

User Experience and Immersion of Interactive Omnidirectional Videos in CAVE Systems and Head-Mounted Displays

Pekka Kallioniemi, Ville Mäkelä, Santeri Saarinen, Markku Turunen, York Winter, Andrei Istudor

► **To cite this version:**

Pekka Kallioniemi, Ville Mäkelä, Santeri Saarinen, Markku Turunen, York Winter, et al.. User Experience and Immersion of Interactive Omnidirectional Videos in CAVE Systems and Head-Mounted Displays. 16th IFIP Conference on Human-Computer Interaction (INTERACT), Sep 2017, Bombay, India. pp.299-318, 10.1007/978-3-319-68059-0_20 . hal-01679797

HAL Id: hal-01679797

<https://hal.inria.fr/hal-01679797>

Submitted on 10 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



User Experience and Immersion of Interactive Omnidirectional Videos in CAVE Systems and Head-Mounted Displays

Pekka Kallioniemi¹, Ville Mäkelä¹, Santeri Saarinen¹, Markku Turunen¹, York Winter², Andrei Istudor²

¹ Tampere Unit for Computer-Human Interaction, University of Tampere, Finland
E-mail: {firstname.lastname@uta.fi}

² Department of Neurobiology, Humboldt-University Berlin, Berlin, Germany
E-mail: {firstname.lastname@charite.de}

Abstract. Omnidirectional video (ODV) is a medium that offers the viewer a 360-degree panoramic video view of the recorded setting. In recent years, various novel platforms for presenting such content have emerged. Many of these applications aim to offer an immersive and interactive experience for the user, but there has been little research on how immersive these solutions actually are. For this study, two interactive ODV (iODV) applications were evaluated: a CAVE system and a head-mounted display (HMD) application. We compared the users' expectations and experience and the level of immersion between these systems. Both indoor and outdoor recorded environments were included. First, the results indicate that the user's experiences with these applications exceed their expectations greatly. Second, the HMD application was found to be more immersive than the CAVE system. Based on the findings of this study, both systems seem to have a great potential for presenting ODV content, thus offering the user an immersive experience for both indoor and outdoor content.

Keywords: Immersion, User Experience, Omnidirectional Video, CAVE, Head-Mounted Displays

1 Introduction

Omnidirectional videos (ODV) have been making their way into the mainstream in the last years. These videos are typically recorded with a set of cameras that cover 360 degrees of the recorded scenery. ODV content has been utilized in several interactive applications, including capturing events such as mountain climbing¹ and musical concerts². As the full contents of these videos cannot be viewed as-is due to the limitations

¹ Mammut #project360. Home page: <https://play.google.com/store/apps/details?id=ch.mammut.project360&hl=en> (Retrieved on 29.7.2015)

² Concert - 360-degree video from ZuZuVideo.
https://www.youtube.com/watch?v=1Kp1_icG328 (Retrieved on 29.7.2015)

in the human field of view, they pose two main design challenges: presentation of the content and interacting with it.

There are several different methods for ODV playback. Often these mediums are some kind of Virtual Reality (VR) applications, ranging from CAVEs (Cave Automatic Virtual Environment) [24] to HMDs [18], but ODV content can also be played with web-based applications (Youtube and Facebook 360 video support) and tablets [33]. In addition to the growing consumer markets, VR applications are used in many domains. For example, they have been found to be a promising tool for treating different kind of phobias, such as acrophobia [6] and agoraphobia [21]. ODV's also have potential in industry use, where they could replace for example 3-dimensional models or content recorded with a single camera, which are often used for demonstrating or training purposes. While numerous interesting solutions and applications exist, thorough understanding of omnidirectional video as a medium and its possibilities in different application domains is yet to be achieved. Our study focuses on iODVs, application that utilize ODV with additional interaction in addition to looking around the scene. This interaction could be, for example, in the form of activating UI elements for more information on different objects in the scene, or transitioning from one ODV scene to another.

One of the most important features of virtual reality applications, also the one's that utilize ODVs, is immersion. For example, in a study by Slater, Alberto and Usoh [27] results indicated that those individuals with a higher sense of immersion achieved better performance overall. The term itself has many definitions in the scientific community, but it is commonly referred to as the feeling of "being there". Our study looked into the differences in the feeling of immersion in two different interactive applications displaying omnidirectional video content – a CAVE system and a HMD application. Both mediums have been studied thoroughly in different contexts but in our study, we wanted to explore these applications further in the context of user experience and immersion. As they are both used extensively, e.g. in industrial use, the results from our study can help in designing future applications. Comparing two different methods of displaying interactive content can be very useful for future designs in this domain. In the two applications we implemented, the user could interact with the environment by activating either exits that took the user to another video or hotspots that offered the user contextual information about the environment. In addition to measuring the sense of immersion, we evaluated the user experience on both applications in order to validate them and to measure the differences in both expectations and experiences between the two systems. The user experience metrics measured the participant's opinion for example on usefulness, pleasantness and clarity of the application. In addition, we compared the different video content types to see if there are any differences between them in the user experience or in the feeling of immersion.

Our main research questions for this study were:

- What are the differences in the user experience between CAVE and HMD applications?

- How immersive are interactive CAVE and HMD applications utilizing omnidirectional videos and are there differences in the level of immersion between these two mediums?

Our findings suggest that the users' experiences exceeded their expectations greatly, especially with the HMD application. The user experience results were very positive in general, and both applications received high scores on the 7-point Likert scale on pleasantness, clearness and performance. One explanation for the contrast between expectations and experiences with the HMD application can be in its "black box" nature, which offers barely any cues on the method of interaction or the overall experience to the user. In the case of CAVE systems, their large size and futuristic look might increase the users' expectations. The positive feedback the HMD application received is also interesting when considering its technical limitations in the presentation of the content: our HMD application offered relatively limited field of view of 60 degrees, which is much more limited than that of the human eye, whereas the CAVE system had no physical limitations on its field of view. Interestingly, none of the users reported this as a limitation.

Regarding immersion, our results indicate that ODV is a very immersive medium. Overall, the HMD application was considered more immersive than the CAVE system with both indoor and outdoor video content. For this difference, we have three explanations: a) HMD obscures the outside world completely from the user, thus allowing them to better focus on the content, b) the sense of depth created by the stereoscopic effect (separate viewports for both eyes), and c) the viewport on the display is based on head orientation, allowing the user to naturally look around.

The motivation for this study stems from the extensive use of CAVE systems in various fields, e.g. in the industry. We argue that HMD systems offer many unique and new application areas requiring immersion, and our results seem to support this argument. The benefits of HMDs come from their portability, as they are often small and mobile, and scalability, as they are less dependent on specific equipment or physical setup. Omnidirectional content could prove useful for example in situations where several people manipulate large objects (such as skylifts) at the same time, as they can show relevant information in multiple directions. CAVE systems also have their uses, for example in situations where the information needs to be presented to multiple persons at the same time.

In the following, we first analyze and summarize the related work in this field of research, which is then followed by a comprehensive description of both applications and their differences. Next, we introduce the methodology used in this study and then report the results of the evaluation along with the discussion on the main findings. We conclude the paper by discussing how these results could be used in designing more immersive interactive ODV applications that offer a better user experience.

2 Related Work

2.1 Immersion in Virtual Environments

The term immersion has many definitions in the scientific community, and there is clearly some discrepancy on what the term actually means. There are no prior evaluations on immersion in interactive ODV applications, and therefore the related work presented below is based on studies on immersion in VR applications. Immersion is an important aspect of virtual reality applications, as it is believed to affect user's behavior with and in these applications [31]. Based on Slater [26], the level of immersion is dependent only on the system's rendering software and display technology. By this definition, immersion is objective and measurable. What some researchers refer to as immersion, Slater defines as presence. According to them, presence is "an individual and context-dependent user response" [26], as in the experience of 'being there'. In short, *immersion* is defined as objective level of sensory fidelity the system provides, whereas *presence* refers to the user's subjective experience and response to the system. Using Slater's definitions, the level of immersion is easier to measure, but restricts the evaluation so that it can be made only on the technological level. This includes only the technical aspects such as field of view (FOV), field of regard (FOR), display size and resolution and the use of stereoscopy. There are several evaluation methods for measuring immersion/presence (based on the definition used), for example the ones by Witmer & Singer [32], Schubert, Friedmann & Regenbrecht [25] and Usoh et al. [30].

Immersion has also been studied extensively in the context of video games, and Brown and Cairns [5] attempted to resolve the disparity with the term. They conducted a qualitative study amongst gamers and talked to them about their experience on playing video games. The study resulted in a grounded theory where immersion was used to describe a person's "degree of involvement with the game". This finding supported the idea that immersion is a cognitive phenomenon. The theory also identified restrictions that could limit the degree of user's involvement, including engagement, engrossment and total immersion.

As the related work shows, immersion can be defined in several ways, depending on many factors such as the emphasis on technology, the research domain and the method of evaluation. With VR related studies, Slater's [26] division of immersion and presence is more prevalent, whereas in video game related studies the term immersion is used more often. In this paper, immersion is referred to as a perceptual phenomenon that is dependent on the individual and the context.

2.2 Omnidirectional Videos

Lot of scientific research has been done to enable the use of omnidirectional video. There exists a large variety of algorithms and devices to capture, construct, project, compress, display and automatically analyze omnidirectional video content.

Application domains, where omnidirectional video has received wider interest include remote operation and telepresence applications [4][20][8], some of which include

automatic situation tracking based on the omnidirectional imagery and directional audio. Another application field identifiable in literature is remote operation of unmanned machines and vehicles, for example drones by using omnidirectional video. Applications where omnidirectional video is used by consumers [17][13][3] provide immersive experiences to cultural contents, e.g., in museums [15][14][19] and theatre [9]. Other application domains include education, e.g., teaching sign language [12], and health care, e.g., relieving stress during medical care [10], and therapy [23]. There has been little research on using ODV in industrial use, for example in demonstrating or training purposes.

From the human-computer interaction perspective, augmenting omnidirectional video with interactive content [2] and UI elements [22] are crucial features in many applications. Another field is multisensory augmentations of video content, e.g., simulated wind [22], to further immerse the viewer and improve sense of presence. Interaction studies have also looked at gesture-based interaction [34][24] and second screen interfaces [33] to interact with omnidirectional video content. For example, Benko and Wilson [1] present the Pinch-the-Sky Dome, which projects a full 360 view of omnidirectional data onto an inner side of a dome-shaped structure. The view is controlled using mid-air gestures from anywhere inside the space, and it supports several simultaneous users. They found that mid-air gestures could enhance immersion in an omnidirectional context.

3 iODV Applications

In this section, we introduce the iODV applications that were built for this study. Both applications used the same ODV content with length of 60 seconds. When the content is finished, it starts again from the beginning. Both applications have two types of user interface elements: exits and hotspots. When activated, an exit takes the user to another video that is linked to that particular exit element, and hotspots provide contextual information about the environment. First, we introduce the video production procedure used for content creation, and then explain the basic features and interaction techniques for both applications. Finally, we compare the main differences between these two applications.

3.1 Video Production

The videos used in this study were recorded with six GoPro 4 cameras attached to a Freedom360 mount on top of a tripod. The resulting six videos from each shot were converted into 4k omnidirectional videos by using AutoPano Video Pro 2 and AutoPano Giga 4 software. Panoramic images and videos are usually divided into either cylindrical (limited vertical field of view – VFOV) or spherical ($360^{\circ} \times 180^{\circ}$) views.

For this study, we produced a total of six videos, three of which were shot indoors, and three in an outdoor environment. Each video was roughly one minute long. Indoor videos were recorded in an industrial hall used for repairing and maintenance of sky-lifts. Each video contains some movement, such as people walking around and working,

and a forklift riding around the hall. Two of the indoor videos were recorded from a top of a ladder to offer a better view of the surroundings. The outdoor videos were recorded in downtown Tampere, Finland. These videos were recorded during quiet hours, but nonetheless contained a relatively large amount of movement, i.e., people walking on the streets.

3.2 cCAVE

For our first experiment, we implemented a multimodal CAVE application, circular CAVE (cCAVE), where the user can explore omnidirectional videos via eight displays set in the form of an octagon. A cylindrical view where the horizontal FOV is 360 degrees and vertical 150 degrees was used in the application. In this system, the user is located at the center of the octagon, sitting on a rotating chair (see Figure 1). The chair has a rotating sensor that sends the rotational axis to the computer. This sensor data is used to update user interface elements on one of the displays, e.g. when the chair is pointing at specific coordinates. The application was developed with Vizard virtual reality software. The omnidirectional video content is then displayed on a 3-dimensional cylinder that is divided between the displays so that each monitor covers 45 degrees of the content.

Each interface element (exits and hotspots mentioned earlier) has a coordinate range (i.e. when the rotating chair is pointed at this range) in which they are shown on the screen. The interface elements are triggered by dwelling, i.e. by focusing an element in the center of the view (by turning the chair towards it) and waiting for five seconds. Dwelling is a relatively common technique for selecting targets with e.g. gaze and mid-air gestures, which is utilized by a number of applications (e.g. [16]). Before the hotspots are activated, they are presented on the screen as blue circles with an exclamation mark inside. Exits are presented as green arrows. During the activation period, the element is scaled up in order to visualize that it is being selected. Users can cancel the activation process by turning away from the element. Similarly, a hotspot dialog is closed by turning away from it.

We used a set of eight Eyevis Eye-LCD 4000 M/W monitors. Each monitor has a screen diagonal of 40 inches with full HD resolution and they were raised 77 cm from the ground. They were 91 cm high, 53 cm wide and 13 cm thick. The bezel between two monitors was 28 mm (14mm in one monitor). These monitors were set up so that they covered an area of 360 degrees around the user. The rotating chair's seat height was adjusted to 50 cm and the distance from the user's head to the monitors was approximately 60 cm. The outer walls of the cCAVE installation were 175 cm wide and 192 cm high. The total resolution for the application was 4320 x 3840 pixels. The monitors were connected to AMD HD 7870 display adapter with 1 GHz processor and 2 gigabytes of GDDR5 memory.



Fig. 1. The cCAVE system. The rotating chair used for interaction is at the center of the system. Eight monitors (only 6 shown in the image) each show 45 degrees of the omnidirectional video content. The two monitors in front are attached to the doors that are opened for entrance and closed during use.

3.3 Amaze360

Amaze360 is an iODV application for HMDs that allows the user to freely observe omnidirectional videos by simply turning one's head in the desired direction. The screen is divided into two separate viewports in order to create a stereoscopic effect, thus creating a sense of depth. This effect is done with the spherical presentation of the video content, as the video content itself is not stereoscopic. The video content used by the application has 360-degree horizontal and 180-degree vertical field of view and the video is projected on a virtual sphere. The viewport's field of view is 60 degrees.

Interface elements (exits and hotspots) in Amaze360 are also triggered by dwelling, but with slight differences. These elements are activated by focusing on an element in the center of the view (by turning the head towards it) and waiting for two seconds. The hotspot and exit icons in Amaze360 are similar to the ones used in cCAVE (blue circle with an exclamation mark inside for the hotspots, and green arrows for the exits). The entire set up and a screenshot of the Amaze360 application with hotspot activation can be seen in Figure 2.

Amaze360 is C# application built on the Unity platform, and it utilizes the Oculus Mobile SDK 1.0.0.0 for iODV features. The application also uses the Easy Movie Texture plugin to enable smooth video playback on mobile devices. It is run on Samsung Note 4 and utilizes the Samsung GEAR headset.



Fig. 2. Top: Amaze360 physical setup. **Bottom:** Amaze360 application view. The video is shown as a stereoscopic presentation. Activated hotspot is shown at the center of screen.

3.4 Differences between the applications

Even though the two applications are intended for the same purpose, there are obvious differences ranging from physical setup and display devices to interaction mechanics. These differences further affected some design choices for both applications. A general overview of the features and differences can be seen in Table 1.

	CAVE	Amaze360
Application Field of View (Horizontal*Vertical, in degrees)	180*150	60*60
Interaction Method	Rotational chair (sensor)	Head/device orientation based activation
Contextual information activation range	X-axis	X- and Y-axis
Contextual information location on the screen	Bottom center	Center
UI Element Activation Time (seconds)	5	2

Table 1. Differences between the two applications

The primary difference between the two applications is in how content is presented – cCAVE shows the ODV in multiple monitors whereas the Amaze360 uses a stereoscopic presentation on a mobile device. In other words, cCAVE always physically displays the full 360-degree view of the content. Therefore, the user sees the content with the full field of view of the human eye. Amaze360, on the other hand, is limited to a 60-degree sector of the content at any given time.

Another major difference is in how the applications are interacted with, i.e. how hotspots and exits are activated. The cCAVE system utilizes the rotation of the chair, and therefore only uses the X axis (chair's rotation relative to the screens) for activating UI elements. Amaze360 relies on head orientation, and hence uses both X and Y axes. For illustration on these differences, see Figure 3.



Fig. 3. The hotspot activation sectors illustrated in both applications. The gray coordinate area represents the coordinate range of hotspots in cCAVE, and the circular area represents the X- and Y- coordinate range used in Amaze360.

Due to the difference in how UI elements are activated, both applications vary in how contextual information is presented. In cCAVE, textual content is shown (when a hotspot is activated) at the bottom of the screen. This design choice was made so that the textual content would not obscure the object it is referring to. In Amaze360, textual information was presented on top of the corresponding hotspot (see Figure 4). This was due to the interaction method: as the user activates hotspots by turning their head towards them, it makes sense that the displayed information is displayed in the same position so that the user does not need to adjust the head once more. Furthermore, this allows closing activated hotspots by turning the head away from them, similar to closing hotspots in cCAVE by rotating the chair to another position.



Fig. 4. Hotspot locations in the two applications. HMD hotspot location is presented in white dotted line and CAVE system hotspot location in black dotted line.

Finally, the activation time for UI elements was also different between the applications because of the conclusions made during pilot testing: a short activation time sometimes caused accidental activations in the CAVE system, whereas with Amaze360 these were not as prevalent. This was caused by the slower interaction with the chair – turning one’s head is much faster and more precise than turning on a chair. The pilot tests verified that the Amaze360 application could have a significantly shorter activation time (2 seconds) for the UI elements than the CAVE system (5 seconds).

4 Experiment of CAVE and HMD

For this study, we conducted two separate experiments which evaluated the user experience, level of immersion and spatial abilities in immersive virtual environments that utilize omnidirectional videos. Experiment 1 was conducted with the CAVE system and Experiment 2 was conducted with a HMD and the Amaze360 application.

4.1 Participants

A total of 34 participants took part in the study, both experiments having 17 participants. The cCAVE was evaluated by 8 females and 9 males aged 30.9 on average ($SD = 5.46$) and the Amaze360 system also by 8 females and 9 males with an average age of 30.7 ($SD = 5.43$). They were recruited from around a university campus and were compensated with a movie ticket for their participation. All participants were naïve with respect to interacting with omnidirectional videos, as in they had not use CAVE, HMD or other type of applications that utilize these type of videos.

4.2 Procedure

In the evaluation scenario the participants were asked to explore the virtual environments that consist of omnidirectional videos. Both indoor and outdoor environments were presented to the user as separate scenarios (one could not move from inside locations to the outside locations, and vice versa). They could move from one location to another after they had spent thirty seconds in one location. The time limitation was set in order to encourage exploration and looking around the scenery instead of just moving quickly from one scenery to another. Each location also contained two hotspots which, when activated, offered contextual information about the object they were referring to. Both indoor and outdoor video content consisted of three different locations and the last location led the user back to the first one, which made it possible for the participant to explore the locations indefinitely.

No specific tasks were given to the participants because we wanted to emphasize the explorative nature of the experiment. This way the participants could concentrate solely on experiencing the virtual environment. The users could use the system under evaluation as long as they wanted to. They informed the researcher when they were finished with each scenario (indoor and outdoor). Participants used each system (both indoor and outdoor scenarios combined) for approximately 10 minutes on average.

In Experiment 1 the participants used the cCAVE system in a laboratory setting while sitting on the rotating chair. In Experiment 2 they used the Amaze360 application also in a laboratory setting while standing and wearing the HMD. Both locations were approximately the same size. For both experiments, conditions were balanced so that half of the users started using the system in indoor locations and the remaining half in outdoor locations. A researcher was present during the procedure for support in case of a technical fault or other disturbance, but did not otherwise intervene with the evaluation.

4.3 Data collection and analysis

We gathered general information from all participants, including age, gender, and experience level with the iODV applications. For the user experience evaluation, we used the SUXES [29] method. It is an evaluation method for collecting subjective user feedback of multimodal systems. In this method, the participants fill out a subjective feedback form about their expectations and experiences on using the system. The form consisted of 9 user experience related claims to which the participants responded on a 7-point Likert scale, where 1 = "Totally disagree", 4 = "Neither agree nor disagree" and 7 = "Totally agree".

Participants filled the expectations form after the user had been informed of the procedure and had been shown to the basics of the system, but before the user personally experienced the system. Then, after they had used the application, users filled out their experiences on a similar form. In addition, after the experiment, participants answered to question regarding their level of immersion during the use of the system ("While using the system, I felt like I was actually standing on the streets/industrial hall"). The same 7-point Likert scale was used for the questions regarding immersion. We decided

to disregard the existing evaluation methods for measuring immersion for practical reasons – our custom-made questionnaire allows us to compare the results with the UX results for different modalities using the SUXES method [29]. Finally, we logged basic interactions with timestamps in both systems, such as start and end times of the application, activations of hotspots, and movements from one video to another. We also considered adding the Santa Barbara Sense-of-Direction questionnaire [11] to the evaluation, but decided against it as the evaluation itself was not about measuring spatial ability.

5 Results

The main research interests in this study were the feeling of immersion and the user experience with the two applications. In addition, we report the results from logged interaction data. For all results, a Bonferroni-corrected independent t-test was conducted to compare the results between the two systems. Here, we treat the disagree-agree-like scale to be equidistant, which is why the t-test for analyzing the results was used. For the statistical analysis, an average UX score of both indoor and outdoor video content was used.

5.1 Expectations versus Experiences

When comparing the UX results of the two experiments, statistically significant differences were discovered between the expectations and the actual user experience on both applications, especially with the HMD. For average UX ratings on all statements in both systems, see Figure 5.

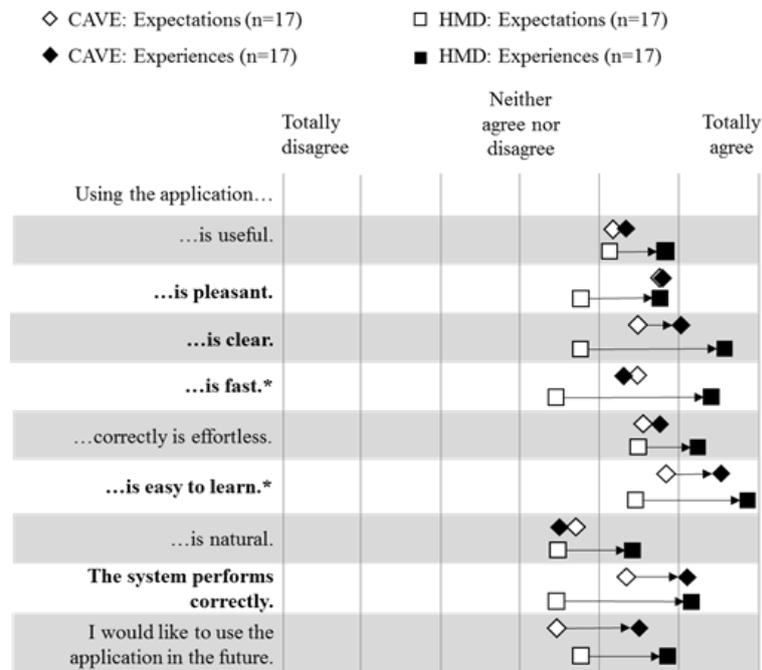


Fig. 5. Average UX ratings for expectations and experiences on both systems. Arrows indicate the direction of the change between expectations and experiences. The statements in bold had statistically significant differences between the applications regarding expectations, and those marked with asterisk in experience.

In almost all metrics the actual use experience exceeded the expectations, especially so with the HMD. Using both systems were considered to be very easy to learn by the participants. All participants except for two in the first experiment and one in the second one agreed (scored either 5, 6 or 7 on the Likert scale) with the statement that the system is useful (Experiment 1, $M = 5.29$, $SD = 1.047$ and Experiment 2, $M = 5.82$, $SD = .883$).

Participants had higher expectations on the cCAVE system used in the first experiment. Statistically significant effects were detected in expectations on pleasantness (Experiment 1, $M = 5.71$, $SD = .920$ and Experiment 2, $M = 4.76$, $SD = 1.251$); $t(32) = 2.499$, $p < 0.05$, and clarity (Experiment 1, $M = 5.53$, $SD = 1.125$ and Experiment 2, $M = 4.71$, $SD = .849$); $t(32) = 2.410$, $p < 0.05$, where the users anticipated more from the CAVE system. cCAVE users were also more optimistic on how fast the system is (Experiment 1, $M = 5.47$, $SD = 1.179$ and Experiment 2, $M = 4.41$, $SD = .939$); $t(32) = 2.896$, $p < 0.05$, and if it performs correctly (Experiment 1, $M = 5.29$, $SD = 1.263$ and Experiment 2, $M = 4.47$, $SD = .874$) was found; $t(32) = 2.210$, $p < 0.05$.

Regarding the user experience, the questionnaire results on both applications were generally positive. With cCAVE, 88 % of the users gave positive feedback (scored either 5, 6 or 7 on the Likert scale) on the system's usefulness. 82 % of the users thought that the system was pleasant to use, and 100 % of the users felt that the use of the system

is easy to learn (where 2.9 % ranked it at 5, 29.5 % ranked it 6 and 67.6 % ranked it at 7 on the Likert scale). The HMD application received even more positive results, where 94 % of the users thought that the system is useful and pleasant to use. Like with cCAVE, all of the HMD users felt that the system is easy to learn.

Comparing the results from the two experiments, some statistically significant findings were discovered. The HMD application ($M = 6.35$, $SD = .862$) was considered to be faster than the cCAVE system ($M = 5.29$, $SD = 1.532$); $t(32) = -2.484$, $p < 0.05$. Participants also felt that the HMD application ($M = 6.88$, $SD = .332$) is easier to learn than cCAVE ($M = 6.53$, $SD = .514$); $t(32) = -2.376$, $p < 0.05$.

5.2 Immersion and System Interaction

The main interest in addition to the user experience was the feeling of immersion experienced during the use. Between the two applications, statistically significant differences were observed with both indoor video content (Experiment 1, $M = 5.18$, $SD = 1.629$ and Experiment 2, $M = 6.18$, $SD = .883$); $t(32) = -2.225$, $p < 0.05$, and outdoor video content (Experiment 1, $M = 5.18$, $SD = 1.510$ and Experiment 2, $M = 6.29$, $SD = .686$); $t(32) = -2.779$, $p < 0.05$. The immersion level of participants for both applications with indoor and outdoor videos can be seen in Figure 6.

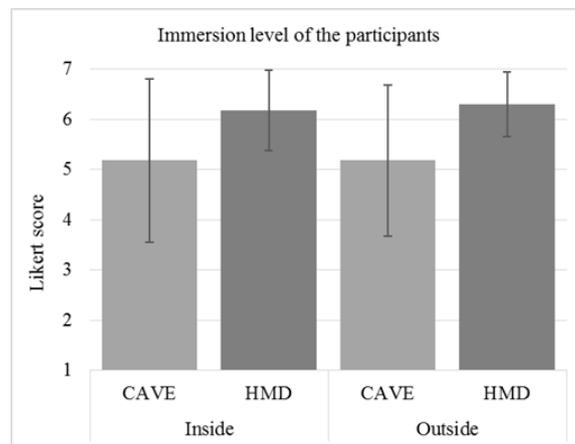


Fig. 6. Average immersion level of the participants in both applications with indoor and outdoor video content

Based on interaction log data, some statistically significant differences in the application use times were observed. cCAVE was used for longer periods of time (in seconds) in total (both outdoor and indoor scenarios combined) than the HMD application (Experiment 1, $M = 884.47$, $SD = 357.91$ and Experiment 2, $M = 561.41$, $SD = 214.52$); $t(32) = 3.193$, $p < 0.05$. Participants also used the CAVE system for longer periods with the indoor video content (Experiment 1, $M = 502.82$, $SD = 303.77$ and Experiment 2, $M = 260.76$, $SD = 92.96$); $t(32) = 3.142$, $p < 0.05$. There was no observed effect with

outdoor video content. The total times spent with both indoor and outdoor video content can be seen in Figure 7.

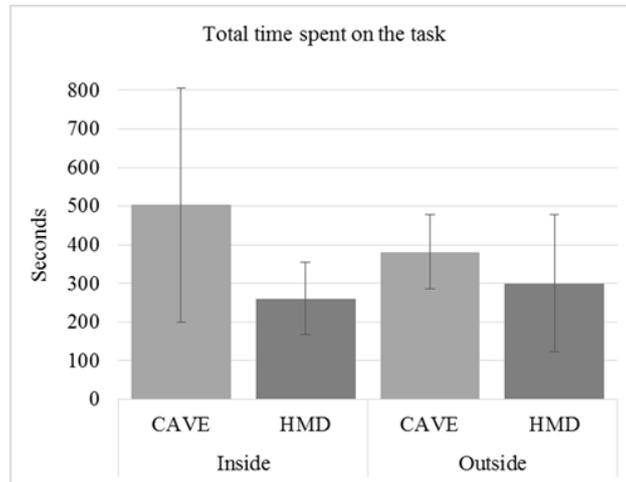


Fig. 7. Total mean time spent on task with indoor and outdoor video content

6 Discussion

6.1 Expectations Versus Experiences

The most interesting finding regarding the user experience was that in almost all metrics the actual use experience exceeded the expectations, especially so in the second experiment with the HMD. The interaction method in the cCAVE experiment can be the reason for the difference in expectations on pleasantness – sitting and interacting with a chair can be expected to be more comfortable for users than standing up while wearing the HMD. In addition, it might be difficult to make any estimates on the pleasantness and clarity on the sort of a “black box” HMD device, which offers no cues on the method of interaction to the user. The cCAVE set up might be more impressive and futuristic looking than HMD devices in general. Another factor to consider is the physical set up of the two applications: cCAVE is a large installation built on a metallic rig with eight monitors, whereas the headset is using the smaller Samsung Gear headset and basic Samsung Note 4 mobile device. The system size difference itself might indicate that the cCAVE is more powerful than the compact HMD device. In addition, a desktop computer can be presumed to have better performance than a smaller mobile device, which might implicate to some participants that the system itself is also better graphics and performance-wise.

The HMD application was considered to be faster than cCAVE, which can be at least partly explained by the interaction method: head turning used with the HMD is much faster to perform than rotating the on cCAVE. As mentioned earlier, both systems were

regarded as easy to learn, but for the HMD this metric was significantly higher. The intuitive method where the viewport is rotated based on the user's head orientation offers the user an efficient way to start interacting with the virtual world immediately after they wear the device. The UI elements draw the user's attention and when they concentrate on these elements, they are activated and animated which again hints the user that something is happening.

The implications of these results are that both CAVE systems and HMD applications utilizing ODVs are regarded as both useful and easy to learn. Both of the applications had very simple interaction methods which were based on dwell-time. This seems to be a meaningful way of interacting with these types of systems, especially when the interaction is kept simple. Nevertheless, more research is required in order to understand the relationship between complex UI elements and different interaction methods.

Overall, the positive feedback on both applications validates their use on this study. The applications were also very robust and had no technical faults during evaluations, which might have also affected the participant's feedback on the user experience. The actual user experience was much more positive than the user's expectations with both systems, but especially so with the HMD application.

6.2 Immersion and User Interaction

Our evaluation suggests that the Amaze360 application is more immersive than the cCAVE with both video content types. There are many possible explanations for this result. First, the headset obscures any other visual stimuli from the view, only showing the contents of the application to the user, whereas in the cCAVE the user can still observe objects outside the screens, including the bezels of the monitors. Second, the HMD provides a stereoscopic effect (coming from the spherical projection, not the video itself) which creates an illusion of depth. This is not provided in cCAVE. Third, since interaction with the Amaze360 is based on head orientation, it does not require any external devices which might enhance the feeling of immersion even further. In the first experiment the aim was to make the interaction as simple as possible with the use of the rotational chair, but it is still not as natural as interaction with the headset. In future implementations a combination of body tracking and gaze tracking could be combined to produce a similar interaction solution as in the HMD application.

Despite the unique advantages of the HMD application, the positive feedback for this application is interesting when considering the current limitations of the technology. For instance, Amaze360 offers a relatively limited 60-degree field of view, which is much smaller than that of the human eye, whereas cCAVE had no such physical limitations. However, none of the users reported this as an issue.

Some cCAVE users had trouble finding the textual content from the bottom of screen even when they were informed about the location beforehand, during the introduction. Participants had no trouble finding or activating the hotspots with the Amaze360 application, but three participants noted that the hotspot text box obscures the visibility of the actual object behind it. One solution for this could be an opaque text box that does not hide the content. These findings indicate that the optimal location for the contextual information is somewhere around the center of the screen where the user is most often

looking at, but that it also should not dominate the viewport. It should also be located close to the actual UI element activating it, so that the user quickly finds it.

One participant using cCAVE remarked that the reflection on the monitors broke the feeling of immersion, as the participant could see the monitors behind him reflected in the monitors in front of him. Four Amaze360 users reported that they felt dizzy during the indoor scenes, which were filmed on a ladder. This interesting finding and its connection to acrophobia could be an interesting topic of research, and has also been looked into by Coelho et. al [6]. None of the participants did not report any motion sickness effects in either applications. Three participants using cCAVE and one user using the Amaze360 stated that the resetting of the omnidirectional video content back to the beginning (due to looping videos) broke the immersion somewhat. In addition, some video production errors that caused distortions were breaking the feeling of immersion for one cCAVE user. These distortions can be eliminated with careful planning of the recording and editing phase of the ODV content. The biggest hurdles in the post-production phase are the color level differences between the cameras, stitching errors where the content between the cameras are not overlapping properly or displaying of the camera equipment in the recording. Also, if the content needs to loop, some attention should be paid to how smoothly the end of the content loops back to the beginning. These problems will likely dissipate once the ODV recording and production technologies advance. When comparing the results between outdoor and indoor video content, there was no significant difference in the feeling of immersion.

The difference in use times with the indoor video content is also an interesting finding. As this same effect was not observed in outdoor environments, one explanation for the difference between the systems could be in the claustrophobic nature of the indoor environment and the limited field of view used in the HMD application. Another explanation could be the filming location of the indoor video content. Two out of three of these videos were recorded from a higher ground, i.e., from a ladder. Four HMD users said that they felt dizzy during these scenes, which might affect the total time used with the indoor videos.

We also note that CAVE systems are diverse and may significantly vary between setups. The cCAVE system was unique but also relatively limited in regards to the rotating chair. It would be interesting to research immersion further with CAVE systems, in particular with larger installations inside which users could walk freely. Also, there are factors that should be taken into account in the future evaluations. For example, evaluating the participant's spatial abilities with Santa Barbara sense-of-direction scale [11] before they use the application.

6.3 Implications for iODV Applications

In the past, CAVE systems have been used extensively in many areas such as the industry [28][7]. However, we argue that HMD systems offer many unique, new application areas because of two reasons. First, due to their small size and easy physical setups, HMDs are easily portable. Second, they are more scalable and adjustable, i.e. less dependent on specific equipment and a specific physical setup. These features could make HMDs a valuable asset in many situations. For instance, we recorded the

indoor video content used in our experiments in a skylift maintenance hall. However, maintenance on skylifts is often conducted in the field. Field technicians could carry HMDs with them and access informative content on-the-spot, in case they needed additional guidance on e.g. how to conduct some specific maintenance procedure on a skylift model unfamiliar to them. We believe omnidirectional video content could prove useful in such situations, as a potentially complicated procedure may be difficult to fully document (and view) on a regular camera, especially if the procedure involves large objects.

7 Conclusion

In this paper, we investigated the user experience and level of immersion in iODV applications that utilize omnidirectional videos. We conducted a comparative study between two applications: a CAVE system, cCAVE, and a head-mounted display application, Amaze360. We collected and analyzed interaction logs and questionnaire data to gain insight on similarities and differences between these two systems and on the feeling of immersion and user experience in iODV applications in general.

Our main findings suggest that in regards to user experience in interactive ODV applications, the experiences exceed the user's expectations. These differences were especially evident with the HMD system, as the users' expectations were exceeded in many aspects such as pleasantness, clarity and performance of the system. Both the CAVE and the HMD applications were considered very easy to learn. Some of the differences in user experience between these two iODV applications can be explained by the different user interaction methods. Head orientation-based interaction used with the HMD is much faster to use than the rotating chair of the CAVE system.

Another interesting take away from our study is that ODV is a very immersive medium. Overall, the HMD application was considered to be more immersive than the CAVE system. This effect was observed with both indoor and outdoor video content. We primarily attribute the immersiveness of the HMD application to a) the head-mount that effectively blocks outside visual stimuli and allows concentration on the content, b) the stereoscopic view creating a sense of depth and c) the viewport on the display is based on head orientation, allowing the user to naturally look around.

As interactive ODV applications are becoming more available in the consumer market, further research on the possibilities of this medium is necessary. For future work, it would be meaningful to study the feeling of immersion on a video content with different heights (skyscraper versus a cave) and different types of background movement (crowded street versus peaceful forest), as these properties were not within the scope of this study. Also, the effect of a moving camera (e.g. a roller coaster or a racing car) and its effects on immersion should be evaluated. This could provide more insight on what kind of ODV content offers the most immersive experience to the user.

8 REFERENCES

1. Benko, H. and Wilson, A. D. (2010). Multi-point interactions with immersive omnidirectional visualizations in a dome. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*. ACM, New York, NY, USA, pp.19-28.
2. Berning, M., Yonezawa, T., Riedel, T., Nakazawa, J., Beigl, M., and Tokuda, H. (2013). 360 degree interactive video for augmented reality prototyping. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication, UbiComp '13*, ACM (New York, NY, USA, 2013), pp. 1471-1474.
3. Bleumers, L., Van den Broeck, W., Lievens, B., and Pierson, J. (2012). Seeing the bigger picture: a user perspective on 360° TV. In *Proceedings of the 10th European conference on Interactive TV and video (EuroITV '12)*. ACM, New York, NY, USA, pp. 115-124.
4. Boult, T. E. (1998). Remote Reality via omnidirectional imaging. In *ACM SIGGRAPH 98 Conference abstracts and applications (SIGGRAPH '98)*. ACM, New York, NY, USA.
5. Brown, E., & Cairns, P. (2004). A grounded investigation of game immersion. *CHI 2004*, ACM Press, pp.1279-1300.
6. Coelho, C.M., Waters, A.M., Hine, T.J. & Wallis, G. (2009). The use of virtual reality in acrophobia research and treatment. *Journal of Anxiety Disorders*, Volume 23, Issue 5, pp. 563-574.
7. Cruz-Neira, C. (1995). Virtual reality based on multiple projection screens: the cave and its applications to computational science and engineering. Doctoral Dissertation, University of Illinois, Chicago, United States.
8. De la Torre, F., Vallespi, C., Rybski, P. E., Veloso, M., & Kanade, T. (2005). Omnidirectional video capturing, multiple people tracking and identification for meeting monitoring. Technical report. <http://repository.cmu.edu/robotics/128/>
9. Decock, J., Van Looy, J., Bleumers, L. and Bekaert, P. (2011). The Pleasure of Being (There?). An Explorative Study into the Effects of Presence and Identification on the Enjoyment of an Interactive Theatrical Performance using Omnidirectional Video. In *Proceedings of the International Society for Presence Research Annual Conference. ISPR 2011*. Edinburgh Napier University, Edinburgh, Scotland, UK.
10. Fassbender, E. & Heiden, W. (2014). Atmosphaeres - 360° video environments for stress and pain management. In: Ma, M., Oliveira, M. & Hauge, J. (Eds.) "Serious Games Development and Applications", Vol. 8778 of *Lecture Notes in Computer Science*, Springer International Publishing, pp. 48-58.
11. Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, 30, pp. 425-448.
12. Järvinen, A. & Ekola, L. (2014). Turning point – A practical assessment of using 360 video in sign language interpreting studies. Diaconia University of Applied Sciences. Degree Programme in Sign Language Interpretation. Bachelor Thesis.
13. Kasahara, S., Nagai, S., Rekimoto, J. (2015). First Person Omnidirectional Video: System Design and Implications for Immersive Experience. In *Proceedings of the*

ACM International Conference on Interactive Experiences for TV and Online Video, June 03-05, 2015, Brussels, Belgium.

14. Kwiatek, K. (2012). How to preserve inspirational environments that once surrounded a poet? Immersive 360° video and the cultural memory of Charles Causley's poetry. In *Virtual Systems and Multimedia (VSMM), 2012 18th International Conference*, pp. 243-250.
15. Kwiatek, K., Woolner, M. (2010). Transporting the Viewer into a 360° heritage story: Panoramic interactive narrative presented on a wrap-around screen. In *Virtual Systems and Multimedia (VSMM), 2010 16th International Conference*, pp. 234-241.
16. Mäkelä, V., Heimonen, T., Luhtala, M., & Turunen, M. (2014). Information wall: evaluation of a gesture-controlled public display. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia (MUM '14)*. ACM, New York, NY, USA, pp. 228-231.
17. Neng, L. A. R. and Chambel, T. (2010). Get around 360° hypervideo. In *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '10)*. ACM, New York, NY, USA, pp. 119-122.
18. Neumann, U., Pintaric, T., and Rizzo, A. (2000) Immersive Panoramic Video. *Proceedings of the Eighth ACM*, pp. 493-494.
19. Okura, F., Kanbara, M., Yokoya, N. (2011). Fly-through Heijo palace site: augmented telepresence using aerial omnidirectional videos. In *ACM SIGGRAPH 2011 Posters (SIGGRAPH '11)*. ACM, New York, NY, USA, Article 78.
20. Onoe, Y., Yamazawa, K., Takemura, H. Yokoya, N. (1998) Telepresence by Real-Time View-Dependent Image Generation from Omnidirectional Video Streams. *Computer Vision and Image Understanding*, Volume 71, Issue 2, August 1998, pp. 154-165.
21. Peñate W., Pitti C.T., Bethencourt J.M., de la Fuente J., Gracia R. (2007). The effects of a treatment based on the use of virtual reality exposure and cognitive-behavioral therapy applied to patients with agoraphobia. *International Journal of Clinical and Health Psychology*, Volume 8, Issue 1, pp. 5-22.
22. Ramalho, J., and Chambel, T. (2013). Windy sight surfers: sensing and awareness of 360° immersive videos on the move. In *Proceedings of the 11th european conference on Interactive TV and video, EuroITV '13*, ACM (New York, NY, USA, 2013), pp. 107-116.
23. Rizzo, A. A., Ghahremani, K., Pryor, L. & Gardner, S. (2003). Immersive 360-degree panoramic video environments: research on creating useful and usable applications. In *Human-computer interaction: theory and practice*, vol 1, J. Jacko and C. Stephanidis, Eds. Lawrence Erlbaum Associates, Mahwah, NJ, USA, pp. 1233-1237.
24. Roveló Ruiz, G. A., Vanacken, D., Luyten, K., Abad, F., and Camahort, E. (2014). Multi-viewer gesture-based interaction for omnidirectional video. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*,

CHI '14, ACM (New York, NY, USA, 2014), pp. 4077–4086.

25. Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, 10, pp. 266-281.
26. Slater M. (2003). A note on presence terminology. *Presence Connect* 3: 3.
27. Slater, M., Alberto, C., Usoh, M. (1995). In the Building or Through the Window? An Experimental Comparison of Immersive and Non-Immersive Walkthroughs, *Virtual Reality Environments in Architecture*, Leeds 2-3rd November 1994, Computer Graphics Society.
28. Talaba, D., Amditis, A. (2008). *Product Engineering: Tools and Methods based on Virtual Reality*. Springer Science & Business Media. pp. 75-77.
29. Turunen, M., Hakulinen, J., Melto, A., Heimonen, T., Laivo, T., and Hella, J. (2009). SUXES – User Experience Evaluation Method for Spoken and Multimodal Interaction. In *Proceedings of Interspeech 2009*. ACM, New York, NY, USA, pp. 109-112.
30. Usoh, M., Catena, E., Arman, S. & Slater, M. (2000). Using Presence Questionnaires in Reality. *Presence: Teleoper. Virtual Environ.* 9, 5 (October 2000), pp. 497-503.
31. Welch, R.B. (1999). How Can We Determine if the Sense of Presence Affects Task Performance? *Presence: Teleoperators and Virtual Environments*. 8(5): pp. 574-577.
32. Witmer, B. G., & Singer, M. J. (1994). *Measuring immersion in virtual environments*. (ARI Technical Report 1014). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.
33. Zoric, G., Barkhuus, L., Engström, A., and Örnevall, E. (2013). Panoramic Video: Design Challenges and Implications for Content Interaction. *Proceedings of the 11th European Conference on Interactive TV and Video*, ACM, pp. 153–162.
34. Zoric, G., Engström, A., Barkhuus, L., Ruiz-Hidalgo, J., and Kochale, A. (2013). Gesture interaction with rich TV content in the social setting. In *Exploring and Enhancing the User Experience for Television*, Workshop of ACM SIGCHI Conference on Human Factors in Computing Systems, CHI'13 (Paris, France, 04/2013 2013).