An Analysis of User Perception Regarding Body-Worn 360° Camera Placements and Heights for Telepresence

Kevin P. Pfeil
University of Central Florida
kevin.pfeil@knights.ucf.edu

Pamela J. Wisniewski
University of Central Florida
pamwis@ucf.edu

Joseph J. LaViola Jr.
University of Central Florida
jjl@cs.ucf.edu

ABSTRACT
Our work investigates body-worn 360° camera placements for telepresence, to balance height and clarity of view. We conducted a user study in a Virtual Reality (VR) simulation, using a 3x3 within-subjects experimental design varying placement and height, with 26 participants. We found that shoulder mounted cameras were significantly less preferable than our other conditions due to the occlusions caused by the wearer’s head. Our results did not show a significant effect of camera height within a range of +/- 12 inches from the user’s natural height. As such, in the context of body-worn 360° cameras, there is leeway for camera height, whereas strategic bodily placements are more important. Based on these results, we provide design recommendations for content creators using wearable cameras for immersive telepresence.

CCS CONCEPTS
• Human-centered computing → User studies; Empirical studies in HCI.

KEYWORDS
360 Camera, Panoramic Camera, Telepresence

ACM Reference Format:

1 INTRODUCTION
Live-streamed video is an increasingly popular content medium. In 2015, Periscope hit a 10 million account milestone [Team 2015], and in 2017, Twitch.tv saw 2 million monthly broadcasters [Twitch 2017]. 360° video streaming is a new technology that is envisioned as the “next big thing” [Bajarin 2015; Graham 2016]. Viewers often watch these panoramic videos via social media and streaming websites [Stout 2018], but there is an opportunity to create an immersive telepresence experience by sharing live-streamed content to users of Virtual Reality (VR) head-mounted displays (HMD). There are a number of use cases where this might be useful. For instance, a geographically distributed family could reconnect with their elderly [Wu et al. 2017]. A person suffering from social anxieties could remotely explore the world with their friends [Rivera et al. 2015]. A bed-bound patient could enjoy a day sight-seeing with their loved ones [Mosadeghi et al. 2016].

We expect this kind of telepresence to become commonplace, enabling people from around the globe to connect like never before. The HCI community has begun to issue recommendations and best practices regarding panoramic video, but there is still work to be done, particularly to optimize viewpoints. In this paper, we aim to identify how various body-worn camera placements and heights affect user experience. While the community has developed an array of prototypes, we are the first to specifically compare and contrast viewer experience among exemplar camera placements. Therefore, our research questions for this work include the following:

- RQ1: What is the optimal body-worn camera placement for reducing visual occlusions created by the video streamer?
- RQ2: What is the optimal body-worn camera height, in relation to the viewer’s own height?
- RQ3: What is the best combination of body-worn camera placement and height to enhance a video viewer’s overall experience?

To help answer these questions, we conducted a 3x3 within-subjects experiment with 26 participants in a VR simulation, varying the factors of Camera Placement and Camera Height. Our main dependent variable was user response regarding satisfaction with the experience. Overall, we found a significant main effect of Camera Placement, such that the video viewers disliked the 360° camera worn on the shoulder, since occlusion due to the surrogate’s head was prevalent. However, we also uncovered an interesting effect where camera placement affected perception of view height. When the camera was placed over the head of the surrogate, even if the height was actually shorter than normal, the participants still felt like the view was tall. Our paper contributes the following:

- A novel user study examining the impact of camera Placement and Height on a user’s viewing experience of body-worn 360° video using an HMD
- An understanding of how vantage points and environmental stimuli affect user perception of camera height
- Design recommendations for content creators, system designers, and camera accessory developers who will prototype future body-worn telepresence experiences

2 RELATED WORK
In this section, we provide a review of relevant literature at the intersection of wearable panoramic video content streaming for...
2.1 Wearable Camera Placements

There has been much work in the HCI community to try out different bodily camera placements, for telepresence or live-streaming. Companies like GoPro sell a variety of mounts and straps to support multiple placements, including backpack-based poles, head and helmet mounts, chest and shoulder straps, hand and wrist straps, and more [GOP 2019; BHA 2019]. Many recent telepresence projects used placements such as on the shoulder [Kashiwabara et al. 2012; Kimber et al. 2014; Saraiji et al. 2018; Tsunaki et al. 2012], the chest [Baishya and Neustaedter 2017; Ishak et al. 2016], the abdomen [Tobita 2017], the crown of the head [Kasahara et al. 2017; Kasahara and Rekimoto 2014, 2015; Lee et al. 2017a,b], over the head (via a pole) [Tang et al. 2017], in a backpack behind the Streamer [Alohali et al. 2016], and even on the Streamer’s face [Misawa and Rekimoto 2015a,b].

Every bodily camera placement has benefits and drawbacks. Cameras worn on the chest, abdomen, or shoulder are not affected by the streamer’s head turns, but the field of view is reduced. Cameras worn on the head allow for a less occluded view, but as the camera and head are coupled, the viewer can be affected by head movements. To combat this, it is possible to implement image stabilization, [Kasahara et al. 2017; Kopf 2016], but this can introduce latency. An overhead vantage point via backpack, mounting pole, and gimbal stabilizer could provide stabilization without latency work, but that means the streamer would need to wear additional gear.

In our study, we analyze perception towards a selection of camera placements that are representative of prior literature - Overhead, Chest, and Shoulder. To our knowledge, we are the first to provide a comparison of user perception on multiple viewpoints. While we don’t expect to find a “one-size-fits-all” solution, we do aim at identifying what types of camera placements are more preferable than others.

2.2 Panoramic Camera Videography

360° cameras are still in their infancy and are on the way to becoming more affordable. As such, there is a lack of definitive guidelines to help panoramic content creators optimize their shots. That said, some general 360° camera tips have emerged, but they do not completely converge. Many seem geared towards taking static, stationary clips. Regarding height, some tips inform creators to place the camera “eye-level with your subject” [Stark 2017], or “at chest-level” [Lavigne 2016], or at “person-height” [Price 2017b]. As far as distance from the subject is concerned, various resources suggest keeping the camera at a balanced distance [Ergürel 2016; Facebook 2016; Price 2017a; Samsung 2016; Sarconi 2017].

Telepresence with body-worn cameras cannot completely subscribe to these tips, because there will often be more than one subject in the scene, the surrogate will often walk around, and the camera viewpoint is coupled with the height of the streamer. For instance, a 6’0” streamer placing the camera over their head would introduce an even taller view. To accommodate a more natural height for a shorter viewer, that surrogate could place the camera on the shoulder; but then that would affect how much of the environment can be seen (see Figure 1). With these issues in mind, our study analyzes the importance of these factors to help determine what viewers prefer.

2.3 View Height in VR and Telepresence

There has been research to help understand how humans estimate heights and distances in VR. Mohler et al. [2006] found that people are more adept at judging distances in the real world, but to help alleviate this, virtual avatars can provide a frame of reference for more accurate distance judgments [Mohler et al. 2010]. Leyrer et al. [2011] found that varying viewpoint heights for a given scene can affect distance judgment as well, as is also implied by Kuhl et al. [2009]. Banakou et al. [2013] studied this effect when participants were given a child avatar body. In our work, we measure how adept our users are at judging heights, but we additionally aim at understanding how various levels of camera height affect the overall user experience. A recent study suggested putting the camera at a constant height of 4’11” (150cm), but was not in the context of body-worn telepresence [Keskinen et al. 2019]. In a robotic telepresence setting, it is possible to move the camera up and down [Matsuda and Rekimoto 2016], but this is not as feasible when the camera is worn on a person’s body, without changing other variables. As such, we expect that camera height and bodily camera placement must be balanced to provide optimal user experience. Our work helps to identify how to achieve this balance.
3 METHODS

3.1 Study Design

We designed a 3x3 within-subjects study varying Camera Placement and Camera Height. We used a simulated virtual environment (VE) for our study. VEs have been used in research for a variety of reasons, including to provide a controlled study space, reduce external and potentially confounding variables, and maintain study feasibility, with little sacrifice to realism [Beidel et al. 2017; Duncan and Murphy 2017; Ragan et al. 2009]. The use of a VE for our study allowed us to maintain variable constancy, such as the surrogate walking gait, level of action, camera steadiness, latency, networking hiccups, etc.

Camera Placement had 3 levels representative of prior literature and common action camera placements - over the head (“Overhead”) [Tang et al. 2017], on the chest near the shirt pocket (“Pocket”) [Baishya and Neustaedter 2017; Ishak et al. 2016], and on the shoulder (“Shoulder”) [Kashiwabara et al. 2012; Kimber et al. 2014; Saraiji et al. 2018; Tsumaki et al. 2012]. For each of our conditions, there are different ways of achieving similar levels of occlusion while being able to manipulate height. For instance, in the Overhead conditions, the surrogate’s body occludes the bottom part of the view; similar views are commonly achieved using a hand-held selfie stick or telescopic pole. In the Pocket conditions, the surrogate’s body blocks the back portion of the view; similar views can be achieved by placing the camera near the abdomen [Tobita 2017] or by using a neck-worn camera (e.g. those used by law enforcement officers). In the Shoulder conditions, the surrogate’s body blocks a piece of the bottom part of the view, and the head blocks the view opposite of the mounted shoulder. The view can be manipulated by including actuators to adjust the camera with 6 degrees of freedom [Kimber et al. 2014; Matsuda and Rekimoto 2016].

Camera Height had 3 levels we felt would give us a good range of exploration - the participants’ natural eye height (“Normal”), their eye height plus 12 inches (“Taller”), and their eye height minus 12 inches (“Shorter”). We chose 12 inches because the difference between an average male and average female in the United States is 6 inches, and two standard deviations of height is 6 inches [Fryar et al. 2016]. Thus, a 12 inch step covers likely ground. We also acknowledge that live streaming is becoming increasingly popular, even at the extremities of average adult human height. For instance, consider the National Basketball Association’s VR app [NBA 2019]. Currently, the app enables viewers to watch live games with a VR HMD, from the view of a static, court-side camera. It seems plausible that the NBA would, in the future, live-stream feeds worn by the players, whose average height in the 2018-2019 season was 6’7” with a standard deviation of 3.3 inches [RealGM 2019]. Additionally, while we do not advocate for minors to wear cameras, we acknowledge that this is a plausible scenario [Everson et al. 2019; Kelly et al. 2012], especially as life-logging tools are becoming more readily available. As such, our height conditions represent a wide range of plausible use cases.

These levels totaled 9 conditions, which were randomized and counter-balanced in a Latin Square design. Our study received IRB approval, and we obtained informed consent from participants before they participated in our study.

Figure 2: Sample snapshot of the environment. Stimuli were distributed and organized in various placements.

3.2 Research Hypotheses

We expect viewers to desire a camera height that matches their own height, but in practice this is not always feasible. Average adult height varies significantly between men and women, and between people from different countries [Cavelaars et al. 2000; Frankenberg and Jones 2004; Fryar et al. 2016; Li et al. 2009; Venkaiah et al. 2002]. To compensate for a difference in height, the streamer could move the camera to a different part of the body; but, this may result in an unnatural or occluded view. We would expect an optimal experience to have a natural viewing height while also affording the clearest viewpoint possible. We thus conducted our study considering these factors, hypothesizing the following:

• H1: We expect a main effect of camera placement such that users will prefer an unoccluded view.
• H2: We expect a main effect of camera height such that users will prefer a viewing angle similar to their natural height.

3.3 Subjects

A priori power analysis using G*Power indicated that we needed a minimum of 22 users to detect a medium effect size [Faul et al. 2007]. We recruited 26 participants for our study from the student body of the University of Central Florida. 21 were male and 5 were female. Their age ranged from 18 to 29 (M = 20.8; SD = 2.74). We measured participant height; the range was 4’9” / 1.44m to 6’3” / 1.91m (M = 5’9” / 1.75m; SD = 4.21in / 10.7cm). All participants had normal vision, or they wore corrective lenses during the study. We asked participants how often they watch 360° videos, and the Median response was “Rarely.” Similarly, the Median response for how often they use VR was “Rarely.”

3.4 Apparatus

We created our virtual environment using the 2017.3.0f3 version of Unity3D. We ran the study on a laptop with Windows 10, Intel core i7-7700HQ at 2.8GHz, with 12GB of RAM, with an Nvidia GeForce GTX 1060. The HTC Vive HMD was used to run the scenario.
3.5 Virtual Environment

We retooled a VE [Shade 2018] to create a virtual art museum (see Figure 2). A museum tour is a plausible example use case, and we were influenced by Tang et al. [2017], who performed a real-life task in which participants needed to search for art during a virtual tour. Our room was rectangular with a dividing wall through the middle, lengthwise. On the walls were famous paintings. The very center of each painting was approximately 5’10” / 1.8m off the floor. Between paintings, there were pedestals that held sculptures. The sculptures rested approximately 2’3” / .7m off the floor. On the floor, ornate rugs were laid out. Scattered in the room were digital human museum-goers who stayed in-place.

For each stimulus type, we proposed a “real” and a “counterfeit” (see Procedure). We tried to balance subtlety with objectivity for the counterfeits, such that the participants would need to inspect the objects yet be able to recognize that something has changed. For the paintings, we performed a web search with the phrase “famous paintings” and selected a subset. To find the counterfeits, we performed a web search for each one, including the word “parody.” We were able to pair every real painting with a parody. For instance, we used Starry Night, and the counterfeit featured Darth Vader [Gilbert 2018]. For the sculptures, we first defined a list of objects typically have one size in the real world, in hopes of giving visual cues for perceiving camera height. For the counterfeits, we used properties of association. For instance, we used a soccer ball, and the counterfeit was a basketball. For the carpets, we performed a web search for “ornate rugs,” and found various patterns. For the counterfeits, we inverted the color scheme but maintained the patterns.

The stimuli were mapped symmetrically. There was a box-shaped path that the virtual surrogate walked, and both sides of the path were as equal as possible in terms of stimuli count and placement. One side of the museum had a door, and the opposite side had a large painting. All paintings and sculptures were scaled to life-size. In total, there were 15 famous paintings, 12 sculptures, 6 rugs, and 4 museum-goers. To simulate common visual artifacts of live-stream cameras, we constructed a virtual Camera Rig that consisted of six cameras, allowing as much time as needed for them to become familiar. Following, we ran the user through a practice trial, in which the camera floated in air (there was no surrogate). The length of this practice run - and each trial - was approximately 2.5 minutes.

The user was then run in the first condition with the surrogate visible. We did not tell the user which condition was being run. During the task, we hand-recorded the user’s audible feedback. After the run was complete, the user filled out a questionnaire. We then told the user how tall the camera was, to untangle confusion which could affect our final survey. We then loaded the next condition and repeated these steps. After all conditions were completed, we gave the user one final questionnaire, to rank all conditions. The time to complete the study was approximately 50 minutes, and participants were given 10USD in cash.

3.6 Procedure

We administered a questionnaire after each condition to measure if the user thought the view was Free from Occlusions and if it had a Natural Height. The measures consisted of the following items, on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree):

- Questions regarding Free-from-Occlusions (FbO):
  - The camera placement allowed me to see everything I needed to see
  - Nothing blocked my view to the point where I became disoriented
  - My field of view was clear, so I could perform the task

- Questions regarding Natural Height (NH):
  - The height of the camera felt natural to me
  - The camera height let me view the environment with ease
  - I liked the height of the camera placement
  - All things considered, please give a score to that camera placement (1 = Terrible, 7 = Excellent)

After each condition, we also asked “Was the view shorter, taller, or equal to your natural height?” Users were also asked to explain what they liked or disliked about each condition. After all conditions were complete, the user was asked to rank all of the conditions from best to worst. They were also asked to tell us the rationale for why they selected the best and worst conditions.
An Analysis of User Perception Regarding Body-Worn 360° Camera Placements...  SAP ’19, September 19–20, 2019, Barcelona, Spain

Table 1: Descriptive Statistics of All Conditions by Dependent Variable

<table>
<thead>
<tr>
<th>Placement</th>
<th>Height</th>
<th>Free-from-Occlusions</th>
<th>Natural Height</th>
<th>Overall Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overhead</td>
<td>Taller</td>
<td>M = 5.923, SD = 1.356</td>
<td>M = 5.551, SD = 1.447</td>
<td>M = 5.846, SD = 1.084</td>
</tr>
<tr>
<td></td>
<td>Natural</td>
<td>M = 6.103, SD = 1.014</td>
<td>M = 5.821, SD = 1.246</td>
<td>M = 6.000, SD = 0.849</td>
</tr>
<tr>
<td></td>
<td>Shorter</td>
<td>M = 5.949, SD = 1.183</td>
<td>M = 5.154, SD = 1.620</td>
<td>M = 5.269, SD = 1.430</td>
</tr>
<tr>
<td>Pocket</td>
<td>Taller</td>
<td>M = 6.000, SD = 1.227</td>
<td>M = 5.564, SD = 1.456</td>
<td>M = 5.654, SD = 1.294</td>
</tr>
<tr>
<td></td>
<td>Natural</td>
<td>M = 6.205, SD = 1.333</td>
<td>M = 6.308, SD = 0.958</td>
<td>M = 6.269, SD = 0.919</td>
</tr>
<tr>
<td></td>
<td>Shorter</td>
<td>M = 6.013, SD = 1.222</td>
<td>M = 5.231, SD = 1.494</td>
<td>M = 5.385, SD = 1.235</td>
</tr>
<tr>
<td>Shoulder</td>
<td>Taller</td>
<td>M = 3.397, SD = 1.854</td>
<td>M = 4.551, SD = 1.649</td>
<td>M = 3.731, SD = 1.185</td>
</tr>
<tr>
<td></td>
<td>Natural</td>
<td>M = 3.000, SD = 1.683</td>
<td>M = 4.538, SD = 1.633</td>
<td>M = 3.385, SD = 1.359</td>
</tr>
<tr>
<td></td>
<td>Shorter</td>
<td>M = 3.244, SD = 1.692</td>
<td>M = 4.321, SD = 1.640</td>
<td>M = 3.423, SD = 1.419</td>
</tr>
</tbody>
</table>

Table 2: Repeated Measure ANOVA Results

<table>
<thead>
<tr>
<th>Construct</th>
<th>ANOVA Result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Main Effect of Camera Placement</td>
</tr>
<tr>
<td>FfO</td>
<td>$F(2, 50) = 71.50, p &lt; .001, \eta^2 = .741$</td>
</tr>
<tr>
<td>NH</td>
<td>$F(2, 50) = 16.58, p &lt; .001, \eta^2 = .399$</td>
</tr>
<tr>
<td>Overall</td>
<td>$F(2, 50) = 58.96, p &lt; .001, \eta^2 = .702$</td>
</tr>
<tr>
<td></td>
<td>Main Effect of Camera Height</td>
</tr>
<tr>
<td>FfO</td>
<td>$F(2, 50) = 0.531, p = .591, \eta^2 = .021$</td>
</tr>
<tr>
<td>NH</td>
<td>$F(2, 50) = 5.946, p &lt; .005, \eta^2 = .192$</td>
</tr>
<tr>
<td>Overall</td>
<td>$F(2, 50) = 2.707, p = .077, \eta^2 = .098$</td>
</tr>
<tr>
<td></td>
<td>Interaction Effect of Camera Height * Camera Placement</td>
</tr>
<tr>
<td>FfO</td>
<td>$F(4, 100) = 1.808, p = .133, \eta^2 = .067$</td>
</tr>
<tr>
<td>NH</td>
<td>$F(4, 100) = 1.182, p = .324, \eta^2 = .045$</td>
</tr>
<tr>
<td>Overall</td>
<td>$F(4, 100) = 1.438, p = .227, \eta^2 = .054$</td>
</tr>
</tbody>
</table>

3.8 Data Analysis Approach

In order to assess the construct validity of the dependent variables, we first tested for internal consistency by calculating Cronbach’s alpha [Cronbach 1951]. $FfO$ and NH were both above the 0.7 threshold for reliability ($FfO: \alpha = .889; NH: \alpha = .824$). Thus, we averaged the values together to form an index per construct, per condition. As a single-item measure, we did not assess Overall Rating for construct validity. Since the DVs were not normally distributed, we used non-parametric tests to test our hypotheses. We anticipated a possible interaction effect between camera placement and height, so we chose to use the Aligned Rank Transform (ART) tool [Wobbrock et al. 2011], so that we could run a repeated measures ANOVA. For our post-hoc tests, we controlled Type I errors by performing Holm’s Sequential Bonferroni Adjustment [Holm 1979]. For qualitative feedback, we used an open coding process to identify emerging themes. We present illustrative quotes to help unpack some of the nuance around the experimental results of our study.

4 RESULTS

We first present the descriptive characteristics of our data, followed by the results of our hypothesis testing, and a summary of our findings.

4.1 Descriptive Statistics

The descriptive statistics of our DVs can be found in Table 1. The following sections describe the results of repeated measures ANOVAs, as shown in Table 2.

4.2 ANOVA Results

4.2.1 H1: Main Effect of Camera Placement. An ANOVA revealed a significant effect of Camera Placement on each of our dependent variables; see Table 2. Post-hoc t-tests revealed significant differences (see Table 3); for the $FfO$ construct, there were differences between Overhead and Shoulder, as well as Shoulder and Pocket. Most of our participants indicated that the Shoulder placement was annoying, frustrating, or generally negative because the surrogate’s head blocked the right side. See Figure 4.

We also found a significant main effect of Camera Placement on NH. Post-hoc t-tests again revealed differences between Overhead...
Table 3: Significant Post-hoc T-Test Results by Main Effect

<table>
<thead>
<tr>
<th>Construct</th>
<th>Condition A</th>
<th>M</th>
<th>SD</th>
<th>Condition B</th>
<th>M</th>
<th>SD</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect of Camera Placement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free-from-Occlusions</td>
<td>Overhead</td>
<td>5.991</td>
<td>1.190</td>
<td>Shoulder</td>
<td>3.214</td>
<td>1.745</td>
<td>( t(25) = 11.35, p &lt; .001 )</td>
</tr>
<tr>
<td>Free-from-Occlusions</td>
<td>Pocket</td>
<td>6.073</td>
<td>1.260</td>
<td>Shoulder</td>
<td>3.214</td>
<td>1.745</td>
<td>( t(25) = 10.13, p &lt; .001 )</td>
</tr>
<tr>
<td>Natural Height</td>
<td>Overhead</td>
<td>5.509</td>
<td>0.941</td>
<td>Shoulder</td>
<td>4.808</td>
<td>0.951</td>
<td>( t(25) = 2.995, p &lt; .001 )</td>
</tr>
<tr>
<td>Natural Height</td>
<td>Pocket</td>
<td>5.701</td>
<td>0.761</td>
<td>Shoulder</td>
<td>4.808</td>
<td>0.951</td>
<td>( t(25) = 5.125, p &lt; .001 )</td>
</tr>
<tr>
<td>Overall</td>
<td>Overhead</td>
<td>5.705</td>
<td>0.881</td>
<td>Shoulder</td>
<td>3.513</td>
<td>1.076</td>
<td>( t(25) = 8.833, p &lt; .001 )</td>
</tr>
<tr>
<td>Overall</td>
<td>Pocket</td>
<td>5.769</td>
<td>0.873</td>
<td>Shoulder</td>
<td>3.513</td>
<td>1.076</td>
<td>( t(25) = 9.730, p &lt; .001 )</td>
</tr>
</tbody>
</table>

Effect of Camera Height

| Natural Height     | Natural    | 5.556   | 1.502  | Shorter     | 4.902   | 1.632  | \( t(25) = 3.004, p < .05 \) |
| Natural Height     | Taller     | 5.560   | 1.447  | Shorter     | 4.902   | 1.632  | \( t(25) = 2.756, p < .05 \) |

4.2.2 H2: Main Effect of Camera Height. An ANOVA revealed a significant main effect on the NH construct, but not on FfO or Overall Rating. Post-hoc t-tests reveal significant differences between Natural and Shorter heights, as well as Taller and Shorter heights; see Table 3. Expectedly, participants found their natural height to be favorable, but unexpectedly, they were unfazed by the taller placement. They did, however, find the shorter camera heights to be less natural. See Figure 5 for illustration.

4.3 Perception of Camera Height

Due to the significant main effect of Camera Placement on the NH construct we conducted a post hoc analysis to help understand why this result emerged. Using our qualitative feedback, we found an interesting result regarding how tall each condition made the participants feel; see Figure 6. During the Overhead conditions, participants often believed the viewpoints were taller than what they actually were, simply because the surrogate’s head was seen underneith them. If we treat the results of this question as a pass/fail item, then only 54% Overhead, 71% Pocket, and 51% Shoulder heights were correctly perceived. Considering our VE, there are only a handful of items with which users could really infer height - paintings, sculptures / pedestals, museum-goers, and the surrogate. The Overhead and Shoulder placements naturally forced the users to see the avatar’s head either below or at eye-level, respectively; but...
An Analysis of User Perception Regarding Body-Worn 360° Camera Placements... SAP '19, September 19–20, 2019, Barcelona, Spain

Figure 6: Perceived Camera Height by Placement. Participants were in-tune with height while using the Pocket placement, but for the Overhead conditions, participants often felt taller regardless of height.

Figure 7: Average Rank by condition with 95% confidence. The Overhead and Pocket placements were well regarded compared to the Shoulder placement. In general, users most preferred the camera to be at their natural viewing height.

for the Pocket placement, the users had to go out of their way to see anything other than the avatar’s legs. Supporting evidence can be found in comments. For the Overhead placement, some participants felt taller simply because there was a head beneath them:

- “I feel taller because I see the dude’s head below”
- “The viewing angle was a bit high, looking down on everything.”
- “I feel like the overall view is lower, but because the avatar is under me, I still feel tall.”
- (As part of positive feedback for the Shorter condition) “I’m tall so the farther off the ground I am the more comfortable”

For the Pocket placement, the participants indicated that the camera wearer did not affect them, and they really only needed to look forward and to their sides, rarely commenting on anything past their immediate view:

- “I didn’t do it, but if I looked back, I would’ve seen the avatar.”
- “The model rarely got in the way of my view...”
- “This was a great, unobscured view of almost everything.”
- “Nothing from the avatar impeded my vision...”
- “I barely noticed the avatar.”
- “The front pocket gives a very clear view of everything in the front.”

Based on this user feedback, it is clear that the virtual streamer was a visual cue that participants used to infer height.

4.4 Feedback from Participants

Next, we analyzed the questionnaire item regarding subjective height perception, and we coded the open response questions in order to determine which placements were regarded positively and negatively, and why.

4.4.1 Ranking Data. Looking at the Ranking data (Figure 7), we find that Shoulder was indeed the worst position of the three. Overhead and Pocket were very positive and comparable. For all placements, we find that the Natural height was viewed as best, followed by a Taller height. This feedback compounds with the quantitative results. Out of all twenty-six (26) users, none of them ranked the Shoulder best, and twenty (20) thought it to be the absolute worst. See Figure 7 for illustration.

There were a number of themes that emerged regarding factors that influenced user satisfaction. A majority of the participants (65%) indicated that a field of view which was free from occlusions was the main benefit of their favorite condition. Forty-two percent (42%) indicated that their favorite placement felt natural to them. Only nine (9) participants responded that height was a major detriment in their least preferred condition. Interestingly, six (6) participants responded that visual features of the virtual avatar was a major drawback:

- “I didn’t like his head being so close to me.”
- “…I felt uncomfortable with the head bouncing.”
- “…I felt like the avatar was bouncing too much.”
- “The avatar’s head was extremely distracting.”
- “It feels a little odd with someone’s head right under my chin...”
- “I didn’t do it, but if I looked back, I would’ve seen the avatar.”
- “The model rarely got in the way of my view...”

In the end, however, occlusion (or lack thereof) is the main contributor to the success of the camera placements.

4.4.2 Virtual Presence and Viewpoint Metaphors. While we often found that our participants simply did not want to see the avatar, some of the users did provide us with interesting feedback, revealing comments that indicate they gained some sense of immersion or presence in the virtual environment. Others provided colorful metaphors within their negative comments for the various camera placements:

- **Overhead:**
  - “It felt like I was riding a horse.”
  - “…it was like piggy-backing on someone’s back.”
- **Pocket:**
  - “It let me view my feet and legs easily...”
  - “Finally know what its like to be tall...”
  - “Made me feel like I was actually walking through.”
  - “I felt like I was being held like a baby.”
  - “This is what its like being in my girlfriend’s body.”
We would expect our finding to hold true in cases where important
would expect users to perceive camera placements more positively if
we found users pointing to negative features of the avatar. This
which can be directly manipulated by the viewer via a GUI system,
where stimuli moves or appears from behind, we would anticipate
5.1 The Clearer the View, the Better
5 DISCUSSION

5.1 The Clearer the View, the Better
Our results indicate a strong disdain for camera placements which
block a significant portion of the front hemisphere. The Shoulder
only had approximately 90° of the environment blocked, and the
Pocket had 180° blocked; but the right side was more important to
our users than the back. Our users were able to complete inspection
before the avatar walked past stimuli, plus they were seated for the
entirety of the study (so it was difficult to turn their head around).
We would expect our findings to hold true in cases where important
stimuli enter the view from the front. For complex environments
where stimuli moves or appears from behind, we would anticipate
placement similar to Overhead being most desirable, to provide the
most opportunity for exploration. For instance, the multi-lens
camera setup in Kasahara et al. seems to meet user needs [Kasahara
et al. 2017; Kasahara and Rekimoto 2014, 2015]. While our users
did not like the Shoulder camera placement, interestingly, prior
researchers did find a similar rig to be well-received [Kimber et al.
2014; Kratz et al. 2015, 2014]; they implemented a gimbal device
which can be directly manipulated by the viewer via a GUI system,
so perhaps it was this sense of control that helped users perceive it
positively [Kimber et al. 2014].

5.2 Users Don’t Want to See the (Virtual)
Streamer
Our participants often pointed out that the virtual streamer had
strange, unattractive, or undesirable traits, or simply didn’t want to
see the model bouncing around while walking. We acknowledge the
Ethan model looks somewhat strange, and this could have made the
participants perceive the avatar negatively. But, user comments also
indicated that camera placement affected overall perception. The
Pocket offered the most unoccluded view, and users had to go out of
the way to see the avatar’s features; but even in these conditions
we found users pointing to negative features of the avatar. This
shows that our participants, for our task and in our environment,
did not want the avatar to appear in their view; as such, our study
cannot answer to a setup where the streamer is someone that the
user knows, e.g. a friend, family member, or celebrity streamer. We
would expect users to perceive camera placements more positively if
a loved one was in the view [Kimber et al. 2014], but users could still
communicate with each other regardless [Baishya and Neustaedter
2017; Ishak et al. 2016].

5.3 Placement Matters More than Height
Based on these results, it seems that camera height may not be a
driving factor for success; while the Shorter height was the least
natural to our users, it didn’t seem to detract from the experience
as much as camera placement. This is contradictory to previous
research, which suggests that lower camera heights are more ac-
ceptable to higher heights [Rothe et al. 2018]. We suspect that there
are external variables not identified here which cause a difference in
height preference; for instance, culture may be such a variable. We
acknowledge that our study only analyzed 3 relative heights, but +/-
12in covers a large range. If we inspected an even wider range, we
would begin including positions alongside the extremities of adult
human height. Our results corroborate with previous findings: our
participants had trouble identifying the “correct” camera height,
which is expected - humans do not excel at judging VR heights and
distances [Asjad et al. 2018; Geuss et al. 2010; Leyrer et al. 2011;
Mohler et al. 2006]. Telepresence between taller and shorter individ-
uals may not be harmed by the disparity in user heights. A drastic
difference may result in a drop in user satisfaction, but our results
imply that it would need to be severe.

5.4 Different Placements Give Different Sense
of Story
User feedback indicated varying experiences through the metaphors
they provided in their comments. It is important to convey a sense
of presence to the user as that allows them to feel as if they were
“actually there” [Bowman et al. 2004; Heeter 1992; Schuemie et al.
2001], which is one of the goals for telepresence. Metaphors have
been used to help describe telepresence setups, such as the user
assuming the role of a parrot sitting on the shoulder [Kimber et al.
2014], or a ghost watching the world from another person’s view
[Kasahara and Rekimoto 2014], or even borrowing another person’s
body [Misawa and Rekimoto 2015b]; but our users sometimes felt like
they were taking a ride on an animal or being carried around
like a child. We find that the Pocket placement helped convey a
sense of active exploration because the users felt like they were the
avatar, due to the character being behind the camera, whereas
the Overhead and Shoulder conditions conveyed a sense of passive
exploration because the users felt like they were watching the avatar
from a third-person viewpoint.

Content streamers have opportunity to give two different types of
experiences, simply by wearing the camera in different spots - one
where the viewer is the “star of the show,” and one where the
viewer watches the events transpire. In a case where the streamer
and viewer are strangers, we would recommend a placement similar
to Pocket. When the streamers know each other and want to have a
communicative experience, an Overhead or even a Shoulder place-
ment could suffice, as the viewer can then clearly see non-verbal
social cues such as upper body gestures or facial features.

6 LIMITATIONS AND FUTURE WORK
While we believe our virtual environment consisted of a good num-
ber of stimuli via the paintings and sculptures, the carpets were not
too interesting, and the counterfeits were easy to spot. While car-
pets make sense for a museum environment, more complex stimuli
such as sidewalk chalk art could be better suited for an inspection
task, and therefore may have revealed a difference in preference
between Overhead and Pocket. We are confident, however, with our
results being representative for environments with much stimuli
that is “eye-level.” We also acknowledge that our virtual environ-
ment may be perceived differently than a “real-life” setup. In the
future, we plan on taking learned outcomes from this study and applying them to a variety of real-world scenarios. For instance, 360° videos are commonly employed for action sports such as skiing and skateboarding, but our test case was a milder indoor exploration task. While our results cannot generalize to all types of telepresence interaction, we plan on using the lessons learned here and analyzing the effects of different walking gaits, speeds, and levels of activity, as well as analyzing how these qualities are perceived by participants. Further, while our users found the Shoulder placement negative here, it may be a good placement when the surrogate is a friend or family member, so the viewer could see their face.

Our study did not contain an aspect of verbal communication, but we plan on taking the current findings and applying them to a real-world task, e.g. a dyadic exploration of a remote environment between friends. We also plan on studying different environments and scenarios in which streamers have begun to explore, to analyze how viewers perceive varying levels of action.

7 CONCLUSION

Telepresence is the culmination of recent technological breakthroughs that we expect to become commonplace in the near future. Our novel investigation contributed a study that revealed how users of similar systems perceive various body-worn 360° camera placements, and we found that there are both positive and negative qualities which can be adjusted for optimal usage. We recommend, if possible, that shorter streamers use an Overhead camera placement, and taller streamers use a Pocket-esque placement, to comfortably accommodate their audience. Our results indicate that this kind of interaction can be used between people of varying heights, in a variety of different cultures, communities, and environments. It is our hope that our investigation will lead streamers onto a path which will provide the best possible experience for their viewers.

ACKNOWLEDGMENTS

This work is supported in part by NSF Award IIS-1638060, Lockheed Martin, Office of Naval Research Award ONRBAIA15001, and Army RDECOM Award W911QX13C0052. We also thank the anonymous reviewers for their insightful feedback. Special thanks to Christy LaPerriere for her graphical design of our Figure 1.

REFERENCES


Tim Bajarin. 2015. Are 360-Degree Cameras the Next Big Thing in Video? Retrieved September 19, 2018 from https://www.pcmag.com/article2/0,2817,2492098,00.asp


Sven Krait, Daniel Avrahami, Don Kimber, Jim Vaughan, Patrick Propp, and Don Severns. 2015. Polly: Being there through the parrot and a guide. In Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services. ACM,


SAP ’19, September 19–20, 2019, Barcelona, Spain


Kevin P. Pfeil, Pamela J. Wisniewski, and Joseph J. LaViola Jr.