,

Kosko, & Ferdig in review). Presence scores were Rasch modeled, and participant scores are logit-based ($M=.79$ $SD=1.02$). The second part of the survey was a map of the classroom. Respondents were asked to pick the top ten locations they focused on while watching the video. Finally, participants were asked to indicate the type of device used to listen to the audio. The options included in-ear (e.g., earbuds), on-the-ear, or over-the-ear.

All 46 subjects completed the pre-survey and post-survey instruments. A total of 24 participants also uploaded videos of their screen-recorded sessions (21 of which were viable files for analysis).

4. Results

RQ1: Ambisonic audio and presence

Research question one set out to determine whether there were any differences in perceived presence between those watching 360 video with ambisonic audio and those with monophonic audio. This question was posed overall in addition to examining demographic data from the users; it was evaluated using the XRPS. Demographic data included gender, self-reported technological savviness, area of study within education, and the type of headphones they were using (e.g., in-ear, on-the-ear, or over-the-ear).

There was no statistically significant difference between monophonic ($M=.89$, $SD=1.24$) and ambisonic ($M=.69$, $SD=.78$) conditions ($t(df=44)=0.65$, $p=.522$). There were also no statistically significant differences for gender or the type of headphones used. However, there was a statistically significant difference between majors preparing to teach math at any K-12 grade ($M=1.52$, $SD=.23$) and those education majors teaching other subjects ($M=0.39$, $SD=.17$) [$t(df=42)=3.930$, $p<.001$]. This magnitude of difference is extremely large with over a logit difference between groups.

In sum, data from this study suggest that the use of ambisonic or monophonic audio for 360 video only impacted perceived presence for users whose area of study was associated with the content area of the video (e.g., mathematics).

RQ2: Ambisonic audio and viewing attributes

Research question two explored whether there were differences in viewing attributes of the 360 video between ambisonic and monophonic conditions. To address this, participants created a self-reported heat map with the 10 places in the classroom they perceived they attended the most. Those self-reported locations were placed onto a map that divided the classroom into four locations (see Figure 2). A fifth category called 'indeterminate' was added as some self-reported locations could not be clearly attributed to one region or another. A Chi-Square analysis resulted in a statistic of 4.366 ($DF=4$; $p=.498$) and was found to not be statistically significant from chance. In other words, the data suggest that there were no significant differences in where participants looked between ambisonic and monophonic groups, at least according to self-perception.

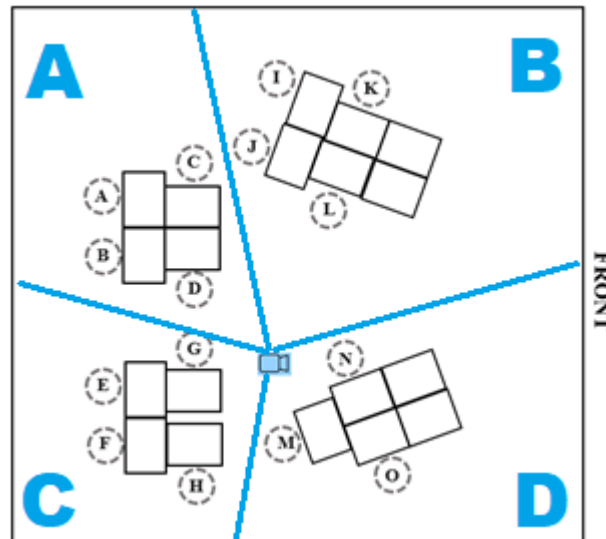


Figure 2: Layout of 360 recorded classroom with regions identified for analysis.

Twenty-one participants (11 for the ambisonic group and 10 for the monophonic group) also provided screen captured video of their 360 video viewing. These videos were used to identify the quadrant of the video attended to by the user (four quadrants and a fifth indeterminate category). Videos were analyzed second by second and then placed into a contingency table similar to the self-reported heat maps for analysis. Results were statistically significant ($\chi^2(df=4)=311.76, p<.001$), indicating that the distribution of where participants attended by condition was not due to chance (see Table 1).

Table 1: Chi-square analysis of viewing locations between groups.

	Region A	Region B	Region C	Region D	Indeterminant	Total
Monophonic	339	1184	525	1301	44	3393
	<i>216.1</i>	<i>1073.94</i>	<i>433.13</i>	<i>1626.81</i>	<i>43.03</i>	
Ambisonic	123	1112	401	2177	48	3861
	<i>245.9</i>	<i>1222.06</i>	<i>492.84</i>	<i>1854.19</i>	<i>48.97</i>	
Total						

Note: Observed counts are in regular text and expected counts are in italics.

In addition to examining where they looked, participants' viewing variability was examined. Their variability was a measure of how many times they changed their perspective. To measure this, an *unlikeability* score was calculated. Unlikeability is a nonparametric measure for nominal variables akin to variance for parametric mean (Kader & Perry, 2007). The U_2 statistic indicates the measure of how alike the frequencies are across multiple categories (range 0 to 1).

To complete the unlikeability analysis, a statistic was calculated for each of the 21 participants with videos and then compared the statistic across conditions (e.g., monophonic vs. ambisonic). Results demonstrated a much higher degree of unlikeability (e.g., variance) for monophonic participants ($U_2=0.68$) than the ambisonic condition ($U_2=0.58$). A Wilcoxon-Mann-Whitney test found this difference to be statistically significant ($W=76, p=.002$).

To better understand these results, a visual comparison of participants' changes in perspective (looking at one region and then another) was conducted. Specifically, Figure 3 illustrates these changes across all participants along the timeline of the 360 video. Although there appear to be periods of significant variance across both groups, participants in the ambisonic condition had sustained periods of focus. Specifically, their field of view remained in the same viewing locations longer than participants in the monophonic condition. Furthermore, these sustained periods of focus were also in similar regions within the ambisonic condition. Stated differently, participants in the ambisonic condition switched perspectives less frequently than those in the monophonic condition and also tended to focus on the same regions within the condition.

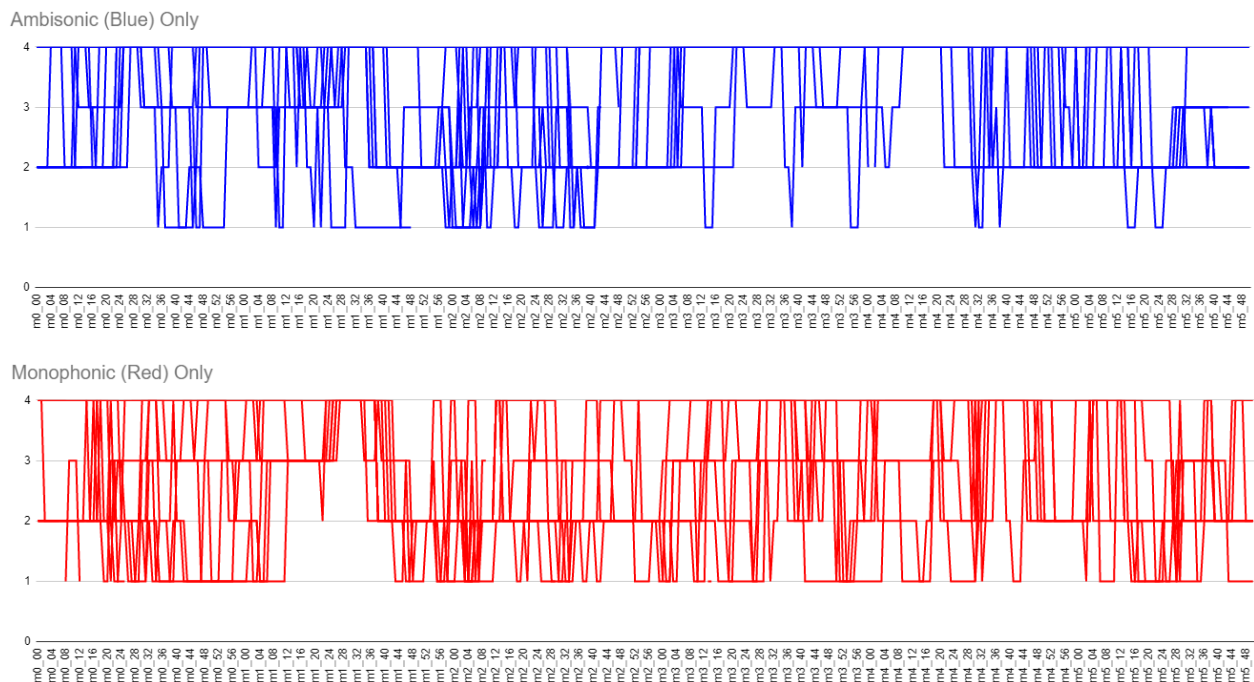


Figure 3: Perspective change between ambisonic (top) and monophonic 360 viewers (bottom).

In sum, there were no statistical differences observed in self-reported 360 viewing locations between ambisonic and monophonic conditions ($\chi^2(df=4)=4.366, p=.498$). However, examining actual viewing data in a second-by-second frame showed statistically significant differences in where participants looked ($\chi^2(df=4)=311.76, p<.001$). Moreover, users in the monophonic condition were significantly more likely to change their perspective as they watched the 360 video and attend to different regions from one another. This was demonstrated by a higher variance as well as the number of times perspectives were switched ($W=76, p=.002$).

5. Discussion

The first research question set out to explore whether there were differences in perceived presence between users viewing 360 video with either ambisonic or monophonic audio. Results from a general analysis of all participants found no statistically significant differences, similar to what de Vries found when examining immersion (2014).

However, data were also examined by sorting group differences according to student major. Results showed that those students who were studying majors related to the content of the video (in this case, mathematics) felt significantly higher levels of presence. This finding deserves further exploration. However, if dual coding and multimodal learning are focused on representations to

support meaningful learning (Mayer & Sims, 1994), it seems likely that those more tied to the content would experience higher levels of presence. This finding can also be related to the relationship between embodied cognition and context of use (Wilson, 2002), implying a different embodied experience due to the more familiar setting.

This lends support for the notion that ambisonic sound could be a valuable investment for instructional purposes, particularly in scenarios where participants are already engaged with or related to the content. As such, future studies are needed to understand the impact of this technology in different academic disciplines and professional development instances, looking at participants' proficiency in specific content areas (e.g., high versus low proficiency). This is also related to the need for building a more structured approach when designing and developing immersive experiences intended to promote presence, from proper representations of practice to specific visual and auditory clues to guide users' attention and focus.

The second research question attempted to examine any viewing differentiation between ambisonic and monophonic viewers of 360 video content. This was assessed in three ways. First, self-reported heat maps were examined to see if participants in different groups looked at content differently in the classrooms. Second, actual screen recordings of participants' viewing of 360 video were analyzed. Third, in addition to examining where participants looked, variances within those viewings were examined for inconsistencies.

There were no statistically significant differences between monophonic and ambisonic groups when analyzing self-reported viewing data ($\chi^2(df=4)=4.366, p=.498$). However, there were significant differences when comparing the second-by-second analysis of screen recordings ($\chi^2(df=4)=311.76, p<.001$). Both groups paid attention to different things happening in the video. Moreover, the ambisonic group stayed more focused ($U_2=0.58$) than the monophonic group ($U_2=0.68$); they had less variance in where they attended and less changing of perspective ($W=76, p=.002$).

More research, obviously, needs to be completed to evaluate whether where participants looked translated into better learning outcomes (in this case, becoming a better mathematics instructor). In other words, the two groups focused on different things. One might question why this was not picked up in the self-reported heat map. A correlation of the proportions between the self-reported heat maps and the recorded video resulted in a Spearman Rho of .67. It could be that the second-by-second analysis was more precise than the broader selection of 10 locations. However, both data sets point to differentiation that deserves further exploration across various learning contexts.

However, the lower variance and the fewer changes in perspective point to an important affordance of ambisonic sound. A participant watching a video of a classroom setting might be focused on a specific table. If a sound were introduced without directionality (e.g., monophonic), it could create cognitive dissonance for the user (Festinger, 1962). It would not happen in similar situations in real life; classroom sound would appear from a direction. Given this dissonance, they might turn to look and see where the sound came from, thus interrupting their focus. For the ambisonic group, a sound coming from behind them would obviously be from a different, and perhaps specific, table. They would be able to tune it out as the audio representation would match the visual interpretation. Such a finding suggests that ambisonic sound may help developers and educators more realistically represent reality in virtual worlds; that, in turn, may positively impact teaching and learning.

Moreover, if there is a correlation between focus and presence (as suggested by Gandolfi, Kosko, & Ferdig, in press), ambisonic audio may further facilitate this sense of immersion. This can also be associated with the sense of control perceived by participants and their ability to supervise their mediated surroundings without moving their heads (e.g., relying on hearing). As such, more studies are needed for better evaluating the level of comprehension of a given situation (e.g., pivotal events happening during a lesson) along with the use of ambisonic audio within a computer-mediated

environment, looking at potential differences. Finally, this reflection can be extended to audio design, questioning what audio inputs (e.g., realistic or artificial) may work better than others, and to other types of sensorial involvement (e.g., by including haptic feedback) able to make a difference.

6. Limitations

One limitation of this study was in the ability to analyze recorded screen captures. Of the 46 original participants, only 21 submitted videos that were able to be analyzed. It would be more beneficial to record such sessions automatically rather than rely on participants to do so. However, such technology was unavailable at the time of data collection. While appropriate statistical measures were implemented to address the smaller size of submitted videos, a larger sample might provide a further understanding of ambisonic outcomes.

A second limitation was that this work was done in a college of education using teacher education candidates. The video used was directly aligned with what they were studying. Other research could examine the use of 360 video not directly aligned or oriented with the beliefs or desired outcomes of the viewers or in learning situations outside of colleges of education (e.g., Ulrich, Helms, Frandsen, & Rafn, 2019). This is an important next step for research as findings on the use of ambisonic audio may have important implications for training environments in and outside of traditional K-12 or postsecondary education, such as healthcare (Pears, Yiasemidou, Ismail, Veneziano, & Biyani, 2020), language learning (Ji & Li, 2018), military environments (Maathuis, Pieters, & van den Berg, 2020), or training with games (Farkaš, 2018).

A third limitation is related to the technology hosting the 360 videos during the study, which was fully online and based on two-dimensional platforms (e.g., desktop or laptop monitors). This may have introduced variance in some of the hardware used that could affect change in perspective (mouse, touchpad, or touchscreen) or audio (in-ear, on-ear, over-ear). Additional hardware like virtual reality headsets (e.g., *Oculus Go*) could have been used as an alternative, which may be more aligned with how participants naturally respond to sight and sound (i.e., turning one's head instead of moving the screen perspective). Future studies may address these innovations and explore possible comparisons (see Rupp et al., 2019).

A final limitation was that this ambisonic study was focused on 360 video. Future work should replicate this study in other augmented and virtual reality environments. For instance, there is new and promising work in the use of holograms for teaching and learning across the lifespan (e.g., Moro, Phelps, Jones, & Stromberga, 2020). Otherwise, digital settings like *Minecraft.edu* are increasingly hosting live lessons in K12 and higher education (e.g., Callaghan, 2016), suggesting a new front to analyze. Studies could compare these new visual representations when paired with ambisonic audio described here.

6.1. Conclusion

Audio and visual representations of practice have been used—and continue to be used—significantly in teaching and learning across the lifespan. Although visual representations have evolved and now include augmented and virtual worlds, research on the use of audio implementation in those environments are sparse. More specifically, audio that matches the visual representation (e.g., 3D audio with 3D video) for learning is a relatively unexplored domain, particularly when examining participant immersion and related feelings of presence.

This study attempted to address this need by examining 360 video with both ambisonic and monophonic audio. Results provided clear differences in where participants looked and their choice to change where they looked. Participants who viewed 360 video with ambisonic audio maintained visual

focus more so than those with monophonic audio, with potential implications in terms of perceived control and situational understanding. These findings support an argument for the usefulness of ambisonic video as a necessary tool for content delivery, particularly when focusing on specific places in 360 video is fundamental to advancing learning of the video content. However, the observable effect of ambisonic audio in such contexts may be subtle, as meaningful differences for self-reported data from participants in this paper were not observed. Therefore, while the role and effect of ambisonic audio in virtual settings are in need of additional study, consideration of the benefits associated with the cost of including such audio is also important.

Future studies are needed to continue to shed light on how ambisonic audio is intertwined with the comprehension of the mediated environment and any related user performance (e.g., noticing specific events or performing a set of tasks). In addition, more research should be directed toward the role of the user's background in supporting presence in extended reality settings, which can co-interact with sensorial stimuli for providing a meaningful and engaging experience.

Acknowledgments

The research reported here received support from the National Science Foundation through DRK-12 Grant #1908159. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Amin, T. G., Jeppsson, F., & Haglund, J. (2015). Conceptual metaphor and embodied cognition in science learning: Introduction to special issue. *International Journal of Science Education*, 37(5-6), 745-758
- Barth-Cohen, L. A., Little, A. J., & Abrahamson, D. (2018). Building reflective practices in a pre-service math and science teacher education course that focuses on qualitative video analysis. *Journal of Science Teacher Education*, 29(2), 83-101.
- Billingsley, G., Smith, S., Smith, S., & Meritt, J. (2019). A systematic literature review of using immersive virtual reality technology in teacher education. *Journal of Interactive Learning Research*, 30(1), 65-90.
- Callaghan, N. (2016). Investigating the role of Minecraft in educational learning environments. *Educational Media International*, 53(4), 244-260.
- Charalambous, C., Yumak, Z., & van der Stappen, A. F. (2019). Audio-driven emotional speech animation for interactive virtual characters. *Computer Animation and Virtual Worlds*, 30(3-4), e1892.
- Chateau, N., Maffiolo, V., Pican, N., & Mersiol, M. (2005, October). The effect of embodied conversational agents' speech quality on users' attention and emotion. In *International Conference on Affective Computing and Intelligent Interaction* (pp. 652-659). Springer, Berlin, Heidelberg.
- Chazan, D., Herbst, P., Fleming, E., & Grosser-Clarkson, D. (2018). Technological supports for practice-based teacher education. *Journal of Technology and Teacher Education*, 26(1), 5-11.

- Cheng, B., Ritz, C., & Burnett, I. (2008, March). A spatial squeezing approach to ambisonic audio compression. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 369-372). IEEE.
- Christ, T., Arya, P., & Chiu, M. M. (2017). Video use in teacher education: An international survey of practices. *Teaching and Teacher Education*, 63, 22-35.
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational psychology review*, 3(3), 149-210.
- Cooper, M., & Taylor, A. (1998). Ambisonic sound in virtual environments and applications for the blind people. In *Second European Conference on Disability, Virtual Reality, and Associated Technologies, ECDVRAT*, 113-118.
- Davidson, A. (2016). Ontological shifts: Multi-sensoriality and embodiment in a third wave of digital interfaces. *Journal of Dance & Somatic Practices*, 8(1), 21-42.
- de Vries, F. (2014) *Examining observer response on environmental audio with ambisonic and stereo sound samples*. Bachelor's Thesis, Artificial Intelligence.
<https://fse.studenttheses.ub.rug.nl/12121/>
- Farkaš, T. (2018). Binaural and Ambisonic Sound as the Future Standard of Digital Games. *Acta Ludologica*, 1(2), 34-46.
- Ferdig R.E., Gandolfi E., & Immel Z. (2018) Educational opportunities for immersive virtual reality. In: Voogt J., Knezek G., Christensen R., Lai KW. (eds.) *Second Handbook of Information Technology in Primary and Secondary Education* (pp. 955-966). Springer International Handbooks of Education. Springer, Cham.
- Ferdig, R.E., & Kosko, K.W. (2020). Implementing 360 video to increase immersion, perceptual capacity, and teacher noticing. *TechTrends*, 64(6), 849-859.
- Festinger, L. (1962). Cognitive dissonance. *Scientific American*, 207(4), 93-106.
- Frommelt, M., Hugener, I., & Krammer, K. (2019). Fostering teaching-related analytical skills through case-based learning with classroom videos in initial teacher education. *Journal for Educational Research Online*, 11(2), 37-60.
- Gandolfi E., Ferdig R.E., & Immel Z. (2018) Educational opportunities for augmented reality. In: Voogt J., Knezek G., Christensen R., & Lai, KW. (eds.) *Second Handbook of Information Technology in Primary and Secondary Education* (pp. 967-979). Springer International Handbooks of Education. Springer, Cham.
- Gandolfi, E., & Clements, R. (2019) Alternative embodied cognitions at play: Evaluation of audio-based navigation in virtual settings via interactive sounds. *The Journal of Virtual Worlds Research*, 12(1), 1-14. <https://doi.org/10.4101/jvwr.v12i1.7328>
- Gandolfi, E., Kosko, K. W., & Ferdig, R. E. (in press). Situating presence within extended reality: Validation of the eXtended Reality Presence Scale (XRPS). *British Journal of Educational Technology*.
- Gellevij, M., Van Der Meij, H., De Jong, T., & Pieters, J. (2002). Multimodal versus unimodal instruction in a complex learning context. *The Journal of Experimental Education*, 70(3), 215-239.
- Janer Mestres, J., Finney, N., Roma Trepas, G., Kersten, S., & Serra, X. (2009). Supporting soundscape design in virtual environments with content-based audio retrieval. *Journal of Virtual Worlds Research* 2(3): 4-6. <https://doi.org/10.4101/jvwr.v2i3.635>

- Ji, S., & Li, K. (2018, December). A Study on the Effects of Binaural Listening Materials on Second Language Listening Comprehension. In *2018 International Joint Conference on Information, Media, and Engineering (ICIME)* (pp. 20-24). IEEE.
- Johnson, G. M., & Cooke, A. (2016). Self-regulation of learning and preference for written versus audio-recorded feedback by distance education students. *Distance Education*, 37(1), 107-120.
- Joseph, D., Nethsinghe, R., & Cabedo Mas, A. (2018). Creating multicultural music opportunities in teacher education: Sharing diversity through songs. *Australian Journal of Teacher Education*, 43(5), 32-47.
- Kader, G. D., & Perry, M. (2007). Variability for categorical variables. *Journal of Statistics Education*, 15(2). <https://doi.org/10.1080/10691898.2007.11889465>
- Kaufman, D., & Ireland, A. (2016). Enhancing teacher education with simulations. *TechTrends*, 60(3), 260-267.
- Kosko, K.W., Ferdig, R.E., & Zolfaghari, M. (2020). Preservice teachers' professional noticing when viewing standard and 360 video. *Journal of Teacher Education*. <https://doi.org/10.1177/0022487120939544>
- Lacey, S., & Campbell, C. (2006). Mental representation in visual/haptic crossmodal memory: evidence from interference effects. *Quarterly Journal of Experimental Psychology*, 59(2), 361-376.
- Lee, K. M. (2004). Presence, explicated. *Communication Theory*, 14(1), 27-50
- Lee, E. A., & Wong, K.W. (2014). Learning with desktop virtual reality: low spatial ability learners are more positively affected. *Computers & Education*, 79, 49-58.
- Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2). <https://doi.org/10.1111/j.1083-6101.1997.tb00072.x>
- Malham, D. G., & Myatt, A. (1995). 3-D sound spatialization using ambisonic techniques. *Computer music journal*, 19(4), 58-70.
- Maathuis, C., Pieters, W., & van den Berg, J. (2020). Decision support model for effects estimation and proportionality assessment for targeting in cyber operations. *Defence Technology*. <https://doi.org/10.1016/j.dt.2020.04.007>
- Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86(3), 389-401.
- Mestre, D. R. (2005). *Immersion and presence*. http://www.ism.univmed.fr/mestre/projects/virtual%20reality/Pres_2005.pdf
- Moro, C., Phelps, C., Jones, D., & Stromberga, Z. (2020). Using holograms to enhance learning in health sciences and medicine. *Medical Science Educator*, 1-2. <https://doi.org/10.1007/s40670-020-01051-7>
- Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.
- Pears, M., Yiasemidou, M., Ismail, M. A., Veneziano, D., & Biyani, C. S. (2020). Role of immersive technologies in healthcare education during the COVID-19 epidemic. *Scottish Medical Journal*, 65(4), 112-119.
- Piwinsky, M., Everett, S., Fulton, L. & Ross, M. (2020). Guiding Undergraduate Students in

- Scripting an Audio Based Story. In D. Schmidt-Crawford (Ed.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 183-188). Online: Association for the Advancement of Computing in Education (AACE). <https://www.learntechlib.org/primary/p/215749/>
- Rupp, M. A., Odette, K. L., Kozachuk, J., Michaelis, J. R., Smither, J. A., & McConnell, D. S. (2019). Investigating learning outcomes and subjective experiences in 360-degree videos. *Computers & Education*, 128, 256-268.
- Sáez-López, J. M., Cózar-Gutiérrez, R., González-Calero, J. A., & Gómez Carrasco, C. J. (2020). Augmented reality in higher education: An evaluation program in initial teacher training. *Education Sciences*, 10(2), 26, 1-12.
- Sathian, K. (2016). Analysis of haptic information in the cerebral cortex. *Journal of neurophysiology*, 116(4), 1795-1806.
- Schreiber, C., & Klose, R. (2017). Mathematical audio-podcasts for teacher education and school. *Teachers and Curriculum*, 17(2), 41-46.
- Theelen, H., Van den Beemt, A., & den Brok, P. (2019). Using 360-degree videos in teacher education to improve preservice teachers' professional interpersonal vision. *Journal of Computer Assisted Learning*, 35(5), 582-594.
- Ulrich, F., Helms, N. H., Frandsen, U. P., & Rafn, A. V. (2019). Learning effectiveness of 360° video: experiences from a controlled experiment in healthcare education. *Interactive Learning Environments*, 1-14. <https://doi.org/10.1080/10494820.2019.1579234>
- Walkoe, J., Sherin, M., & Elby, A. (2020). Video tagging as a window into teacher noticing. *Journal of Mathematics Teacher Education*, 23(4), 385-405.
- Webster, R. (2016). Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments*, 24(6), 1319–1333.
- Widodo, H. P., & Rozak, R. R. (2016). Engaging student teachers in collaborative and reflective online video-assisted extensive listening in an Indonesian initial teacher education (ITE) context. *Electronic Journal of Foreign Language Teaching*, 13(2), 229-244.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625-636.
- Zolfaghari, M., Austin, C.K., Kosko, K.W., & Ferdig, R.E. (2020). Creating asynchronous virtual field experiences with 360 video. *Journal of Technology and Teacher Education*, 28(2), 315-320.