Investigating a Simple Technique to Alter Viewpoint Height in Immersive 360° Video

Evan G. Center[1] Adhi Widagdo[1] Matti Pouke[1] Markku Suomalainen[1] Katherine J. Mimnaugh[1]
Abdulsatar Aboud[1] Timo Ojala[1] Steven M. LaValle[1]

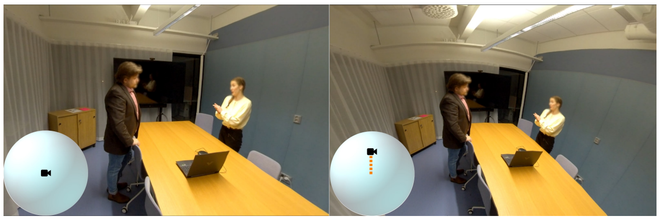

Fig. 1. Two co-workers having a meeting; screenshots of the 360° video content used in the experiment. Left: A view from the normal condition. Right: A view from the shifted-height condition. Insets depicting condition are for illustrative purposes only and were not visible to participants.

***Abstract*— A user's perceived height can have a significant impact on their experience in an immersive telepresence environment. However, virtually manipulating the user's height (if physical adjustment of the camera is not possible) introduces distortions which may counteract positive effects caused by an adjusted height. In a user study of 68 participants, we implemented a simple method for virtually adjusting a user's height in an immersive telepresence meeting which was pre-recorded via a 360° camera to observe the trade-off between the height shift and its ensuing distortions. The shifted-height condition was created via software by changing the position of the virtual camera within the 3-D projection sphere, a simple technique which introduces mild visual distortions. Participants were asked to attend two meetings in immersive telepresence, at normal and increased heights. Our results indicate that while participants were able to detect the visual distortions at an above chance rate, these distortions had little influence on the participants' preferences between conditions, supporting this technique as a viable method of virtually altering a user's height in immersive telepresence environments.**

I. INTRODUCTION

Immersive telepresence technology offers a unique experience for remote collaboration. By leveraging the combination of a 360° camera-equipped mobile robot at the remote location and a virtual reality (VR) head-mounted display (HMD) at the user's location, the technology can simulate the feeling of being in the same physical space as remote participants and grant users the ability to attend a conference or visit a loved one from afar. The ability to manipulate the virtually simulated environment allows developers to then customize the experience to meet the needs of particular users and situations. One critical factor that affects the user experience in telepresence is the perception of height. Height perception is crucial in creating a sense of space [1], [4], [5], [11] and facilitating comfortable interaction with remote participants [10], [19]–[21], which can both contribute to a user's overall quality of experience (QoE) in immersive telepresence.

Many telepresence robots use a "tablet on a stick" design, composed of a set of wheels on the ground to provide mobility and a camera with an accompanying display to provide bidirectional video, which are linked together by a pole. Although this implementation tends to keep the robot lightweight, it can lead to instability, and significant sway and vibration at the tip of the pole where the camera is located.Two popular commercially available telepresence robots come from Double, standing at an adjustable 120 to 150 cm, and Gobe, standing at 161 cm. Note, however, that the global average heights of adult human females and males are 159 and 165 cm, respectively, with significant variability among and within countries [22]. Clearly one size does not fit all, and even adjustable solutions at the hardware level, complicating the design and increasing the cost, may fail to match significant swaths of the global population.

What if, instead, we could virtually alter a user's height at the software level, and perform that transformation regardless of the particular hardware implementation being used? This is the approach taken in this research. By shifting the location of the virtual camera within the 360° video's 3-D projection sphere, we seek to provide the sensation that a user's height has changed without the need to make hardware changes. This technique does come with a potential cost, however. The 3-D projection sphere is constructed with the intention that the virtual camera will be placed in its exact center, and disrupting this placement will then break the intended mapping between the viewpoint and all points on the projection sphere, causing distortions whereby straight lines may appear curved (see Figure 1, right). Although it is admittedly unusual to purposely introduce distortions into a display, the gained flexibility to alter a user's height may entirely outweigh the side effects of whatever subtle visual distortions are produced.

In this study, the main goal is to research the trade-off of allowing the users to observe the world through their preferred height while, unavoidably, introducing certain distortions in the surrounding world. Previous studies have altered the physical height of recording devices, or have changed the virtual height of fully simulated VR experiences, though to the best of our knowledge, no study has virtually manipulated height in 360° video without further scene augmentation to investigate its impact on user experience. Here we artificially raise the camera height in immersive telepresence to fill this gap and test whether participants notice the accompanying visual distortions. Furthermore, we evaluate whether our

This work was supported by BF project HUMOR 3656/31/2019, the RCF projects PIXIE 331822, PERCEPT 322637, and the ERC project ILLUSIVE 101020977.

[1]Center for Ubiquitous Computing, Department of Information Technology and Electrical Engineering, University of Oulu, Finland

manipulation degrades their experience, regardless of whether they notice any distortions, by exploring its impact on VR sickness, the user's sense of presence, and their overall QoE.

II. Related Work

A. QoE Factors in 360° Videos

Obtaining and presenting 360° videos is a non-trivial process with many stages, including stitching, compression/decompression, geometry projection, and rendering, and errors or inadequacies at any stage can lead to various artifacts [2]. These artifacts are usually avoided outright and their potentially detrimental effects on the a user's QoE has not often been specifically examined. Tran et al. [29] manipulated a variety of characteristics of 360° videos that were viewed by participants, and found that video resolutions below 2.5k greatly reduced presence (the sense of "being there" in the virtual environment [27]) and QoE, and particularly so when videos were viewed within VR headsets as opposed to on mobile phones. Later complementary studies then investigated compression factors and the effects of "freezing events" where videos briefly paused, and found harsh drop-offs in QoE with compression past a critical threshold, and at any introduction of freezes [32], [33]. These effects, however, did not seem to be associated with visual attention as indexed by eye-tracking [33]. Doumanoglou et al. [6] as well emphasize the roles of subpar resolution and the presence of lag in reducing QoE.

Some have examined the relationship between motion and QoE in 360° videos. Here, motions may be "camera" motions deriving from movements of the 360° videos during acquisition of the content, and/or users' own head motions as they watch the playback of the content. Tran et al. [29] found that medium camera motion related to better QoE than static or high motion. Regarding users' own head motions, Singla et al. [26] found that the relationship between head motion and QoE was highly dependent upon the type of content present in the 360° video, such that motion was well-tolerated in more static environments, but tended to cause problems in highly dynamic environments. Many prior studies lack large samples and diverse 360° video content. With this limitation in mind, Jun et al. [8] performed a large study of 511 participants who were each exposed to five of 80 unique potential 360° videos. The researchers reported complex interactions between presence, arousal, exploratory behavior, sickness, and content, making a strong case that any effects found for a particular 360° video must be interpreted through the context provided by the scene.

B. Effects of Height in VR, Telepresence, & 360° Videos

A number of studies have explicitly manipulated viewpoint height in either 360° videos or to investigate its effects on some aspect of perception and/or comfort. Leyrer and colleagues have shown that increasing a participant's viewpoint height in VR can cause systematic underestimations of room sizes [13], [14], a process which is also thought to be mostly dependent on postural cues from the participant's own body [15]. Deng & Interrante [4] found that participants were highly sensitive to changes in height in VR environments, particularly when familiar size cues were present. Asjad et al. [1] tasked participants with climbing and descending stairs in virtual environments, and found that participants tended to significantly overestimate their final distance. However, similarly to Deng & Interrante, distances were more accurate when the familiar size cue of a representation of participants' shoes was present in the scene.

Rae et al. [21] examined height manipulations from a social dimension using a setup where participants interacted with a confederate who was live-streamed via a telepresence robot, which could either be configured to stand taller or shorter than the sitting participant. Participants found the confederate to be less persuasive in the shorter configuration. In terms of 360° videos viewed through HMDs, Rothe et al. [23] found that viewpoint heights that were too low were generally better accepted than those that were too high, and that sitting positions were easier to work with than standing positions. Keskinen et al. [10] had participants interact with live actors via 360° video and found that participants' preferred heights bore little relationship to their own true heights, though they also reported negative effects for very low heights and situations where the actors came too close to the camera. Pfeil et al. [19] as well found a large tolerance for different camera viewpoint heights in VR, though they found camera placement relative to potential occluding factors to be critical. In a follow-up study, the researchers also found significant interactions with viewpoint height regarding the presence of other avatars and gender, such that having others in the scene led to taller heights being preferred, and women generally preferred lower heights [20].

C. Arbitrary viewpoint translations in 360 videos

In contemporary VR experiences, users have the affordance to not only look around by rotating their heads but to also translate their viewpoint, resulting in full six-degrees-of-freedom (6DOF) motion. However, 360° videos and photographs are captured from a fixed point of view, so that multiple images are stitched together and 2-D projected either as a sphere map or a cube map. Here, the usual approach to movement in the video is to simulate only rotational head motions even if the VR hardware itself supports 6DOF tracking; the viewpoint cannot be translated, except by moving the physical camera. To overcome this limitation, various computer vision approaches have been developed to simulate 6DOF tracking in inherently 3DOF experiences. For example, by sampling multiple viewpoints around a fixed point of view, one can generate stereoscopic representations of 360 photographs with 6DOF tracking (e.g., Luo et al. [16]). Additionally, structure-from-motion algorithms allow the construction of 3D scenes from video tracks even when captured with a single monoscopic 360 camera; the data can be then used to preprocess and warp the original video track for stereoscopic 6DOF playback (e.g. [7]). Stereoscopic cameras provide depth information, which has also been utilized for augmenting 360 video tracks with 6DOF viewpoint manipulation [12].

To our knowledge, there has been no previous work that investigates the perceptual effects of distortions that occur when the viewpoint is simply translated without warping the 360 projection sphere to account for the translation. Although this approach introduces visible distortions, we argue that this simple solution can be enough to simulate changes in height when more advanced methods are not feasible, such as in real-time telepresence meetings using monoscopic 360 cameras. We pre-registered the following hypotheses (see osf.io/gvwep): H1) Participants *will* notice the shift in height, and H2) Participants *will not* notice the distortions.

III. Methods

A. Study design

We conducted a user study with participants who were asked to attend a telepresence meeting using 360° videos and a VR headset. The scenario was designed to simulate an office meeting, and the participants were asked to attend the meeting as if they were in the same physical space. The 360° videos were recorded in 8k resolution at 60 fps with an Insta 360 Pro 2 camera using live paid actors to reproduce a realistic business meeting environment in a real office setting. The native audio recording by the Insta 360 Pro was found to be too quiet and was replaced by concurrently recorded audio captured by a conference microphone on the center of the table that was attached to a laptop. The videos were filmed in collaboration with Ameliate at the Helsinki XR Center. Videos from the Insta 360 Pro 2 were stitched using their “Insta360Stitcher” software version 3.1.3 and rendered in 4k resolution at 60 fps. A software implementation was developed in Unity for displaying the equirectangular 360° videos with an HMD and to provide a means for our experimental height manipulation. A Varjo XR3 headset was used to display the implementation in the “high” resolution setting at 35 pixels per degree (PPD) per eye, using two Valve Base Stations placed in corners of the room to provide continuous head tracking at 90Hz. The participants were asked to stand in place but could look around the scene by rotating their heads.

The scenes in the 360° videos portrayed two office workers having discussions in a meeting room (see Figure 1). The dialogues included bits of emotional content (e.g., disagreements on whether to hire a new candidate, or celebrating the recent success of coworker) in order to try to keep participants engaged in the meeting. Each video lasted roughly 2 minutes and 20 seconds. The reported heights of the actors were 171 cm and 157 cm for the man and woman, respectively. We captured two unique videos that differed only in terms of the content of the meetings and the placement of the curtains in the meeting room, however the same two actors appeared in both videos, and appeared in the same positions relative to the camera. The camera was placed near a corner of the room, which allowed for a view of both actors from a single viewpoint angle while still occupying a valid standing location for a hypothetical third participant in the room, respecting typical office space proxemics.

The camera was placed at a height of 155 cm. The software implementation allowed us to change the vertical offset of the participant’s viewpoint inside the sphere, effectively changing his/her height compared to the scene in the 360° video. In this way, the perspective in either video could then be virtually raised to approximately 195 cm by translating the location of the virtual camera. This change in height (from the standing height of one of our lab’s most used telepresence robots to a height towards the right tail of the global height distribution) was intentionally drastic and selected such that the height increase should be obvious to participants, and the distortions should become rather noticeable, at least if users knew where to look. A continuous range of heights was initially considered, but was discarded in order to maximize simplicity and statistical power. This “extreme-case” scenario can serve as a strong test of our hypotheses, as it would simulate the experience for taller users who would then also experience the strongest distortions. We may then assume that if the distortions fail to significantly detract from the experience in this case, they should carry even less of an effect for milder height changes. Continuous eye-tracking was obtained throughout viewings of the 360° videos at 200 Hz using the Varjo XR3’s native eye-tracking capabilities, and gaze durations were obtained using custom counters that tracked the overlap between the forward eye vector and researcher-defined colliders in Unity placed over the actors’ faces and bodies, and prominent distortion locations near the floor and ceiling (see Figure 2).

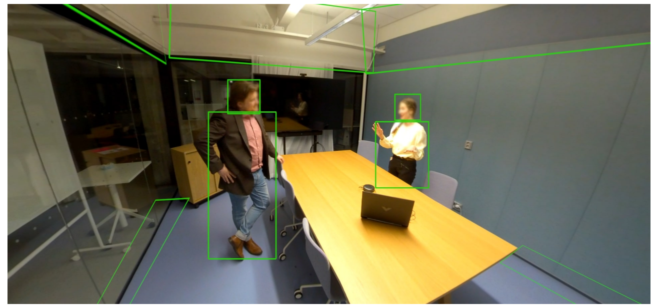

Fig. 2. A screenshot from the normal condition with depictions of colliders overlaid used to collect eye-tracking data, which were invisible to participants.

There were thus two conditions resulting from the height manipulation: normal (“N”; without distortions) and shifted (“S”; with distortions). We used a within-participants design where each participant experienced both conditions at different points in the experiment. Each of two 360° videos (“A” and “B”) could either be displayed as normal or shifted. The order of videos accompanying height perspectives was fully counterbalanced among participants, resulting in four potential video sequences for any individual participant to experience.

B. Participants

Participants from the broader university community were recruited via flyers, university mailing lists, and from a university course. Participants were compensated with university merchandise or partial course credit. We preregistered a sequential sampling approach using Bayesian statistics [25] where we would collect an initial sample of 48 participants and add participants in batches of four until Bayes factors

(BFs) for all four analyses corresponding to our preregistered hypotheses surpassed critical thresholds, whereby the associated BF for each analysis was either less than 1/5 (for evidence in favor of the null hypothesis) or greater than 5 (for evidence in favor of the alternative hypothesis), or, until we had reached 100 participants. Our initial sample size of 48 participants was guided by 1) traditional null hypothesis significance testing (NHST) *a priori* power analysis of the cybersickness effect (based on SSQ total score) from a pilot experiment ($n = 20$), such that we should have an 80% chance of detecting an effect of the estimated magnitude or larger, and 2) the average number of participants needed to find evidence for the null hypothesis given our chosen parameters as indicated by the simulations reported in [25].

C. Procedure

All experiment procedures were approved by the local ethical review board. Participants were welcomed and informed of their rights regarding privacy and data management, and made aware that they could leave the study at any time without penalty. All participants gave written informed consent to take part in the study. The researcher explained that the participant would wear a VR headset to join a currently ongoing meeting in Helsinki, and that they should pay attention to the contents of the meeting. Participants were informed that they would then leave the meeting and take off the headset to complete a questionnaire, then put the headset back on to join one additional meeting, followed by another additional questionnaire outside of the headset. The researcher instructed the participant to stand in a central location in the room marked by tape on the floor, then helped the participant to put on the headset and a pair of over-ear noise-canceling headphones which provided audio for the meeting. Volume was set to 60% of its maximum output. Participants who normally wore glasses were instructed to wear them in the headset. The Varjo XR3 then automatically adjusted for the participant's interpupillary distance (IPD) and performed eye-tracking calibration. Participants experienced both meetings in one of the four possible counterbalancing sequences, answering questionnaires after each.

Each questionnaire began with the simulator sickness questionnaire (SSQ) [9], followed by a 7-point Likert scale item asking how likely they could see themselves participating in an immersive telepresence meeting in the future, a request to summarize the contents of the meeting to the best of their ability, and the extended Slater-Usoh-Steed questionnaire [28], [30] to assess their sense of presence. After participants had experienced both videos, they were then given the following items, in order: a forced-choice yes or no question regarding whether they noticed any differences between the two meetings (beyond the content of the meetings), a four-option multiple choice question asking which language was used in the meetings (as a "catch" question), a forced-choice question asking which of the two meetings was more comfortable, a forced-choice question asking which of the two meetings gave a better sense of being in the remote location, a 7-point Likert scale question regarding the proximity of the other people in the meeting, a 7-point Likert scale question regarding their height relative to others (separately for each meeting), a forced-choice question asking in which meeting they felt taller, a multiple-choice question asking whether they noticed any visual distortions in either of the experiences, a forced-choice question asking which meeting they preferred, and a multiple-choice question asking whether the reason for their preference was primarily related to "height," "distortions," or "other." All relevant questionnaire items were followed by open response boxes that allowed participants to explain the reasoning behind their choices. The last section of the questionnaire concerned demographic information, including self-reported height and previous experience with VR and video games.

D. Analyses

Support for hypotheses was evaluated using two analyses for each hypothesis, respectively. To ascertain whether participants noticed that a height shift occurred, we analyzed the forced choice question, "In which experience did you feel taller?" with choices "First experience" or "Second experience," and computed a BF using the "proportionBF" function from the "BayesFactor" R package [18]. We also analyzed the Likert scale rating for the questions, "Thinking back to the [1st/2nd] experience, how did you feel about your height relative to the other people?" with choices from 1 to 7 with 1 representing "too short" and 7 representing "too tall," for which we computed a BF using the Wilcoxon-signed rank Bayes factor implementation by van Doorn et al. [31].

We tested whether participants noticed the distortions by analyzing the forced-choice question, "Did you notice any visual distortions in either of the experiences?" with options "First experience," "Second experience," "Both experiences," and "Neither," which were re-coded into a two by two table of row dimensions with "normal" and "shifted" height perspectives, and column dimensions with "yes" and "no" to indicate the noticing of distortions. A BF was then computed using the "contingencyTableBF" function from the "BayesFactor" R package [18]. Additionally, the total gaze duration for portions of 360° video defined as containing prominent visual distortions (straight lines near floor and ceiling which become warped by height shift technique) were compared between normal and shifted videos using eye tracking data collected from the Varjo XR3 and Unity. We first tested the differences of the gaze distribution pairs for normality using a Shapiro-Wilk test and applied the appropriate within-groups test ("ttestBF" function from the "BayesFactor" R package [18] if the distribution satisfied conditions of normality, or the Wilcoxon-signed rank Bayes factor implementation by van Doorn et al. [31] otherwise).

All Bayes factors were computed using Cauchy prior distributions with width scale parameter $r = 1$, following recommendations from [24]. Evidence was interpreted as support for alternative hypotheses after surpassing a Bayes factor threshold of 5, and as support for null hypotheses after falling below a Bayes factor threshold of 1/5, representing five to one odds in favor of either hypothesis over the other.

We used a sequential Bayesian hypothesis testing approach [25]. We collected an initial sample of 48 participants and computed Bayes factors, then, if all four hypotheses had not yet surpassed the thresholds, we collected four more participants and recomputed. This process continued until all Bayes factors concerning the four preregistered analyses surpassed evidence thresholds (or would continue until we reached 100 participants). Data from participants who did not complete the experiment due to technology malfunctions or early withdrawal, or who exhibited disruptive behavior (e.g., trying to walk through the virtual environment after being instructed to stand still) or failed to follow instructions provided by the experimenter (e.g., prolonged purposeful disengagement from the tele-meeting content) would be excluded. Data from participants who incorrectly answered the "catch" question of which language was being spoken in the tele-meeting would be excluded as well. Excluded participants would be replaced before conducting analyses.

Finally, we also explored the effects of our height manipulation on cybersickness (as measured by the SSQ total score after each video), presence (as measured by the extended Slater-Usoh-Steed questionnaire after each video, using the original intended calculation procedure [28], [30]), preference (forced choice only after both videos had been viewed), comfort (forced choice only after both videos had been viewed), and relationships between preference and comfort in comparison to participants' true heights. Correlations and their associated Bayes factors were derived using the "correlation" R package [17]. In exploratory analyses, informal Bayes factor thresholds of 3 and 1/3 were taken as moderate support for or against alternative hypotheses, respectively, with values falling in between interpreted as uninformative. Open responses were thematically coded independently by two authors and then cooperatively revised for agreement [3].

IV. Results

A. Sample Characteristics

Three analyses surpassed evidence thresholds within the initial 48 participants, while the analysis probing in which condition participants noticed visual illusions required 68 participants in total (five more batches of four participants each) to pass a threshold. Hence, our final sample contained data from 68 participants. No participants were excluded on the basis of incorrectly answering the multiple choice question which prompted participants to identify the language spoken in the 360° videos. Six participants were excluded in total and were replaced prior to performing analyses. Four of those excluded were due to hardware or software malfunctions, and two were due to loud, unexpected construction noise occurring in a nearby room during 360° video viewings. Participants had a mean age of 28.1 years (range: 20-57) and a mean height of 169 cm (range: 150-193). In terms of gender, 30 participants identified as female and 38 identified as male. In terms of VR experience, 11 reported never having used VR previously, 33 once or twice ever, 8 once or twice per year, 9 once or twice per month, 5 once or twice per week, 2 several times per week, and 0 every day. As for video game experience, 10 reported never having played video games previously, 8 once or twice ever, 8 once or twice per year, 13 once or twice per month, 9 once or twice per week, 7 several times per week, and 13 every day. All participants had normal or corrected-to-normal visual acuity.

B. Confirmatory Results

1) Perception of Height: When asked in which experience participants felt taller, 59 out of 68 chose the shifted height condition, $BF = 8.68 \times 10^7$. Similarly, regarding their height relative to others, participants felt about right in the normal condition ($mdn = 4$) and too tall in the shifted height condition ($mdn = 6$), $Z = 5.60$, $BF = 50,464$. Taken together, we can conclude that the increased height of the shifted height condition was readily noticeable by participants.

TABLE I
NOTICING OF VISUAL DISTORTIONS

Height	Did Not Notice	Noticed
Normal	52	16
Shifted	37	31

2) Perception of Distortions: When participants were asked whether they noticed any visual distortions, 23 reported seeing distortions in either only their first or second experience, 12 reported seeing distortions in both, and 33 reported seeing distortions in neither. In total, only 19 of 68 correctly reported seeing visual distortions in the shifted height condition and not the normal condition. A table recoded by condition is shown in Table I. Although most participants did not claim to see visual distortions in the normal condition, participants were nearly split evenly between noticing and not noticing visual distortions in the shifted height condition, and the imbalance of these proportions did not support our hypothesis that the noticing rate between conditions would be the same, $\chi^2(1, N = 68) = 6.37$, $BF = 7.66$. Thus, participants were more likely to report visual distortions in the shifted height condition when warping was truly present, even if the noticing rate was relatively modest at 45.6%.

Regarding eye tracking data, the difference in the duration spent looking at potentially distorted regions between the two conditions was non-normal, $W = 0.85, p < .001$, and non-parametric tests were used. Participants spent less time looking at potentially distorted regions (summed gaze duration of the ceiling and floor colliders) in the normal condition ($mdn = 3.95s$) than in the shifted height condition ($mdn = 5.03s$), $Z = 2.79$, $BF = 11.26$ (Figure 3, right). Evidence for this difference remained strong even if an outlier who spent 24 more seconds viewing distortions in the shifted height condition than the normal condition was removed, $Z = 2.64$, $BF = 7.67$. Data from both self-report and gaze measures thus support that participants did tend to notice the presence of the visual distortions created by the height shift transformation.

C. Exploratory Results

1) Noticed Differences: The first unique question participants answered after having experienced both conditions was

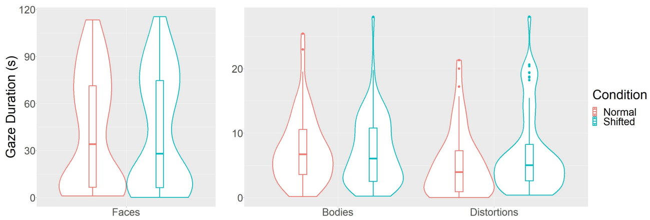

Fig. 3. Violin plots of gaze duration data with box plot insets, for colliders covering actors' faces and bodies, and regions of the scene with long straight lines that become more easily distorted in the shifted condition (ceiling and floor), for both the normal (red) and shifted (blue) conditions. Note the change in the y-axis scale between face colliders and other collider types.

whether they noticed any differences between them at all, beyond the conversation content of the meetings. Fifty-six out of 68 participants responded affirmatively, $BF = 587,406$. When prompted to explain exactly what they noticed to be different, the most common responses included 35 mentions of height (e.g., "The height of the panoramic camera was much higher in the first meeting. I felt like a giant compared to the people in the meeting.") and 17 mentions of the environment (e.g., "There werent curtains, you could see other offices and there was a huge whiteboard on the left."), while there were only 6 mentions of distortions (e.g., "...in the second meeting the environment seemed to be warped a bit.").

2) Gaze Duration: Gaze duration data for all colliders is shown in Figure 3. As can be noted from the graph, participants spent the most time looking at the actors' faces, regardless of condition, to the extent that face gaze duration data needed to be plotted on its own scale to allow the other distributions to be readily discernible. Our evidence was indecisive in terms of the time participants spent looking at the actors' faces between the normal ($mdn = 34.10s$) and shifted height ($mdn = 27.94s$) conditions, $Z = 1.98$, $BF = 0.44$ (Figure 3, left). A couple of participants noted that they thought that body shapes became distorted in one session. We examined the gaze durations for the actors' bodies and here found support for the null hypothesis of no difference in body gaze duration between the normal ($mdn = 6.70s$) and shifted height ($mdn = 6.04s$) conditions, $Z = .66$, $BF = 0.11$ (Figure 3, middle). We suspected that participants who spent more time staring at potentially distorted regions might have been more likely to accurately report the presence or lack of visual distortions in each condition, however it is unclear whether that was in fact the case, $r = .20$, $BF = 0.60$. Furthermore, neither participants' previous experience with VR ($r = -.04$, $BF = 0.16$) nor video games ($r = .11$, $BF = 0.22$) was associated with accurately identifying distortions.

3) VR Sickness: Participants completed the SSQ [9] immediately after each condition. Given that the experiences were stationary meetings, we did not expect to see differences in SSQ total scores. Results from a Bayesian Wilcoxon-signed rank test were inconclusive, as the difference between the normal condition ($mdn = 7.48$) and the shifted height condition ($mdn = 11.22$) lacked strong evidence for or against the null hypothesis, $Z = 1.79$, $BF = 0.74$.

4) Presence: We were curious whether our height shift manipulation would alter a participant's sense of presence in the virtual environment. SUS scores from the normal condition ($mdn = 1.5$) were similar to SUS scores in the shifted height condition ($mdn = 1$), $Z = 1.33$, $BF = 0.24$. A participant's difference in SUS scores between the two conditions was not associated with their true height, $r(66) = -.03$, $BF = .18$. When asked directly which of the two experiences gave a better sense of being in the remote location, 45 out of 68 participants chose the normal condition over the shifted condition, $BF = 5.33$. There lacked strong evidence in either direction of a relationship between participants' choices and their true heights as tested by a point-biserial correlation, $r(66) = -0.15$, $BF = 0.32$. When describing the reasoning behind their choice, the most common responses included 23 mentions of height (e.g., "Being in more of a eye level with the people in the meeting made me feel more like I was actually there."), 8 mentions of realism (e.g., "[The] first meeting was more alike my real-world experience."), and 8 mentions of familiarity (e.g., "Because it was the second experience of the meeting with the HMD on."); yet, there were no mentions of distortions.

5) Comfort: When asked to choose which of the experiences was more comfortable, 48 out of 68 participants selected the normal condition, $BF = 52.09$, however it is unclear whether this choice was associated with participants' true heights, $r(66) = -.23$, $BF = 1.09$. The most common reasons given to distinguish comfort among the two experiences included 23 mentions of presence (e.g., "I felt more like I was in that meeting as a participant and less like an observer."), 21 mentions of height (e.g., "My height was more natural."), and only 4 mentions of distortions (e.g, "In real life rooms do not shrink along orthogonal axes.").

6) Preference: When asked directly to choose which experience they preferred, 51 out of 68 participants selected the normal condition, $BF = 952$, although it was again unclear whether participants' preferences were associated with their true heights, $r(66) = -.16$, $BF = .39$. When asked directly whether their preference was mostly related to height, distortions, or "other," 44 chose height, 8 chose distortions, and 16 chose "other." When asked to describe any distortions if any were noticed, 8 mentioned specific objects or locations without actually describing the nature of the distortion (e.g, "I felt the corner of the room behind me being very distorted."), 5 mentioned line curvature (e.g., "I felt like my field of vision was curved, like watching security camera footage."), and 3 mentioned body proportions (e.g. "In the second experience the people in the room looked slightly compressed vertically as if they shrunk with me. It was not a very strong distortion and was not particularly annoying or disorienting.").

7) Continued Use: After each session, participants were asked to indicate on a 7-point Likert scale whether they could see themselves partaking in an immersive telepresence meeting in the future, with options ranging from "strongly disagree" to "strongly agree." A test of the difference between ratings in the normal condition ($mdn = 6$) and shifted height condition ($mdn = 5$) was inconclusive, $Z = 1.94$, $BF = 1.37$.

V. Discussion

We investigated the effects of an easy method of virtually altering height in immersive telepresence. In an experiment with 68 participants, one condition took the normal perspective of the 360° camera, while in another, we raised the perspective by changing the position of the virtual camera within the projection sphere of the 360° video. We predicted that participants would notice the shift in height but not notice the accompanying visual distortions. We found support for our first hypothesis; however, we could not support our second hypothesis that participants would not notice the distortions, as noticing rates differed between the two conditions and participants spent more time looking at distorted regions in the shifted-height condition than they spent looking at those same regions of the normal condition. The evidence in favor of noticing the distortions (on the order of about 10 to 1) was, however, modest in comparison to the magnitude of the evidence in favor of noticing the height shift (on the order of many thousands to one).

In terms of participants' explicit reports of noticing whether any visual distortions were present, we observed a base rate of roughly 24% of participants claiming to notice distortions in the normal condition, compared to roughly 46% claiming to notice distortions in the shifted-height condition. Although this represents a significant increase, it shows that less than half of the participants noticed any distortions at all in the shifted-height condition. Of those that reported noticing distortions, only eight accounts specifically mentioned visual features that could plausibly relate to the height shift manipulation (descriptions mentioning line curvature and body proportions). When asked directly whether height, distortions, or other factors were most related to their preferred experience, only 12% of participants selected distortions as their primary motivating factor. Both gaze durations on actors in the scene and presence scores were found to be the same between the two conditions, while evidence regarding VR sickness and continued use was ultimately inconclusive. The normal condition was found to be more comfortable and more preferred than the height-shifted condition, although the qualitative data revealed that the reasons for these choices were largely due to the height-shifted condition being too tall for most users' preferences.

Our eye-tracking results show consistency with van Kasteren et al. [33] in terms of a focus on faces in the scene, though they stand in contrast at first glance in terms of their relationship between gaze and visual anomalies. We suspect at least two reasons could have attributed to their null effect and our positive one. The most obvious is that their visual anomalies were different in nature, comprising of visual quality degradations and freezing events which would equally affect the entire scene and would not therefore necessarily attract visual attention towards any particular location. Our distortions, although technically distributed widely throughout the scene, were most salient in particular areas containing long, straight lines. Another interacting factor could be statistical power, as van Kasteren et al. powered their experiment to detect mean opinion score (MOS) differences of a particular magnitude, rather than focusing on eye-tracking outcomes. Our approach instead used Bayesian sequential sampling, a process which only terminates after enough evidence has accumulated, resulting in a final sample size that was over double that of van Kesteren et al. Still, even with more specifically localized distortions in a larger sample, our gaze effects were somewhat modest.

In sum, we may infer that our height shift manipulation is effective at creating noticeable height differences, and though it creates distortions, they do not significantly impact the user experience. As in previous work, participants in our sample preferred a lower height to a taller one [23]. Interestingly, we also find concordance with previous work showing little relationship between a user's true height and our outcome measures [10], [19]. Our selected viewpoint heights (155 and 197 cm) fall near the range extremes of our sample's true heights (150 to 193 cm), yet we note that stronger conclusions regarding the relationships to true height could be drawn from a design where participants were permitted to continuously adjust to their preferred height. Nevertheless, regardless of a user's preferred height, we advocate our technique as a viable method of effortlessly achieving that target height. In addition to achieving a strong sense of the height change, the technique benefits from not requiring a physical change to the remote hardware, which even allows users to fix pre-recorded videos acquired at an undesirable height after the fact. Importantly, the visual distortions that are introduced are minor and do not require complex post-processing techniques to counteract them in order to maintain a user's sense of presence in the scene. Although we used prerecorded 360° videos in our study to guarantee that all participants received the same experiences, we have no reason to believe that the technique would introduce lag or decrements in resolution in live settings; two of the most disruptive factors in maintaining high QoE [6], [32], [33].

Finally, although results regarding VR sickness, presence, and likelihood of future use did not yield conclusive results, note that it is only through the use of Bayesian statistics that we may quantitatively conclude that we do not have enough information to draw strong conclusions. Using traditional NHST, we would have failed to reject the null hypothesis in each of these three occasions ($p = .074$, $p = .18$, $p = .052$, respectively), which might indicate to some researchers that our manipulation does not impact these factors. What we may infer in these cases, however, is that if these factors are impacted, the impact is likely quite small; a conclusion supported by not only the Bayes factors previously reported but also effect size estimates and 95% bootstrapped confidence intervals associated with VR sickness ($r = 0.22$, $CI = [0.01, 0.44]$), presence ($r = 0.13$, $CI = [0.01, 0.33]$), and likelihood of future use ($r = 0.25$, $CI = [0.03, 0.47]$).

A. Limitations and Future Work

A limitation of the study is only comparing two heights. The motivation was to choose a manipulated height that clearly created distortions, but was still sensible for a human user to request, while permitting a design with strong statistical power. This way, we managed to answer the most important research question; even if users noticed the distortions, they did not bother them enough to overcome the positive effects of the height shift, showing that a height shift can be enacted even without cutting-edge virtual-parallax shifting methods. Still, future work is needed to determine how shorter heights and height changes of smaller magnitude might affect the noticing of distortions and user preferences.

It is important to note that, given the dependencies of 360° video manipulations on content type [8], we have thus far only used one set of videos as a first pass that contain actors of particular heights, performing one particular type of activity, and in one particular setting. In this case, our results show that the distortions do not greatly deteriorate the subjective experience of the users in telepresence, and that users may be willing to tolerate distortions to have the robot at a preferred height. Future work could explore further height-distortion trade-offs in different settings.

References

[1] N. S. Asjad, H. Adams, R. Paris, and B. Bodenheimer. Perception of height in virtual reality: a study of climbing stairs. In *Proceedings of the 15th acm symposium on applied perception*, pp. 1–8, 2018.

[2] R. G. d. A. Azevedo, N. Birkbeck, F. De Simone, I. Janatra, B. Adsumilli, and P. Frossard. Visual distortions in 360° videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(8):2524–2537, 2020. doi: 10.1109/TCSVT.2019.2927344

[3] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.

[4] Z. Deng and V. Interrante. Am i floating or not?: Sensitivity to eye height manipulations in hmd-based immersive virtual environments. In *Acm symposium on applied perception 2019*, pp. 1–6, 2019.

[5] E. Dima, K. Brunnström, M. Sjöström, M. Andersson, J. Edlund, M. Johanson, and T. Qureshi. View position impact on qoe in an immersive telepresence system for remote operation. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–3. IEEE, 2019.

[6] A. Doumanoglou, D. Griffin, J. Serrano, N. Zioulis, T. K. Phan, D. Jiménez, D. Zarpalas, F. Alvarez, M. Rio, and P. Daras. Quality of experience for 3-d immersive media streaming. *IEEE Transactions on Broadcasting*, 64(2):379–391, 2018.

[7] J. Huang, Z. Chen, D. Ceylan, and H. Jin. 6-dof vr videos with a single 360-camera. In *2017 IEEE Virtual Reality (VR)*, pp. 37–44. IEEE, 2017.

[8] H. Jun, M. R. Miller, F. Herrera, B. Reeves, and J. N. Bailenson. Stimulus sampling with 360-videos: Examining head movements, arousal, presence, simulator sickness, and preference on a large sample of participants and videos. *IEEE Transactions on Affective Computing*, 2020.

[9] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, Jul 1993. doi: 10.1207/s15327108ijap0303_3

[10] T. Keskinen, V. Mäkelä, P. Kallioniemi, J. Hakulinen, J. Karhu, K. Ronkainen, J. Mäkelä, and M. Turunen. The effect of camera height, actor behavior, and viewer position on the user experience of 360 videos. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*, pp. 423–430. IEEE, 2019.

[11] J. Kim and V. Interrante. Dwarf or giant: The influence of interpupillary distance and eye height on size perception in virtual environments. In *Proceedings of the 27th International Conference on Artificial Reality and Telexistence and 22nd Eurographics Symposium on Virtual Environments*, pp. 153–160, 2017.

[12] P. K. Lai, S. Xie, J. Lang, and R. Laganière. Real-time panoramic depth maps from omni-directional stereo images for 6 dof videos in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 405–412. IEEE, 2019.

[13] M. Leyrer, S. A. Linkenauger, H. H. Bülthoff, U. Kloos, and B. Mohler. The influence of eye height and avatars on egocentric distance estimates in immersive virtual environments. In *Proceedings of the ACM SIGGRAPH symposium on applied perception in graphics and visualization*, pp. 67–74, 2011.

[14] M. Leyrer, S. A. Linkenauger, H. H. Bülthoff, and B. J. Mohler. Eye height manipulations: A possible solution to reduce underestimation of egocentric distances in head-mounted displays. *ACM Transactions on Applied Perception (TAP)*, 12(1):1–23, 2015.

[15] M. Leyrer, S. A. Linkenauger, H. H. Bülthoff, and B. J. Mohler. The importance of postural cues for determining eye height in immersive virtual reality. *PloS one*, 10(5):e0127000, 2015.

[16] B. Luo, F. Xu, C. Richardt, and J.-H. Yong. Parallax360: Stereoscopic 360 scene representation for head-motion parallax. *IEEE transactions on Visualization and Computer Graphics*, 24(4):1545–1553, 2018.

[17] D. Makowski, M. S. Ben-Shachar, I. Patil, and D. Lüdecke. Methods and algorithms for correlation analysis in R. *Journal of Open Source Software*, 5(51):2306, 2020. doi: 10.21105/joss.02306

[18] R. D. Morey and J. N. Rouder. *BayesFactor: Computation of Bayes Factors for Common Designs*, 2022. R package version 0.9.12-4.4.

[19] K. Pfeil, P. Wisniewski, and J. J. LaViola Jr. An analysis of user perception regarding body-worn 360 camera placements and heights for telepresence. In *ACM Symposium on Applied Perception 2019*, pp. 1–10, 2019.

[20] K. Pfeil, P. J. Wisniewski, and J. J. Laviola Jr. The effects of gender and the presence of third-party humans on telepresence camera height preferences. In *ACM Symposium on Applied Perception 2020*, pp. 1–10, 2020.

[21] I. Rae, L. Takayama, and B. Mutlu. The influence of height in robot-mediated communication. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 1–8, 2013. doi: 10.1109/HRI.2013.6483495

[22] M. Roser, C. Appel, and H. Ritchie. Human height. *Our World in Data*, 2013. https://ourworldindata.org/human-height.

[23] S. Rothe, B. Kegeles, and H. Hussmann. Camera heights in cinematic virtual reality: How viewers perceive mismatches between camera and eye height. In *Proceedings of the 2019 acm international conference on interactive experiences for tv and online video*, pp. 25–34, 2019.

[24] J. N. Rouder, P. L. Speckman, D. Sun, R. D. Morey, and G. Iverson. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*, 16:225–237, 2009.

[25] F. D. Schönbrodt, E.-J. Wagenmakers, M. Zehetleitner, and M. Perugini. Sequential hypothesis testing with bayes factors: Efficiently testing mean differences. *Psychological methods*, 22(2):322, 2017.

[26] A. Singla, S. Fremerey, W. Robitza, and A. Raake. Measuring and comparing qoe and simulator sickness of omnidirectional videos in different head mounted displays. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6. IEEE, 2017.

[27] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.

[28] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3(2):130–144, 1994.

[29] H. T. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang. A subjective study on qoe of 360 video for vr communication. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 1–6, 2017.

[30] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence: Teleoperators & Virtual Environments*, 9(5):497–503, 2000.

[31] J. van Doorn, A. Ly, M. Marsman, and E.-J. Wagenmakers. Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and spearman's ρ. *Journal of Applied Statistics*, 47(16):2984–3006, 2020.

[32] A. van Kasteren, K. Brunnström, J. Hedlund, and C. Snijders. Quality of experience assessment of 360-degree video. In *Human Vision and Electronic Imaging 2020*, vol. 11, 2020.

[33] A. van Kasteren, K. Brunnström, J. Hedlund, and C. Snijders. Quality of experience of 360 video–subjective and eye-tracking assessment of encoding and freezing distortions. *Multimedia tools and applications*, 81(7):9771–9802, 2022.