Hotspot Interaction in Omnidirectional Videos Using Head-Mounted Displays

Pekka Kallioniemi, Tuuli Keskinen, Ville Mäkelä, Jussi Karhu, Kimmo Ronkainen, Arttu Nevalainen, Jaakko Hakulinen, Markku Turunen
Tampere Unit for Computer-Human Interaction (TAUCHI), Faculty of Communication Sciences
University of Tampere
firstname.lastname@uta.fi

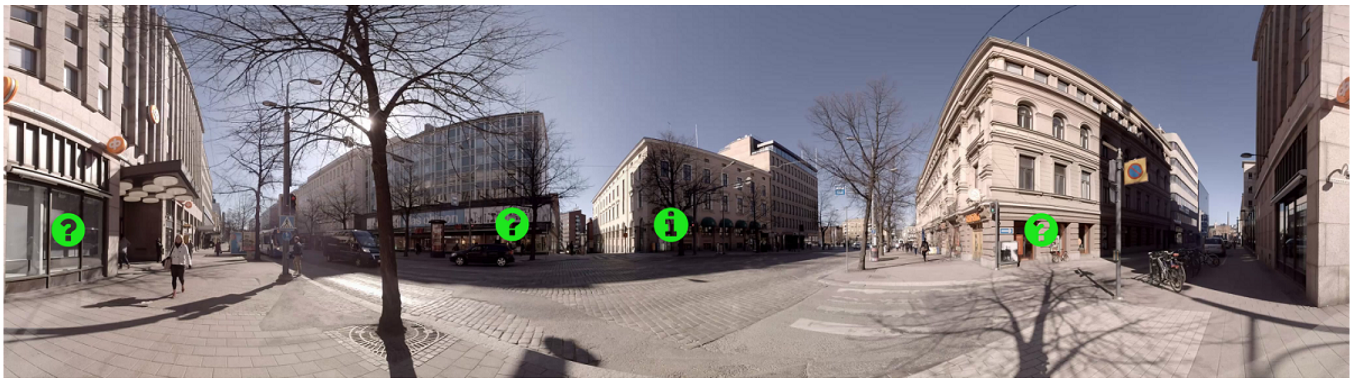

Figure 1: A snapshot from an omnidirectional video displaying four interactive hotspots. When activated, the hotspots display additional information of the buildings they are attached to. The image has been cropped from top and bottom to better fit the media, and the hotspots are larger than in the actual application to maintain clarity.

ABSTRACT

Omnidirectional videos (ODVs) – or 360° videos – are often viewed using a head-mounted display (HMD), allowing users to look around the scene naturally by turning their head. These videos may be populated with visual markers, hotspots, that can contain additional information of points of interest or provide additional functionality. However, optimal ways for interacting and presenting these hotspots remain unclear. We conducted a 16-participant user study where participants interacted with hotspots with different properties while exploring ODVs. We tested 1) two hotspot activation types, *dwell time* and *fade-in*, 2) small and large hotspot icon sizes, and 3) centered and left-edge alignments for the appearing content. We found that a combination of fade-in activation, centered alignment, and large icons generally performs the best, but also that user preferences are diverse. Based on our results, we provide recommendations for optimizing the designs for systems utilizing hotspot interactions with HMDs.

Mindtrek 2018, October 10–11, 2018, Tampere, Finland

ACM ISBN 978-1-4503-6589-5/18/10…$15.00
https://doi.org/10.1145/3275116.3275148

CCS CONCEPTS

• **Human-centered computing → Virtual reality**;

KEYWORDS

Interactive omnidirectional videos; 360° videos; head-mounted displays; virtual environments; virtual reality; hotspot interaction; user experience.

ACM Reference Format:
Pekka Kallioniemi, Tuuli Keskinen, Ville Mäkelä, Jussi Karhu, Kimmo Ronkainen, Arttu Nevalainen, Jaakko Hakulinen, Markku Turunen. 2018. Hotspot Interaction in Omnidirectional Videos Using Head-Mounted Displays. In *Academic Mindtrek 2018 (Mindtrek 2018), October 10–11, 2018, Tampere, Finland.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3275116.3275148

1 INTRODUCTION

The use of omnidirectional video (ODV) - also known as 360° video - has been on the rise in the recent years. These videos can offer an immersive user experience, especially when viewed with a head-mounted display (HMD) [13, 14]. Users can view the video content by looking around naturally, as if they really were at the location.

Despite their immersive nature, omnidirectional videos are static, that is, they cannot be directly interacted with. Therefore, interaction is often presented in the form of *hotspots* [8, 12, 13]. Hotspots are visual markers that are placed on or near points of interest visible in omnidirectional videos (Figure 1). These hotspots can then be activated to launch certain integrated functionalities. For instance, after triggering a hotspot, a text box might appear that gives the user more information of the corresponding point of interest. These types of interactions are the most fundamental features in virtual environments (also in those utilizing ODVs) [1, 5].

A natural and easy way for interacting with hotspots is to utilize the head tracking integrated into HMDs. A common method for activating hotspots with HMDs in this manner is to use *dwell time*, that is, focus the center of the viewport on a hotspot and wait for a fixed period of time [8, 12, 13]. However, it remains unclear if this is the best way, as no other head tracking -based methods are commonly used. It further remains unclear whether other details, such as the size of the hotspots and the positioning of the hotspot content, affect the use and experience of head-tracking interactions.

Therefore, we conducted a user study to investigate the details of head-tracking interaction with HMDs. Our main research question was: *How to design efficient, easy-to-use and pleasant head-tracking interactions?* In particular, we studied three variables:

- *Activation type.* We compared dwell time activation to a fade-in solution, wherein the hotspot's content automatically fades in as the user's focus moves closer to the center of the hotspot. In this study, interacting with hotspots is referred as activation, whereas it is sometimes referred to as selection (see e.g. [9, 10, 17]).
- *Content alignment.* We compared a centered alignment of the hotspot content to an alignment where the content's left edge is on the hotspot, presumably blocking less of the corresponding point of interest.
- *Icon size.* We compared two different sizes of hotspots, to investigate the balance between easy recognition and possible obfuscation of the ODV content.

Our results suggest that overall, a combination of fade-in activation, centered content alignment, and large icons performs the best in terms of efficiency and user experience. However, we also observed that these properties are context-dependent, and that the participants' opinions varied.

Fade-in activation was generally perceived as easier to use and more pleasant than dwell time, but 56% of participants still preferred dwell time over fade-in. Despite this distribution, both activation types received favorable feedback and therefore we argue that both can be used, with the optimal choice depending on the exact need. For the content alignment, most participants preferred the centered alignment in the overall feedback, but this preference was not evident in the condition-based user experience questionnaires. With regards to icon size, users had a slight preference to larger hotspot icons, but they were also reported to sometimes cover too much of the ODV content. The results therefore suggest that while centered alignment and large icons are generally preferred, both can be changed when needed without significantly affecting the users' experiences.

2 RELATED WORK

2.1 Omnidirectional Videos

Although ODVs have been around for a while now (e.g. [5]), they have been researched relatively little until the recent years. As HMDs have become cheaper and more available, many enterprises, such as YouTube and Facebook, have started offering their own platforms for viewing ODVs. This has raised many questions regarding the usability and interaction methods with this type of video content, making them more common research subjects. ODVs have been used in many contexts, including remote vehicle operations [5, 13], cultural contexts [14], collaborative wayfinding tasks [12], and as a communication tool for long-distance couples [23].

Applications that in addition to viewing the video content include more interactive elements, such as user activated hotspots, are called iODVs (interactive omnidirectional videos) [13]. There are many methods for implementing the interaction in these applications, for example, using mid-air gestures [4], second screen interfaces [24], and position-based interaction with a dwell-timer [8, 13] or using a rotating chair [7, 13]. Some guidelines for production and recording of iODV content has already been established, see for example [2] and [22]. These guidelines also include guidance for designing user interface elements.

2.2 Interaction Using Head-Mounted Displays

In this study, we focus on interaction using head orientation. Hence, here we discuss the relevant terms and currently common interaction methods used with head orientation.

A common method for selecting objects in virtual environments (VEs) and iODVs is ray casting [1, 15, 16], where the object that intersects with a virtual ray is selected. This is a natural abstraction of finger pointing in real life, making it an intuitive method of interaction within VEs. All ray casting techniques are essentially 2-dimensional interaction techniques, where the selection is made on the image plane [21]. There are various methods for ray casting selection, including head-directed, gaze-directed, and hand-directed ray casting, as well as combinations of these methods. Modern HMDs utilize the head directed method when external controllers are not used.

Qian and Teather [21] conducted a comparison study between three different eye and head interaction methods. By using Fitt's law and ISO 9241-9, they studied the accuracy and speed of these interaction methods. Based on their results, head-only selection is faster and more accurate than gaze selection and the combination of these two. Regarding the user experience, participants rated head-input higher than gaze on 8 out of 9 items. Hansen et al. [10] came to similar conclusions, but also stated that gaze pointing is less strenuous than head pointing.

To confirm a selection of an object (i.e., to *activate* the object), a common method is to use *dwell time* in combination with ray casting. With dwell time, the ray must remain steadily focused on the target for a period of time before the selection is confirmed. This is done to avoid unintended selections of items that the user is focusing on. This phenomenon, often referred as the "Midas Touch" problem, was grounded by Jacob [11] in their early research. Dwell time has been commonly used as an activation method with head position based input, such as with HMDs [8, 12, 13, 22]. It has also been used in other contexts that commonly utilize ray casting, such as in gaze interaction [9, 17], and freehand pointing [19, 20].

In summary, the combination of head orientation and dwell time is currently the common method to select objects with HMDs, especially when interacting with content embedded in ODVs. In this study, we compare dwell time interaction to a new method, fade-in, that utilizes head orientation based ray casting to automatically

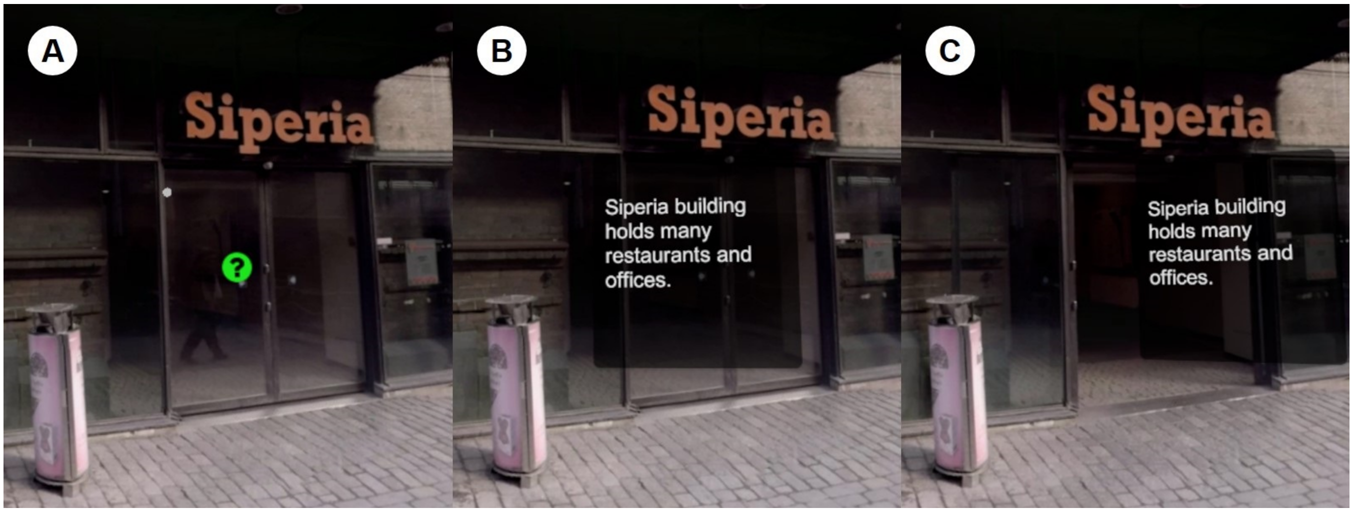

Figure 2: Different parameters: A) The smaller icon; B) Horizontally centered location for the pop-up content; C) Left edge aligned location for the pop-up content.

fade in the content of a hotspot when the head is turned close to it. We hypothesize that fade-in is a faster activation method, but it remains unclear how the automatic opening of content affects the user experience, for instance, whether it is perceived as intrusive.

In addition to the above-mentioned activation types, in this study we also investigate the positioning of the hotspot content as well as the size of the hotspot icons. We are not aware of existing research that deals with these topics in the context of HMDs and ODVs.

3 EXPERIMENT

To investigate different solutions for hotspot interaction on HMDs, we ran a controlled laboratory experiment. For this, an HMD-based omnidirectional video application was built. We examined the activation type, the pop-up content's alignment, and the activation icon size, each having two levels. The experiment was conducted following a within-subjects design.

3.1 Participants

We recruited 16 participants (8 male, 7 female, 1 other) from a summer school course on the fundamentals in human-technology interaction. Either the experiment, or another additional assignment, was a part of the course, and thus, the participants received no extra compensation for their participation. The participants were 19–37 years old (M=24.9, SD=5.17) with a range of reported mother tongues (5 Finnish, 11 other). The self-reported English skills were basics for 2, good for 4 and fluent for 10 of the participants. Half (8) of the participants were studying human-technology interaction or similar field, 2 computer science or similar, 2 social sciences, and 4 other fields. Their earlier experience with omnidirectional videos as well as with virtual reality technologies in general was rather limited: 12 had watched ODVs and 10 had used VR technologies a couple of times at most before the experiment. Not a single participant reported to have more earlier experience than this.

3.2 Application and Parameters

We built a virtual reality application that displays 360-degree videos in a sphere around the user. Users can rotate their heads around naturally, and the viewport will turn accordingly. The application was built on the Unity platform using the SteamVR SDK. The application supports populating omnidirectional video scenes with interactive hotspots and supports various designs for hotspot interaction (see Figure 2 and Figure 3 for examples). In the following, we describe the details and implementation of the three evaluated parameters, the properties of which were decided based on a brief pilot test, and on related work when possible.

Activation Type. We tested *dwell time* and *fade-in* activation. Dwell time started (and kept running) when the head cursor's distance to a hotspot icon's center was 2.8 degrees or less. Dwell time was set to 2 seconds like in previous work [13]. The progress of the activation was represented as a filling ring. Before the actual activation area was reached and the activation started, the head cursor (a white dot at the center of the viewport) became visible when the distance to the icon's center was less than 12 degrees. Figure 3 shows a visualization of the dwell time activation.

Fade-in activation started when the head cursor's distance to the icon's center was 16 degrees or less, and the maximum visibility (alpha 0.8) of the pop-up content was reached when the distance was 9.6 degrees or less. The pop-up content was also gradually faded out after the head cursor's distance from the center grew. With fade-in, all activation and closing settings, i.e., the triggering distances between the head cursor and the icon's center, were the same regardless of the icon size.

With both activation types, the opened content closed when the head cursor's distance from the center of the activation icon grew larger than 16 degrees. In the experiment, all hotspots were located on the same horizontal plane.

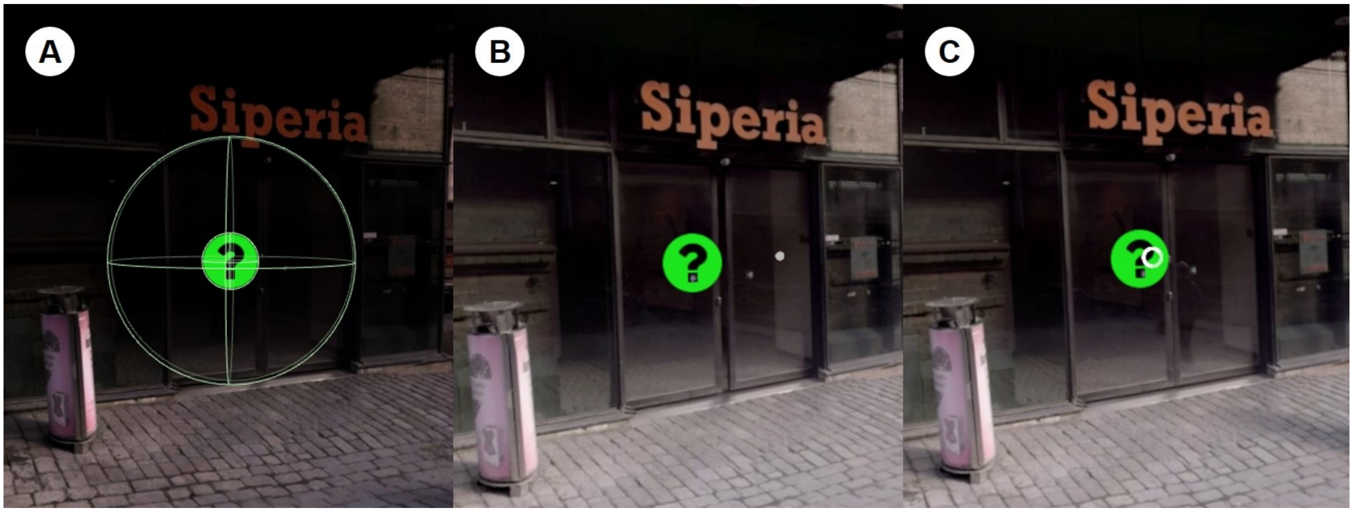

Figure 3: Visualization of the dwell time activation: A) The large icon, the outer limit circle (not visible to the user), and the inner limit circle (in this example the inner circle is at the hotspots icon's border); B) The head cursor becomes visible (the white dot on right side of the icon) when the head position is within the outer limit circle; C) The dwell time starts to run when the head position is within the inner limit circle and is represented as a filling ring (overlaid on the icon).

Table 1: The conditions and the related parameters.

Condition	Activation	Alignment	Icon
D-C-S	Dwell Time	Centered	Small
D-C-L	Dwell Time	Centered	Large
D-L-S	Dwell Time	Left	Small
D-L-L	Dwell Time	Left	Large
F-C-S	Fade-in	Centered	Small
F-C-L	Fade-in	Centered	Large
F-L-S	Fade-in	Left	Small
F-L-L	Fade-in	Left	Large

Content Alignment. For the pop-up content's location, we tested horizontally centered and left-edge aligned content with all activation types (Figure 2B-C) in relation to the activation icon's center point. The pop-up content's and the activation icon's centers were always vertically aligned.

Icon Size. We tested small and large icons, the properties of which were decided with brief pilot tests. The small icon's diameter was 2.8 degrees and the large one's diameter was 5.6 degrees. The small icon can be seen in Figure 2A and the large icon in Figure 3.

3.3 Materials

The three parameters each having two levels resulted in a 2x2x2 within-subjects design, and eight different conditions. Details can be seen in Table 1. The order of the conditions between participants was counterbalanced using a Latin square.

For each condition, there were four trials with four different videos. The same set of four videos was used for all conditions, only the parameters varied. The videos were 2D omnidirectional videos without camera movement. They were shot in different locations in a city center. Audio from the captured videos, i.e., ambient city sounds, was included and played back in low volume during the trials.

For each video, four hotspots were added in seemingly randomized horizontal positions from the user's point of view to induce actual exploration. The vertical positions varied only a little below and above approximate gaze level. In each trial, one of the hotspots was the correct target, indicated by the letter 'i', and the other three were indicated by question marks. The correct activation icon per video was varied across conditions. For each hotspot, there was about one sentence of real, textual information which the participant read out loud after activating the (correct) icon.

The experiment was conducted in a laboratory setting using a head-mounted display, HTC Vive with Tobii eye-tracking functionality and integrated Vive Deluxe Audio Strap.

3.4 Procedure

In the beginning of the sessions, participants were given a brief introduction to the upcoming study tasks. After the introduction, the participants sat on a free-rotating chair, put the HMD on and made the needed adjustments. Then, participants had a chance to try moving their head and rotating the chair, i.e., practice exploring the VR environment with a non-interactive example video.

During the tasks, participants experienced and interacted in all eight conditions in a counterbalanced order to complete simple activation trials. For each condition (see the previous subsections for details) there were four trials, where the participants explored the video to locate and activate the correct icon and read out loud the appearing text. After the participant had read the correct text, the moderator pressed a dedicated button on a keyboard to move to the next trial.

The participants answered statements after each condition. Participants answered them verbally so that the HMD could be kept on and thus save time. If the participants felt uncomfortable they were allowed to remove the headset, though. The statements were read to the participants one by one, and the participants gave their answer as a number from 1 to 5 based on how much they agreed with the statement. While collecting the feedback, the participants were shown the same non-interactive video as in the practice phase, but with a scale figure indicating that number 1 corresponds to "Totally disagree", number 3 corresponds to "Neither agree or disagree", and number 5 corresponds to "Totally agree". In verbal instructions by the moderator, also the intermediate options, 2 for "Somewhat disagree" and 4 for "Somewhat agree", were mentioned.

After all the conditions were complete, participants filled out an online questionnaire concerning, e.g., their background information, and overall user experience and preferences regarding the experienced conditions. The experiment took an average of 45 minutes per participant. All material as well as verbal discussions during the sessions were in English.

4 RESULTS

4.1 Condition Experiences

We investigated whether there are statistically significant differences in the user experience statements between the eight conditions. Due to the data being based on several repeated measures and of ordinal scale, and also not normally distributed, non-parametric Friedman's test was used. Statistically significant ($p<0.05$) differences between the conditions were found in three of the five user experience statements inquired after each trial set: ease of use, speed, and pleasantness of opening the textboxes. Boxplot presentations of these statements can be seen in Figures 4-6.

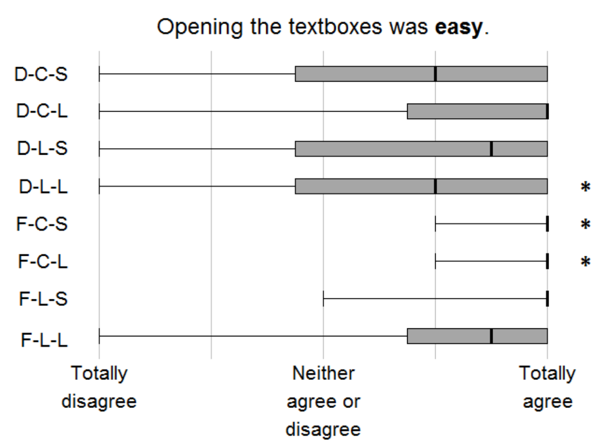

Figure 4: Participants' user experiences about the ease of opening the textboxes. The whiskers represent the minimum and maximum answers, the grey boxes represent the interquartile ranges and the thick, black vertical lines represent the medians. Conditions which showed statistically significant differences in the Friedman's test are marked with asterisks.

The statement "Opening the textboxes was easy" (Figure 4) revealed statistically significant differences in two pairwise comparisons: Fade-in, centered with both small (F-C-S) and large icon (F-C-L) was rated higher than dwell time, left aligned and large icon (D-L-L). In fact, for fade-in, centered conditions only 2/16 participants gave it the already high minimum answer of 4, i.e., somewhat agreed with the statement, so a clear majority agreed that opening the text boxes was easy.

Unsurprisingly, for the dwell time based conditions opening the textboxes was considered slower than the fade-in based conditions (Figure 5). The statement regarding speed was statistically significantly less agreed upon for D-C-S as compared to fade-in, centered conditions regardless of the icon size (F-C-S and F-C-L). Also, for D-C-L, D-L-S, and D-L-L the statement was less agreed as compared to fade-in, centered conditions (F-C-S and F-C-L), but also compared to fade-in, left aligned, small icon (F-L-S). In addition, for fade-in, left aligned, large icon (F-L-L) opening the textboxes was considered statistically significantly slower than for all other fade-in based conditions (F-C-S, F-C-L, and F-L-S).

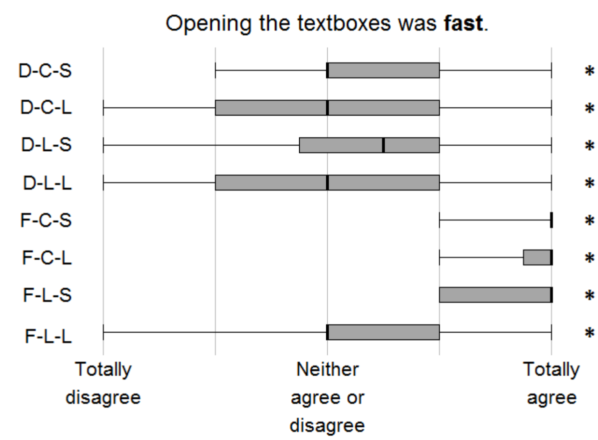

Figure 5: Participants' experiences about the speed of opening the textboxes. (See Figure 3's caption for representation details.)

The answers for the statement "Opening the textboxes was pleasant" can be seen in Figure 6. Friedman's test did not identify any statistically significant differences in the pairwise comparisons of the conditions, but it showed there are some statistically significant differences overall. Also here, the trend seems to be that the statement was agreed more regarding the fade-in based conditions, but the differences and answer distributions are not as clear as for the ease of use and speed of opening the textboxes.

Considering the additional two statements inquired about for each condition – "The textboxes were situated logically" and "Finding the correct icons was easy" – there were no statistically significant differences between the conditions. For both of these the median experience was rather high, 4–5, regardless of the condition. For the logical placement of the textboxes the minimum agreeing level was 2 (somewhat disagree) only for one condition, while it was 3 (neither agree or disagree) for all other seven conditions, meaning that the participants' answers were heavily on the neutral–agreeing side and the selected textbox locations (centered or left-aligned) in relation to the hotspot are both logical. Also, based on the participants' subjective experience ratings, the variation between the small and large icon was not clear enough to show statistically significant differences between the conditions in how easy it was to find the correct icons.

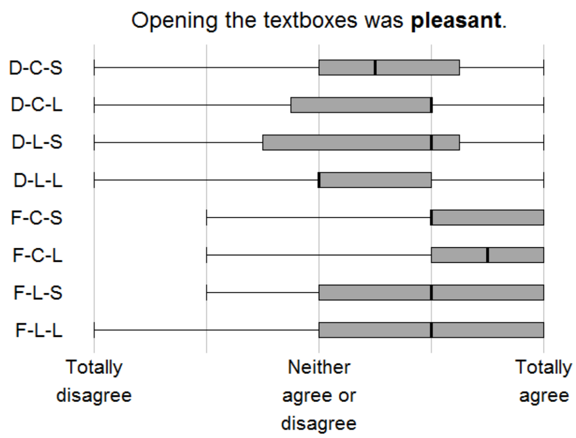

Figure 6: Participants' experiences about the pleasantness of opening the textboxes. Friedman's test revealed there are statistically significant differences between the conditions but none were identified in pairwise comparisons. (See Figure 3's caption for representation details.)

4.2 Overall User Experiences

Considering the subjective parameter preferences, the pop-up content's location showed the clearest result as 13/16 (81%) participants preferred the centered alignment. For the activation icon size, 10 out of 16 (63%) participants reported to prefer the large icon. Some participants reported that it was easier to detect and activate the large icons, but some also said that they were too big as they covered the environment more. The activation type distributed opinions more evenly: a slight majority, 9/16 (56%) participants, preferred the dwell time over fade-in. Based on the open question answers, the participants considered fade-in to be generally faster – both in positive and negative as some preferred it due to its speed, and some considered it to be too fast. Some mentioned that dwell time, in turn, was too slow and some that it was more rewarding. It was also reported by a couple of participants that dwell time is better because icons are not as easily activated by mistake.

The overall assessments about the virtual experience can be seen in Figure 7. The participants generally thought exploring the virtual environments was interesting and the virtual experience was pleasant as a whole, both statements reaching a median of 5 (totally agree) and having only a few respondents on the disagreeing–neutral side. Also, most participants would be interested in exploring places around the world with such VR technologies, but here the answers were distributed more widely. Some mentioned about blurriness in the content which indicates that the HMD may not have been properly adjusted. These issues, including the blurring of the lenses due to sweating, were also brought up by Kallioniemi et al. [13].

Taking into account the participants' very limited earlier experience with VR this is well possible, despite the instructions by the moderator. The blurriness may also explain the results considering nausea. Although the answers were mainly on the disagreeing–neutral side, they distributed across the scale. This somewhat conflicts with the findings of Hakulinen et al. [8] on their ODV-based rally simulator experienced with an HMD, for example. Over 75% of their nearly 300 respondents totally disagreed with a similar statement. It is worth noting, however, that Hakulinen et al. used a different technical setup for their experiment.

4.3 Performance and Error Rates

In addition to subjective feedback, we measured the speed and error rates of the techniques. We measured the speed from when a task began (when a video and the hotspots were shown), until the participant had located and activated the correct hotspot and read out loud the revealed content. We did this because of the fundamental differences between the two activation types. With fade-in activation, simply revealing the content once is not necessarily a reliable indication of successful activation. Rather, a more definite indication is when the user is able to see and read the revealed content. Nonetheless, this method of measurement may introduce a

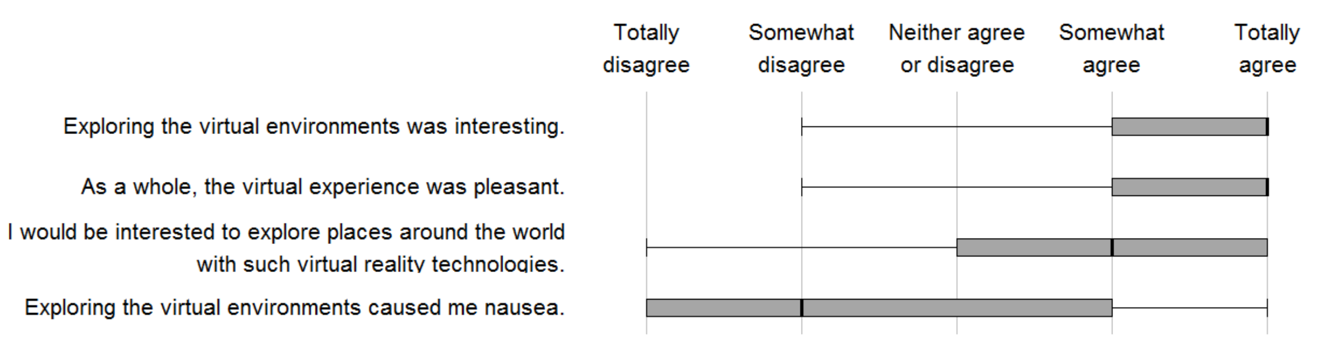

Figure 7: Participants' assessments about the virtual experience in general. The whiskers represent the minimum and maximum answers, the grey boxes represent the interquartile ranges and the thick, black vertical lines represent the medians.

margin for error, as it is also dependent on, e.g., the speed at which the user reads the text.

Task completion times are presented in Figure 8. A repeated measures ANOVA with a Greenhouse-Geisser correction revealed a significant main effect of the used condition on task completion time ($F(1.934, 100.558) = 281.038$, $p < 0.0005$). Pairwise comparison with Bonferroni correction showed statistically significant differences between around half the pairs. Most importantly, F-C-L was significantly faster than the other techniques ($p < 0.05$), expect for D-C-S, the difference for which was near-significant ($p = 0.052$). The slowest technique was F-L-L, followed by D-L-S, D-C-L, and D-L-L. There were no significant differences between these four pairs.

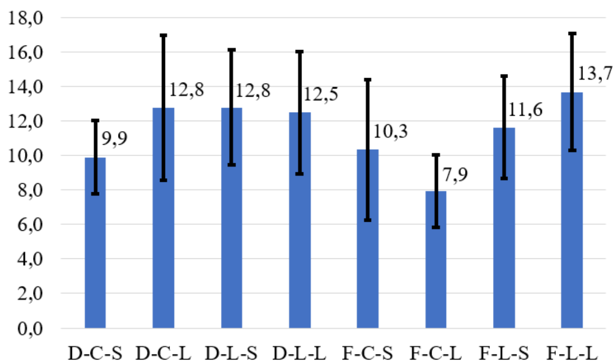

Figure 8: Average task completion times and standard deviations for all conditions (x-axis: conditions; y-axis: time in seconds).

With the four conditions using dwell time activation, users did not trigger incorrect hotspots at all. For the remaining conditions that utilized fade-in activation, we argue that it makes little sense to measure errors in a similar manner. Because fade-in activation automatically shows the hotspot content as the user's orientation draws closer, then naturally hotspots that the users cross on their way to the intended target are also opened. As such, these instances should not be treated as errors per se.

5 DISCUSSION

Overall, the combination of fade-in activation, centered content, and large icons (F-C-L) performed the best. However, a single solution is unlikely to work in every context. Fortunately, our tested techniques generally performed well and received favorable feedback, and therefore we argue that there is room for tailoring our tested properties to the specific needs of the application. Next, we provide a more detailed discussion on the choices regarding hotspot activation, content alignment, and icon size, after which we discuss the results on a larger scale.

Hotspot Activation. Fade-in activation was faster than dwell time and received slightly higher scores in terms of ease of use and pleasantness, although the differences were mostly not statistically significant. User preferences were distributed rather evenly, as 56% preferred dwell time over fade-in. Therefore, we argue that the choice between the two is context-dependent, and the two techniques might even work in the same application for different purposes. Fade-in is the better choice when efficiency is a priority. Dwell time may be more suitable when the hotspot content might interfere with the view or other hotspots, for instance, when several hotspots are placed in close proximity.

Despite the positive results of the fade-in technique, it is unlikely suited for other functionalities besides displaying additional information. Because fade-in lacks distinct activation behavior, it is prone to the Midas touch issue [11], which was also pointed out by our study participants. Fade-in should therefore not significantly change the state of the application. Nonetheless, our results suggest that fade-in is a valid technique for activating information hotspots using HMDs. We are not aware of any other work that utilizes fade-in in HMD interaction, while the other tested technique, dwell time, has been used actively (e.g. [8, 12, 13]).

We observed a surprising amount of variance in the speed of the tested designs, especially with those utilizing fade-in activation. Indeed, F-C-L was the fastest design, while F-L-L was the slowest. This suggests that the speed of the fade-in technique is sensitive to other aspect of the design. For instance, it may be that for content that is not aligned with the target hotspot, it is more difficult to focus correctly on the hotspot. Therefore, there is room for investigating the fade-in technique further.

Content Alignment. A clear majority of participants preferred the centered alignment of content to the left alignment (81%). Therefore, the centered alignment is the safe option in most cases. However, no clear differences were found in the condition-based user experience questionnaires. Hence, our results suggest that different alignments can be used when needed without significantly affecting the experience, e.g., when the opened content would otherwise cover something important. Note, however, that content alignment seemed to affect the speed of the fade-in technique.

Icon Size. Users somewhat preferred the large icons over small ones (63%). In the condition-based user experience questionnaires, large and small icons performed evenly. We note that the large icons were larger only visually, i.e., they were not easier to activate. Hence, the preference for large icons is likely explained by them being easier to notice, as pointed out by participants. However, it was also reported that they can sometimes cover too much of the scene. In summarizing the results, we conclude that emphasis should be put on larger icons, but again, as the overall differences are not great, it is safe to use small icons when needed, e.g., to prevent the icon from covering important parts of the scene, or when several hotspots are close together. Michalski, Grobelny and Karwowski [18] came to the same conclusion for more traditional displays and mouse interfaces in their experiment.

Impact of the Results. Our results provide insight into hotspot interactions with head-mounted displays, by showing the effect of different properties on performance and user experience. This work has two primary contributions. First, we provide the *fade-in* technique, a clearly valid alternative for activating information hotspots with HMDs, as opposed to the currently prevailing method of dwell time. Second, we show that even though larger icons and centered hotspot content are generally preferred, changing them has little effect on the user experience, which suggests that there is room for tailoring these properties to the application's specific needs.

We argue that this research is important because interaction by turning one's head is fundamentally different from other, more traditional ways of interaction. Moreover, the output form of a HMD is different from, e.g., a standard display or a tablet. Therefore, it is not clear whether existing research revolving around similar properties in different platforms holds in HMD interaction. For instance, in traditional desktop settings, visual search on the vertical plane is more efficient than on the horizontal plane [3]. This is unlikely true for head tracking, as turning the head horizontally is more natural and less stressful. In addition, when HMDs are used to view ODVs especially with urban content, horizontal exploration generally makes more sense than vertical exploration.

An interesting discussion is to what extent our results apply beyond omnidirectional videos. We believe that our results apply well in other types of virtual environments that are explored using a HMD, if the user's viewpoint and the hotspots remain stationary. However, when locomotion is introduced, it likely influences how easily users can focus on the hotspots, and how well the displayed content can be discerned. This discussion can also be approached from the point of view of ODVs – how would the results apply when ODVs are viewed on a different platform such as a tablet? Because of the static nature of ODVs, the need for hotspots as a way to present more functionality and information still applies. Therefore, it may be that the results regarding the icon size and content alignment hold relatively well, whereas the head tracking based activation techniques (dwell time and fade-in) are not needed in environments that have more explicit and direct ways for selecting objects, such as touch.

It is also worth noting that interaction in this experiment was limited to the activation of individual icons and did not include more complex user interface structures such as menus or sliders. These solutions and their interaction have been explored in previous work, for example by using gestures for multi-level menu interaction [6]. This type of menu structures could also be used when multiple hotspots are located in a small area in order to avoid accidental activations, as the menu selection could be used to activate the desired hotspot. These types of more complex selections in head position based interaction would be an interesting subject for future studies.

6 CONCLUSIONS

The study presented in this article examined visual markers called hotspots and their activation in applications that utilize omnidirectional videos (ODVs) and head-mounted displays (HMDs). We conducted a user study where we evaluated 1) two hotspot activation types, dwell time and fade-in; 2) small and large hotspot icon sizes; 3) centered and left-edge alignments for the appearing content.

Overall, a combination of fade-in activation, centered alignment, and large icons performed the best in terms of speed, user experience and user preferences. However, we also found that user preferences vary, allowing the designers to tailor the tested properties to suit their needs. Fade-in activation was perceived as easy-to-use and pleasant, but 56% of participants still preferred dwell time. Our results also suggest that content alignment does not significantly affect users' experiences. Finally, users had a slight preference to larger icons, but they were also reported to occasionally cover too much of the ODV content. Therefore, icon size can be changed depending on the context of use. These findings help in the design of successful hotspot interactions for future head-mounted display applications, especially when utilizing omnidirectional videos.

ACKNOWLEDGMENTS

This work was done as a part of *Immersive Media Disruption* (IMD) project, financed by Tekes — the Finnish Funding Agency for Innovation (later Business Finland). We thank all the participants.

REFERENCES

[1] Ferran Argelaguet and Carlos Andujar. A survey of 3d object selection techniques for virtual environments. *Computers and Graphics*, 37(3):121 – 136, 2013. ISSN 0097-8493. doi: https://doi.org/10.1016/j.cag.2012.12.003. URL http://www.sciencedirect.com/science/article/pii/S0097849312001793.

[2] L. Argyriou, D. Economou, V. Bouki, and I. Doumanis. Engaging immersive video consumers: Challenges regarding 360-degree gamified video applications. In *2016 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS)*, pages 145–152, Dec 2016. doi: 10.1109/IUCC-CSS.2016.028.

[3] Richard W Backs, Larry C Walrath, and Glenn A Hancock. Comparison of horizontal and vertical menu formats. In *Proceedings of the Human Factors Society Annual Meeting*, volume 31, pages 715–717. SAGE Publications Sage CA: Los Angeles, CA, 1987.

[4] Hrvoje Benko and Andrew D. Wilson. Multi-point interactions with immersive omnidirectional visualizations in a dome. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 19–28, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0399-6. doi: 10.1145/1936652.1936657. URL http://doi.acm.org/10.1145/1936652.1936657.

[5] Terrance E. Boult. Remote reality via omni-directional imaging. In *ACM SIGGRAPH 98 Conference Abstracts and Applications*, SIGGRAPH '98, pages 253–, New York, NY, USA, 1998. ACM. ISBN 1-58113-046-5. doi: 10.1145/280953.282215. URL http://doi.acm.org/10.1145/280953.282215.

[6] Matthew M Davis, Joseph L Gabbard, Doug A Bowman, and Dennis Gracanin. Depth-based 3d gesture multi-level radial menu for virtual object manipulation. In *Virtual Reality (VR), 2016 IEEE*, pages 169–170. IEEE, 2016.

[7] Jan Gugenheimer, Dennis Wolf, Gabriel Haas, Sebastian Krebs, and Enrico Rukzio. Swivrchair: A motorized swivel chair to nudge users' orientation for 360 degree storytelling in virtual reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1996–2000. ACM, 2016.

[8] Jaakko Hakulinen, Tuuli Keskinen, Ville Mäkelä, Santeri Saarinen, and Markku Turunen. Omnidirectional video in museums–authentic, immersive and entertaining. In *International Conference on Advances in Computer Entertainment*, pages 567–587. Springer, 2017.

[9] John Paulin Hansen, Anders Sewerin Johansen, Dan Witzner Hansen, Kenji Itoh, and Satoru Mashino. Command without a click: Dwell time typing by mouse and gaze selections. In *Proceedings of Human-Computer Interaction–INTERACT*, pages 121–128, 2003.

[10] John Paulin Hansen, Vijay Rajanna, I Scott MacKenzie, and Per Bækgaard. A fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In *Proceedings of the Workshop on Communication by Gaze Interaction*, page 7. ACM, 2018.

[11] Robert JK Jacob. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 11–18. ACM, 1990.

[12] Pekka Kallioniemi, Tuuli Keskinen, Jaakko Hakulinen, Markku Turunen, Jussi Karhu, and Kimmo Ronkainen. Effect of gender on immersion in collaborative iodv applications. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, pages 199–207. ACM, 2017.

[13] Pekka Kallioniemi, Ville Mäkelä, Santeri Saarinen, Markku Turunen, York Winter, and Andrei Istudor. User experience and immersion of interactive omnidirectional videos in cave systems and head-mounted displays. In *IFIP Conference on Human-Computer Interaction*, pages 299–318. Springer, 2017.

[14] Karol Kwiatek and Martin Woolner. 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 on*, pages 234–241. IEEE, 2010.

[15] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. *3D user interfaces: theory and practice.* Addison-Wesley Professional,

2017.

[16] Sangyoon Lee, Jinseok Seo, Gerard Jounghyun Kim, and Chan-Mo Park. Evaluation of pointing techniques for ray casting selection in virtual environments. In *Third international conference on virtual reality and its application in industry*, volume 4756, pages 38–45. International Society for Optics and Photonics, 2003.

[17] Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. Fast gaze typing with an adjustable dwell time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 357–360. ACM, 2009.

[18] Rafal Michalski, Jerzy Grobelny, and Waldemar Karwowski. The effects of graphical interface design characteristics on human–computer interaction task efficiency. *International Journal of Industrial Ergonomics*, 36(11):959–977, 2006.

[19] Ville Mäkelä, Tomi Heimonen, Matti Luhtala, and Markku Turunen. Information wall: Evaluation of a gesture-controlled public display. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia*, MUM '14, pages 228–231, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-3304-7. doi: 10.1145/2677972.2677998. URL http://doi.acm.org/10.1145/2677972.2677998.

[20] Ville Mäkelä, Jobin James, Tuuli Keskinen, Jaakko Hakulinen, and Markku Turunen. "It's natural to grab and pull": Retrieving content from large displays using mid-air gestures. *IEEE Pervasive Computing*, 16(3):70–77, 2017. ISSN 1536-1268. doi: 10.1109/MPRV.2017.2940966.

[21] Yuan Yuan Qian and Robert J Teather. The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pages 91–98. ACM, 2017.

[22] Santeri Saarinen, Ville Mäkelä, Pekka Kallioniemi, Jaakko Hakulinen, and Markku Turunen. Guidelines for designing interactive omnidirectional video applications. In *IFIP Conference on Human-Computer Interaction*, pages 263–272. Springer, 2017.

[23] Samarth Singhal and Carman Neustaedter. Bewithme: An immersive telepresence system for distance separated couples. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 307–310. ACM, 2017.

[24] Goranka Zoric, Louise Barkhuus, Arvid Engström, and Elin Önnevall. Panoramic video: design challenges and implications for content interaction. In *Proceedings of the 11th european conference on Interactive TV and video*, pages 153–162. ACM, 2013.