Closer Object Looks Smaller: Investigating the Duality of Size Perception in a Spherical Fish Tank VR Display

Qian Zhou, Fan Wu, Sidney Fels
University of British Columbia
{qzhou, wufan95, ssfels}@ece.ubc.ca

Ian Stavness
University of Saskatchewan
ian.stavness@usask.ca

ABSTRACT

Fish Tank Virtual Reality (FTVR) displays provide compelling 3D experiences by rendering view-dependent imagery on a 2D screen. While users perceive a 3D object in space, they are actually looking at pixels on a 2D screen, thus, a perceptual duality exists between the object's pixels and the 3D percept potentially interfering with the experience. To investigate, we conducted an experiment to see whether the on-screen size of the 2D imagery affects the perceived object size in 3D space with different viewing conditions, including stereopsis. We found that the size of on-screen imagery significantly influenced object size perception, causing 83.3% under/overestimation of perceived size when viewing without stereopsis and reducing to 64.7% with stereopsis. Contrary to reality, objects look smaller when the viewer gets closer. Understanding the perceptual duality helps us to provide accurate perception of real-world objects depicted in the virtual environment and pave the way for 3D applications.

Author Keywords

Fish tank virtual reality, spherical display, 3D perception

CCS Concepts

•**Human-centered computing → Virtual reality;**
•**Computing methodologies → Perception;**

INTRODUCTION

Fish Tank Virtual Reality (FTVR) displays create 3D illusions by rendering to the viewer's perspective. The view-dependent imagery provides multiple visual cues about the depicted virtual objects. This includes 3D cues (depth cues), such as motion parallax and binocular stereo, as well as 2D cues (on-screen cues), such as the position and size of the 2D projection on the screen. While the importance of these depth cues has been long appreciated in the practice of FTVR [32, 18, 8], few works have investigated the influence of on-screen visual cues. As the virtual objects are depicted via the 2D projection on the screen, these on-screen cues might influence the perception of virtual objects. For example, when a 3D ball is rendered on the screen based on the viewer's position, the 2D size of its projection is an on-screen cue to the viewer as shown in Figure 1. When judging the ball's size, they may judge it based on the actual size of the ball, or, the 2D size of its projection on the screen. In this case, the on-screen size cue might interfere with the visual interpretation of the ball.

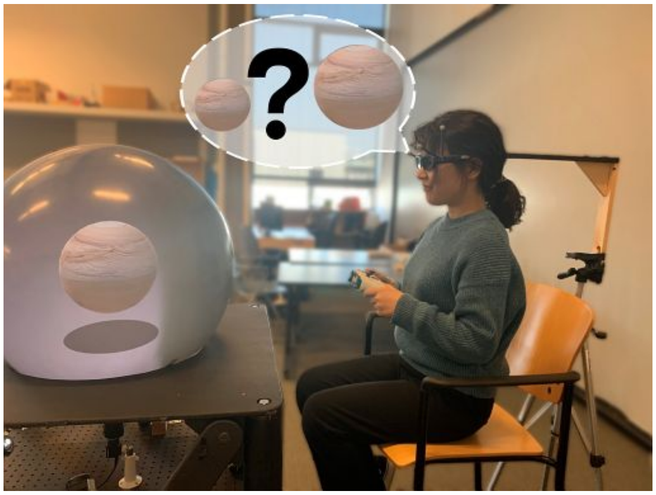

Figure 1. A user interacting with the Spherical Fish Tank VR display while performing the 3D perception task by judging the size of a virtual ball. The virtual ball looks smaller as user gets closer to the screen.

For FTVR displays, the depth cues and on-screen cues are inherently coupled due to the nature of rendering: all cues are presented via the screen. Hence a perceptual duality exists such that users can perceive the virtual object either based on depth cues or on-screen cues. In vision science, similar ambiguity has been previously referred as "duality of the depth perception in picture" [12, 13, 32]. They found that adding depth cues in pictures can make one see in 3D rather than in 2D [32]; though their work focuses on static pictures. It is an open question whether these findings could be applied to screen-based 3D displays such as FTVR displays.

To investigate whether the on-screen visual cues can influence users' perception, we conducted an experiment that measures users' size perception on a spherical FTVR display. We evaluated the perceptual duality by measuring the perceived size with different on-screen imagery on a spherical FTVR display. We focus on the spherical form factor in this study because it has been most widely adopted for FTVR displays [8]. We found the on-screen cues significantly affected perceived size of users. With equivalent retinal images, smaller/larger on-screen imagery caused 83.3% under/over-estimation of the

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376601

perceived size respectively when viewing without the stereo cue. Adding stereo cue reduced the under/over-estimation rate to 64.7%. Contrary to reality, participants had a tendency to report objects as smaller when they moved closer to the object.

This study is the first to evaluate and provide insights on the perceptual duality with 3D displays. While it is conducted with a spherical FTVR display, the result applies to most screen-based 3D displays such as a CAVE [5]. Our study establishes a fundamental limitation for a broad range of "screen-based" 3D displays. All screen-based 3D displays, like FTVR and CAVE, approximate holograms by rendering perspective on the surface with the assumption that if the perspective is geometrically correct, the perception will be correct. But our study shows that under some circumstances, this assumption may not hold and the approximated "hologram" causes perceptual bias with visual artefacts. Understanding the perceptual bias helps us to provide accurate perception and pave the way for 3D applications.

RELATED WORK

This section provides an overview of FTVR displays and perception studies in the virtual environment. Our work investigated the perceptual duality with FTVR displays.

Fish Tank Virtual Reality

Early FTVR display provides multiple depth cues with a single planar desktop display to improve user's spatial perception of various tasks [2]. Using this technique, a number of works [24, 19, 17] have been introduced. An important extension of the FTVR concept utilized multiple 2D displays or projectors to make it scalable with high resolution. Enclosed as a geometric shape, such as a box [24] or a sphere [9], FTVR can render 3D objects on geometric screens that allows viewers seeing the virtual world situated in the real environment. Among different shapes of displays, the spherical FTVR display has a promising shape as it is seamless between screens to provide an unobstructed view from all angles [36]. Recent research progress with spherical FTVR displays have introduced improved calibration and rendering techniques [29, 36]. These advances have increased the fidelity of 3D FTVR experience.

The extra depth information provided by FTVR displays can potentially facilitate 3D understanding. A number of works have been conducted to investigate the relative importance of different depth cues with FTVR displays in a variety of tasks. Stereopsis and motion parallax have been found to improve 3D performance to understand the 3D graph structure [2, 26], perceive exocentric depth [33, 8], perceive 3D contour [31], select 3D targets [1, 8] and improve presence [18]. A comprehensive review of the task-dependent depth cue theory can be found in [32].

Perceptual Duality in FTVR displays

When users perceive 3D objects in FTVR displays with various depth cues, they are actually looking at pixels on a 2D screen. A perceptual duality exists between the object's pixels and 3D percept such that users can either perceive the object in a 3D space, or as a 2D representation on the screen. Similar perceptual duality has been well-studied by vision science researchers. They found that the inherent dual reality in paintings or photographs enables viewers to perceive a scene as 3D at the same time see the flat surface of the picture [12, 13, 27]. In the virtual environment, few work has investigated this duality. Ware [32] discussed the duality of size perception in the virtual environment as "a choice between accurately judging the size of a depicted object as though it exists in a 3D space and accurately judging its size on the picture plane". Benko et al. [3] mentioned this ambiguity as object presence. Using projector-based 3D displays, they investigated when visualizing without the stereo cue, whether uses could perceive the presence of virtual objects as spatial rather than as 2D projections on the screen surface. Recently, Zhou et al. [34] investigated the screen shape factor on the size constancy in FTVR displays. They discussed different visual cues in FTVR displays, including 2D cues, such as the on-screen size of the 2D imagery, as well as 3D cues, such as stereopsis and motion parallax. They believed the duality of size perception is caused by the coexistence of 2D cues and 3D cues. Following [34], our work seeks to examine the role of the 2D on-screen imagery on the perceived size of virtual objects. To the best of our knowledge, it is the first study that investigates the duality of size perception in FTVR displays.

Size Perception in the Virtual Environment

Most of the research investigated size perception with a Head Mounted Display (HMD) using a size-matching [20] and size-judgment task [25]. Eggleston et al. [6] found size-constancy is weak in a VE compared to the real world using an HMD. Kenyon et al. [20] further found depth cues like stereopsis can help provide more accurate size perception in the VE. They used a size-matching task to measure size perception in different viewing conditions. Ponto et al. [21] investigated the size perception using a shape-matching task and found that accurate perceptual calibration will significantly improve the size perception. Kelly [15] investigated the re-scaling effect caused by walking through the VE also using a size-matching task. They found walking through a VE causes rescaling of perceived space with an HMD. Benko et al. [3] evaluated size perception with a projector-based 3D display and found participants are able to deduce the size of a virtual object using a size-judgment task. Elner and Wright [7] reported a direct measure of VE visual quality in a distance and size estimation task with an HMD. Stefanucci et al. [25] used size-judgment tasks to assess the perceived size of virtual objects. They found the size in the virtual environment is underestimated compared to the real world; the addition of stereopsis alleviated the underperception. Zhou et al. [34] evaluated size perception with FTVR displays and found that the screen shape factor influenced size perception using a size-matching task.

As our goal is to investigate whether the on-screen size cue can affect the perceived object size with the FTVR display, we measured the size perception using a size-judgment task due to the effectiveness demonstrated from previous research as mentioned. Our present study seeks to provide evidence on spatial perception with a better understanding of the role of the on-screen 2D imagery. It also provides some insight on

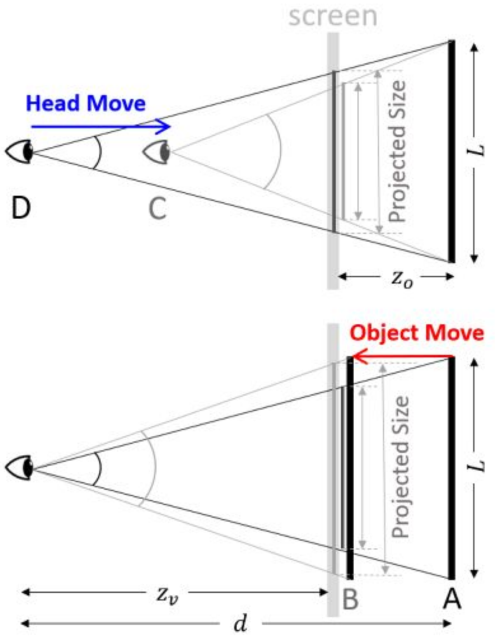

Figure 2. Perspective projection of a traditional planar FTVR display. (Top) The projected size on the screen decreases as the viewpoint moves towards the screen from D to C in *HeadMove*. (Bottom) The projected size increases as the object moves towards the screen from A to B in *ObjectMove*. The visual angle on the retinal image increases in the same way in both cases, thus a viewer moving toward an object versus object moving towards the viewer has the same impact on the retinal image.

size perception with FTVR displays as there has been few size perception studies with FTVR displays.

USER EXPERIMENT

We conducted a user study to examine the duality of size perception on a spherical FTVR display. The purpose of this study is to investigate whether different sizes of the on-screen imagery can affect perceived size of virtual objects. As the perceived object size can be greatly affected by the visual angle on the retinal image [16], we defined two types of movements (*HeadMove* and *ObjectMove*) to provide consistent retinal images across conditions with different on-screen size as described below.

Projection Models

As shown in Figure 2, the on-screen imagery is computed based on the perspective projection between the virtual object and viewpoint. To keep consistent retinal images, we define two types of movements: *HeadMove* and *ObjectMove*. We show that *HeadMove* and *ObjectMove* provides distinct on-screen imagery while maintaining the same visual angle.

In *HeadMove* (Figure 2 (Top)), the viewpoint moves closer to the virtual object via forward head movements toward the screen. The projected size on the screen can be computed as:

$$ProjectedSize = L - \frac{Lz_o}{d} \tag{1}$$

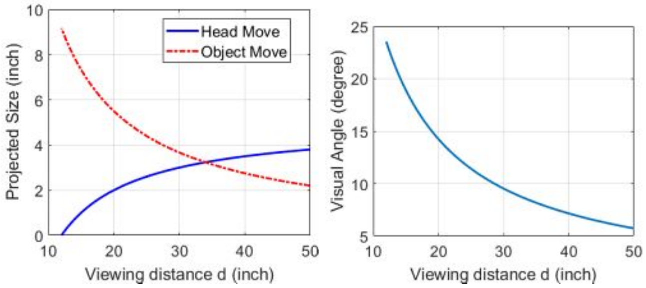

Figure 3. (Left) Projected size on the screen and (Right) visual angle on the retinal image as functions of the viewing distance d. In *HeadMove*, the projected size decreases as d decreases, while in *ObjectMove*, the projected size increases as d decreases. In both *HeadMove* and *ObjectMove*, the visual angle increases independent of movement types as d decreases.

where d is the viewing distance between object and viewpoint, z_o is the distance between object and screen, and L is the virtual object's size. The on-screen projected size gets smaller as d decreases as shown in Figure 3 (Left). In *ObjectMove* (Figure 2 (Bottom)), the virtual object moves closer to the viewpoint. Similarly, it can be computed as:

$$ProjectedSize = \frac{Lz_v}{d} \tag{2}$$

where L is the virtual object's size, z_v is the distance between the virtual object and screen. Contrary to *HeadMove*, the on-screen projected size increases as d decreases, shown in Figure 3 (Left). In both *HeadMove* and *ObjectMove*, the visual angle α can be computed as:

$$\alpha = 2arctan\frac{L}{2d} \tag{3}$$

As the visual angle α only depends on d and L, it is independent of movement types; thus, the retinal images are the same across *HeadMove* and *ObjectMove* as shown in Figure 3 (Right). The primary difference between two movements is the on-screen size of the 2D imagery. While the example shown in Figure 2 uses a planar screen, the same projected phenomena occurs for arbitrary screen shapes, as each pixel on the screen follows the same rule of the perspective projection in the view frustum. In this study, we tested whether the oppositely varied size of the on-screen imagery results in different size interpretations with equivalent retinal images.

Task

Participants visualize a virtual ball while getting closer to the ball, via either *HeadMove* or *ObjectMove*. The task for participants is to judge whether the size of the ball has changed or not by making a three-alternative forced choice via answering the question: "Is the size of the ball getting smaller, larger or unchanged?". Early pilots of this experiment showed that presenting the virtual ball without modifying its size would trivialize the task; to make the task nontrivial, the size of the virtual ball is adjusted by making it smaller, larger or unchanged so that ball's size was varied at the same time as the

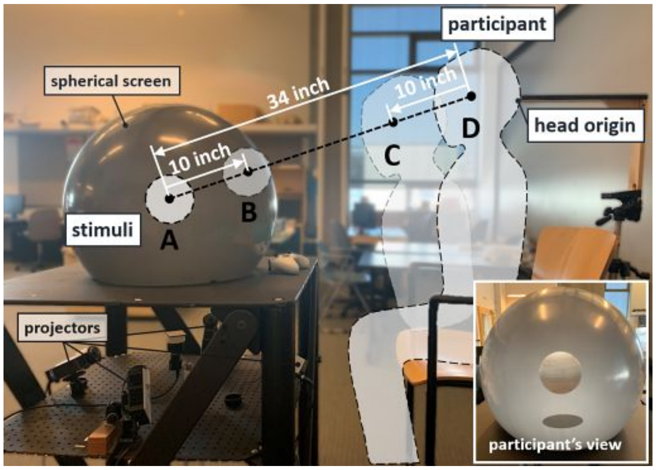

Figure 4. Experimental setup of the study. In *HeadMove*, participants move their head from D to C when the stimuli stays at A. In *ObjectMove*, participants move the stimuli from A to B when the head stays at the origin D. We use a spherical display (24" diameter) with projectors rear-projecting through a projection hole at the bottom of the screen. We track the head position to render view-dependent imagery shown in the right corner, and ensure the movement magnitude is 10 inches in both *HeadMove* and *ObjectMove*.

head/object was moving (Figure 5). If participants' perception is influenced by the on-screen imagery, their answers will be biased towards one side.

Experimental Design

We followed a 2x2 within-subjects design with two independent variables as:

- **C1** The movement condition, which could be *HeadMove* or *ObjectMove*. In *HeadMove*, participants move the head towards the object while in *ObjectMove* they move the object towards them.

- **C2** The viewing condition, which could be *Stereo* or *NonStereo*. In *Stereo*, participants visualize stereoscopic imagery while in *NonStereo* they visualize monocular imagery set to the mid-point of two eyes as suggested by [8].

The four conditions are shown in Figure 6. Subject performance was evaluated based on the measure *BiasError*, defined as the difference of scores between the reported and expected answer, with the score of Small, Same and Large equal to -1, 0 and 1 respectively. Hence, a positive value of *BiasError* indicates overestimation of size while a negative value means underestimation.

We hypothesized that:

- **H1** There is a difference of size perception on *BiasError* between the *HeadMove* and *ObjectMove*;

- **H2** *Stereo* will have lower *BiasError* than *NonStereo* because of its additional depth cue.

These hypothesis were based on the combination of previous research on the duality of size perception in pictures [7, 32] and the observations made using FTVR displays in lab.

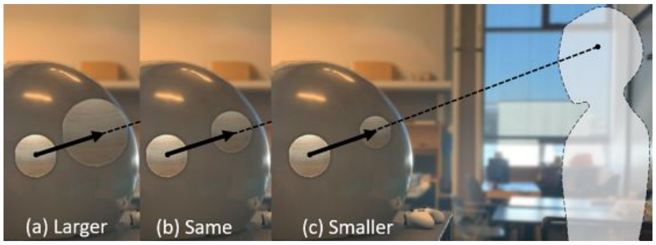

Figure 5. In *ObjectMove*, the task for participants is to judge whether the ball's size is (a) becoming larger (b) the same, or (c) becoming smaller at the same time as the ball was moving towards them. While this figure is a showcase of *ObjectMove*, the task is the same for *HeadMove*. The only difference is that participants move towards the ball in *HeadMove*.

Stimuli

Similar to other size perception studies [14, 22, 34], we chose a spherical stimuli due to its isotropic shape. We use the same texture and shadow pictorial cues across conditions to help users perceive ball depth. A shadow is dropped to appear on a plane overlapping the physical black surface holding the display as shown in Figure 4. It is necessary to render a shadow to indicate the ball's position. In particular, without stereopsis, this is the primary visual cue that indicates ball depth. We chose a wooden texture so viewers do not have an obvious size feature. This encourages viewers to perceive the size based on the volume in 3D rather than 1D or 2D features like the length of a checkerboard pattern. In real life, people do see wooden balls but have no prior knowledge about their exact size, which minimizes prior size bias.

Procedure

Participants started by filling out the consent form after verbal explanations of the study. We measured the interpupillary distance (IPD) of each participant with a ruler tape and calibrated the viewpoints based on the IPD [28]. Prior to the study, they underwent a stereo acuity test to confirm the eligibility [11]. Then they were seated on a fixed chair in front of the spherical display. They were instructed to place their head in a position where the head could gently touch a wooden bar rigidly attached on the chair to ensure the viewing distance d of 34 inches as shown in Figure 4. To ensure the consistency of the moving velocities in *HeadMove* and *ObjectMove*, participants were instructed to perform forward movements toward the screen paced by an audible electric metronome at 1.5 Hz similar to [4, 20]. We measured the velocity of their head movements before the study and set the measured velocity to all conditions.

As shown in Figure 4, in *HeadMove*, participants were required to judge the size change of the ball placed at A, while moving their heads forward towards the display from D to C. They moved their head 10 inches from 24" to 34" with the velocity paced by the metronome. The viewing distance is chosen to match the 24" diameter spherical display. Participants were presented with two successive stimuli per trial and they had to choose among three alternatives of smaller, unchanged and larger using a controller.

In *ObjectMove*, participants were required to judge the size change of the ball while pressing a button on the controller

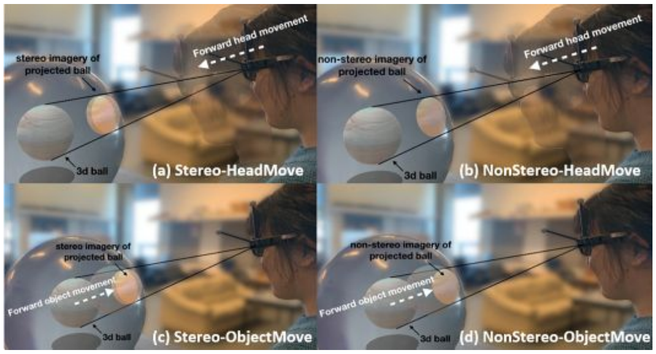

Figure 6. We followed a 2x2 within-subjects design with four conditions: (a) *Stereo-HeadMove*, (b) *NonStereo-HeadMove*, (c) *Stereo-ObjectMove*, and (d) *NonStereo-ObjectMove*. In *Stereo*, participants visualize stereoscopic imagery while in *NonStereo* they visualize monocular imagery set to the mid-point of two eyes. For illustrative purposes, the ghosted sphere shows the stereo condition while the ghosted user shows head movement.

to move the ball towards them. They moved the ball 10 inches from A to B at the pre-measured velocity paced by the metronome while keeping their heads stationary at the origin D. Likewise in *HeadMove*, they were presented with two successive stimuli per trial and reported the answer among three alternatives using a controller.

In both conditions, the movement of head/object causes the viewing distance d to decrease from 34 inches to 24 inches. This is to ensure the visual angle changes in the same way across conditions. Each participant conducted 12 data trials plus 3 practice trials per condition, resulting in 48 $(12x4)$ data observations for *BiasError*. The 12 data trials always contain 4 larger, 4 smaller and 4 unchanged stimuli in a random sequence at the resizing ratios of either 0% (4 unchanged), 15% (2 larger, 2 smaller) or 30% (2 larger, 2 smaller). The initial diameter of the ball is randomized between 4 and 6 inches. For each condition, we presented two likert scale questions (confidence and realism) to participants and asked them to rate each with a number in the range -2 (“totally disagree”) to 2 (“totally agree”).

Early pilots of this experiment showed that repeated toggling between *HeadMove* and *ObjectMove* was disorienting; to minimize this disorientation, participants only toggled between *HeadMove* and *ObjectMove* once and switched viewing conditions within each movement condition. Hence we counterbalanced the movement conditions (2), and also counterbalanced the viewing conditions (2) within each movement condition (2), resulting in 8 $(2x2x2)$ different sequences.

Participant

Seventeen participants (12 male and 5 female) from a local university were recruited to participate in the study with compensation. All participants passed the stereo acuity test. The average age of all participants was 29 years old from 18 to 35 years old. All participants completed written informed consent. We also asked participants to provide their previous experience and usage with Virtual Reality displays with a scale from 1 (“never”) to 4 (“regularly”). The average score of all participants was 2.4.

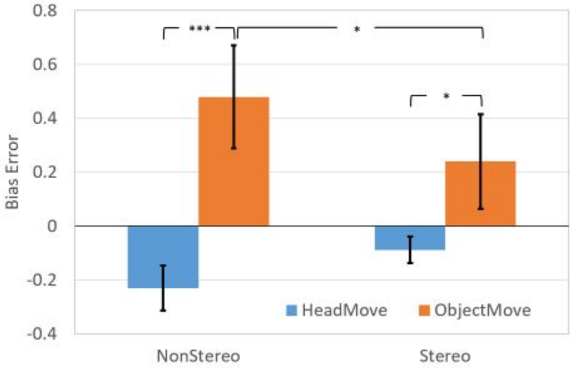

Figure 7. *BiasError* with means and 95% confidence intervals (CI). *BiasError* of *HeadMove* are below zero while *BiasError* of *ObjectMove* are above zero, showing a trend of underestimation in *HeadMove* and overestimation in *ObjectMove*. Significance values are reported in brackets for $p < .05(*)$, $p < .01(**)$, and $p < .001(***)$.

Apparatus

We implemented the *spherical* FTVR display with four mini-projectors and an acrylic, specially coated, spherical screen based on [35]. As shown in Figure 4, four projectors back-project onto the spherical screen through the projection hole at the bottom of the sphere. The 24” diameter spherical screen has a projection hole of 18” inch diameter. The projectors are Optoma ml750st with $1024x768$ pixel resolution with the frame rate of $120Hz$. Shutter glasses are synchronized with the projector to generate stereo images with each eye of $60Hz$. A host computer with an NVIDIA Quadro K5200 graphics card sends rendering content to the projectors. We employ head-tracking to generate perspective-corrected views. The viewer is tracked using OptiTrack (NaturalPoint Inc., Corvallis, OR) optical tracking system with passive markers attached to the stereo glasses as shown in Figure 1. We use Unity3D to create our 3D content of the study with a two-pass rendering approach to generate perspective-corrected imagery based on tracked viewpoints [10].

To render the correct and undistorted imagery based on the user’s viewpoint, there are two important calibration procedures: a screen calibration method to blend multiple projections as well as generate undistored images, and a perceptual calibration method to accurately register the tracked viewpoint with respect to the display. We calibrate the spherical display using an automatic calibration approach [36] with an on-surface error less than 1mm. We use a pattern-based viewpoint calibration [28] to register user-specific viewpoint with respect to the display with average angular error of less than one degree. The total latency is between 10-20 msec [10].

Result

Data did not meet the normality assumption of anova. A Friedman ranked sum test was performed and revealed a significant difference across conditions in *BiasError* ($\chi^2(3) = 32.0,\ p < .001$). Pairwise post-hoc Wilcoxon signed rank test for multiple comparisons with Bonferroni correction shows *BiasError* for *HeadMove* $(M = -0.23, SE = 0.043)$ is significantly lower ($W = 151,\ p < .001$) than *ObjectMove* $(M = 0.48, SE = 0.098)$ when viewing in *NonStereo*; *BiasError* for

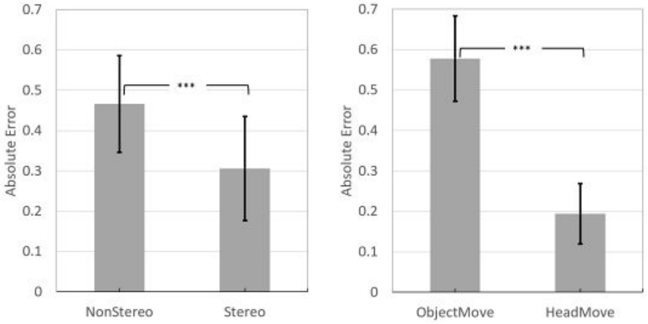

Figure 9. ***AbsoluteError*** **with means and 95% confidence intervals.** ***NonStereo*** **shows higher error than** ***Stereo*****;** ***ObjectMove*** **shows higher error than** ***HeadMove*****. Significance values are reported in brackets.**

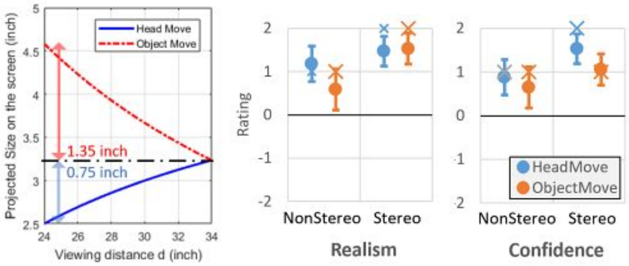

Figure 8. (Left) The projected size on the spherical screen varied when the viewing distance decreased from 34" to 24" during the experiment. The absolute gradient of the projected size in ***ObjectMove*** **is 1.8 times larger compared to** ***HeadMove*****. (Right) Participants' rates with means (circle), medians (cross) and 95%CI from -2 "totally disagree" to 2 "totally agree" to the questions of to how confident and real they felt about the reported result and stimuli.**

HeadMove $(M = -0.088, SE = 0.025)$ is significantly lower $(W = 119,\ p < .05)$ than *ObjectMove* $(M = 0.24, SE = 0.090)$ when viewing in *Stereo*. *BiasError* for *Stereo* $(M = 0.24, SE = 0.090)$ is significantly lower ($W = 94,\ p < .05$) than *NonStereo* $(M = 0.48, SE = 0.098)$ when performing *ObjectMove*, but not when performing *HeadMove* ($W = 20.5,\ p = .108$). The mean *BiasError* with 95% CI is shown in Figure 7. Given the result, we reject the NULL hypotheses of `H1` and `H2` and accept them.

An One-sample Wilcoxon signed rank test was performed on *BiasError* indicated a significant underestimation of size ($W(\mu < 0) = 136,\ p < 0.001$) when performing *HeadMove* and overestimation ($W(\mu > 0) = 140,\ p < 0.001$) when performing *ObjectMove*. The mean underestimation rate in *HeadMove* and overestimation rate in *ObjectMove* is 83.3% when viewing in *NonStereo*, and reduce to 64.7% when viewing in *Stereo*.

A Friedman ranked sum test was performed on the likert scale questions of the confidence and realism. Results revealed a significant difference across conditions in the confidence ($\chi^2(3) = 10.9,\ p < 0.05$) and realism ($\chi^2(3) = 8.86,\ p < 0.05$). Post-hoc Wilcoxon signed rank test for multiple comparisons with Bonferroni correction did not show significant difference between any pairs. Results of the mean, median and 95% CI are shown in Figure 8(Right).

DISCUSSION

Results show that participants systematically underestimated size in *HeadMove* and overestimated size in *ObjectMove* when retina images are the same across conditions. This indicates the on-screen size of the 2D imagery affects the perceived object size in 3D space. In particular, when participants moved closer to the object in *HeadMove*, they had a tendency to report objects as smaller. This contradicts reality as we usually expect closer objects look larger due to a larger visual angle. In addition, the bias appears to be stronger in *ObjectMove* than *HeadMove* shown in Figure 7. One potential explanation is that the absolute gradient of the projected size along the viewing distance is steeper in *ObjectMove* than *HeadMove*, with the on-screen imagery in *ObjectMove* gradient 1.8 times larger compared to *HeadMove* as shown in Figure 8(Left). To encourage perception of 3D features, users were instructed to attend to the change of the ball. Note that if they focused on the change of the shadow or the distance between shadow and ball instead, their performance will still be consistent because they are visualizing the scale of the entire scene and any geometries in the scene are subject to the projection model. However, it may impact the extent of the effect, which may account for the performance variance shown in Figure 7.

Knowing the effect of the on-screen imagery helps us to understand spatial perception in the screen-based 3D displays. One of the perceptual errors is the size underestimation of virtual objects [14, 20, 25]. Virtual objects are usually located behind the screen in the screen-based 3D displays [23, 25, 34], rendering an on-screen imagery smaller than the actual size as shown in Figure 2. If the size perception regresses towards the on-screen imagery similar to the phenomenal regression to the real object [7], viewers will have a tendency to report a modified value for the perceived object size, biased towards the on-screen size, resulting in the underestimation of the reported value. Hence the on-screen imagery could be a possible source for the size underestimation. On the other hand, if virtual objects are rendered in front of the screen, the on-screen imagery will be larger than the actual size. Therefore, we expect viewers will overestimate the size, which requires future experiments to investigate the effect of the on-screen when rendering on different side of the screen.

Adding the stereo cue significantly mitigated the systematic bias in *BiasError* as shown in Figure 7. This is consistent with Stefanucci's study [25] in which they found the size underestimation was alleviated by the addition of stereo as a depth cue in the display. As *BiasError* does not directly reflect the accuracy [1], to better understand the effect of *Stereo* on the performance, we also computed the *AbsoluteError*, defined as the average absolute difference between the reported answer and the correct answer. As shown in Figure 9, *AbsoluteError* is significantly lower ($F(1, 16) = 21.1,\ p < .001$) with *Stereo* than *NonStereo*, suggesting the addition of the stereo cue improved the size perception with better accuracy and less systematic bias. The result is consistent with Ware's discussions on the duality of depth perception in picture: the amount and effectiveness of the depth cues can help viewers to judge the size

---

[1] *BiasError* can be negative or positive so that the average *BiasError* can be zero while the absolute mean error is much greater than zero.

of a depicted object in a 3D space rather than on the picture plane [32]. This is similar to the scenario when people view a painting in the real world. They can choose to see the depicted object as a 3D structure, or as a 2D picture surface. The painter creates spatial vividness by adding various pictorial depth cues in their work to strengthen the illusion. Similarly, the stereo cue in our study is also an additional depth cue, making it easier to see in a '3D mode', rather than in a '2D mode', reducing sensitivity to the on-screen cues. Hence, it is suggested that future designs of FTVR displays should include stereopsis to alleviate the influence of on-screen cues.

AbsoluteError of *ObjectMove* is also significantly higher ($F(1,16) = 45.8,\ p < .001$) than *HeadMove*. Consistent with the performance data, participants' rating on the confidence of their performance in *HeadMove* is also slightly higher than *ObjectMove* as shown in Figure 8(Right). In the study, we provided limited pictorial cues with a plain background to keep the visual stimuli simple. One possible explanation for the performance difference is that the plain background did not provide sufficient depth information when the object moved towards participants in *ObjectMove*, compared to *HeadMove*, in which they might have a better depth perception via self movement. As depth and size perception are closely related, the lack of depth cues might be the cause of the difference in *AbsoluteError*. Additionally, *HeadMove* provides extra proprioception cue compared to the object movement. Participants could see the changes of real-world objects as they approached the screen, while the only visual cue in *ObjectMove* is the virtual object. Therefore, it is suggested that the head movement should be considered for better size perception.

We designed the study using *HeadMove* and *ObjectMove* to provide equivalent retinal images across conditions while rendering oppositely varied size of the on-screen imagery. Naturally, in this study, we cannot decouple the on-screen imagery from the movement itself, making it ambiguous to interpret the result as whether the bias is caused by different on-screen cues or different movement types. In reality, when our eyes get closer to an object, via either head or object movement, we always feel closer objects look larger due to the increased visual angle. Hence it seems unlikely that the movement types would be responsible for the opposite *BiasError* observed in our study, though future studies are required to separate the movement types with the on-screen cues.

Finally, as our study found that the on-screen imagery influenced users' spatial perception, we summarize the following design recommendations for FTVR displays. First, it is suggested to include the stereo cue to reduce the effect and make sure users perceive 3D scenes in a way as expected. Similar to the stereo cue, other depth cues such as pictorial cues or motion parallax cue may help to reduce the bias. Second, we expect different projection functions, such as orthographic or weak perspective projection, may help to reduce the effect, because the perspective projection is the primary cause for the on-screen changes. Projection functions with a stable on-screen imagery such as the orthographic projection, would provide more consistent size perception, though at the potential cost of viewing naturalness [30]. Third, the head movement should be considered to provide better size perception for 3D applications that require realistic visual perception.

LIMITATIONS AND FUTURE WORK

There are several limitations to the experiment presented in this paper. We chose an abstract task with simple stimuli on a plain background. While it reduced complexity and eliminated variance, the task does not directly represent real-world use cases. We chose closer/farther head/object movement as it represents a typical scenario when users move closer for a careful inspection of 3D content, such as in computer aided design and scientific visualization. However, in the real-world scenario, head movements may be up-down, left-right, or a combination of all possible directions. Future studies with a more realistic task are required, such as a walk-around visualization task with free head movement to further characterize the perceptual trade-offs.

Our study was performed on a spherical FTVR display. We chose a spherical display as it is the most common form-factor for recent FTVR displays. However, the design of the task, the analysis of the projection model, and the object/head movement are independent of the display shape and generalize to all screen-based 3D displays. We expect the findings extend to other display shapes such as planar or cylindrical displays. We also expect the perceptual bias to be more pronounced when the size of the stimuli is comparable to the size of the screen, because the screen can serve as a reference that makes the local changes of the on-screen imagery more noticeable. It is also interesting to consider how these findings could be transferred to HMDs. Unlike FTVR displays, users move together with HMD screens. Ideally, there is no relative movement between the viewpoint and screen when users wear HMDs tightly. Therefore we expect results will be different from our study as the projected size will increase in both *HeadMove* and *ObjectMove*, making closer objects look larger. To generalize beyond spherical FTVR displays, it is necessary to conduct similar studies on other FTVR and screen-based 3D displays.

CONCLUSION

FTVR displays render perspective-corrected imagery on a 2D screen. Because users perceive a 3D object by looking at pixels on the 2D screen, there exists a perceptual duality between the on-screen pixels and the 3D percept. In this work, we presented an empirical study evaluating the effect of the on-screen imagery interfering with users' size perception using a size-judgement task in different viewing conditions, including stereopsis. We found that the size of on-screen imagery significantly influenced object size perception, causing 83.3% under/ over-estimation of perceived size when viewing without stereopsis and reducing to 64.7% with stereopsis. Contrary to reality, objects look smaller when viewers get closer. Understanding the effect of on-screen cues helps us to provide accurate perception of real-world objects in the virtual environment.

ACKNOWLEDGMENTS

The authors wish to thank reviewers for their constructive comments and NSERC for funding this project.

REFERENCES

[1] Roland Arsenault and Colin Ware. 2004. The importance of stereo and eye-coupled perspective for eye-hand coordination in fish tank VR. *Presence: Teleoperators and Virtual Environments* 13, 5 (2004), 549–559.

[2] Kevin W Arthur, Kellogg S Booth, and Colin Ware. 1993. Evaluating 3d task performance for fish tank virtual worlds. *ACM Transactions on Information Systems (TOIS)* 11, 3 (1993), 239–265.

[3] Hrvoje Benko, Andrew D Wilson, and Federico Zannier. 2014. Dyadic projected spatial augmented reality. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM, 645–655.

[4] Mark F Bradshaw, Andrew D Parton, and Richard A Eagle. 1998. The interaction of binocular disparity and motion parallax in determining perceived depth and perceived size. *Perception* 27, 11 (1998), 1317–1331.

[5] Carolina Cruz-Neira, Daniel J Sandin, and Thomas A DeFanti. 1993. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*. ACM, 135–142.

[6] Robert G Eggleston, William P Janson, and Kenneth A Aldrich. 1996. Virtual reality system effects on size-distance judgements in a virtual environment. In *Virtual Reality Annual International Symposium, 1996., Proceedings of the IEEE 1996*. IEEE, 139–146.

[7] Kevin W Elner and Helen Wright. 2015. Phenomenal regression to the real object in physical and virtual worlds. *Virtual Reality* 19, 1 (2015), 21–31.

[8] Dylan Fafard, Ian Stavness, Martin Dechant, Regan Mandryk, Qian Zhou, and Sidney Fels. 2019. FTVR in VR: Evaluation of 3D Perception With a Simulated Volumetric Fish-Tank Virtual Reality Display. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 533.

[9] Dylan Fafard, Andrew Wagemakers, Ian Stavness, Qian Zhou, Gregor Miller, and Sidney S Fels. 2017. Calibration Methods for Effective Fish Tank VR in Multi-screen Displays. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 373–376.

[10] Dylan Brodie Fafard, Qian Zhou, Chris Chamberlain, Georg Hagemann, Sidney Fels, and Ian Stavness. 2018. Design and implementation of a multi-person fish-tank virtual reality display. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. ACM, 5.

[11] Davide Gadia, Gianfranco Garipoli, Cristian Bonanomi, Luigi Albani, and Alessandro Rizzi. 2014. Assessing stereo blindness and stereo acuity on digital displays. *Displays* 35, 4 (2014), 206–212.

[12] James J Gibson. 1971. The information available in pictures. *Leonardo* 4, 1 (1971), 27–35.

[13] Ralph Norman Haber. 1980. How We Perceive Depth from Flat Pictures: The inherent dual reality in pictorial art enables us to perceive a scene as three dimensional at the same time we see that the painting or photograph is actually flat. *American Scientist* 68, 4 (1980), 370–380.

[14] Jonathan W Kelly, Lucia A Cherep, and Zachary D Siegel. 2017. Perceived space in the HTC vive. *ACM Transactions on Applied Perception (TAP)* 15, 1 (2017), 2.

[15] Jonathan W Kelly, Lisa S Donaldson, Lori A Sjolund, and Jacob B Freiberg. 2013. More than just perception–action recalibration: Walking through a virtual environment causes rescaling of perceived space. *Attention, Perception, & Psychophysics* 75, 7 (2013), 1473–1485.

[16] Robert V Kenyon, Daniel Sandin, Randall C Smith, Richard Pawlicki, and Thomas Defanti. 2007. Size-constancy in the CAVE. *Presence: Teleoperators and Virtual Environments* 16, 2 (2007), 172–187.

[17] Kibum Kim, John Bolton, Audrey Girouard, Jeremy Cooperstock, and Roel Vertegaal. 2012. TeleHuman: effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2531–2540.

[18] Sirisilp Kongsilp and Matthew N Dailey. 2017. Motion parallax from head movement enhances stereoscopic displays by improving presence and decreasing visual fatigue. *Displays* 49 (2017), 72–79.

[19] Billy Lam, Yichen Tang, Ian Stavness, and Sidney Fels. 2011. A 3D cubic puzzle in pCubee. In *3D User Interfaces (3DUI), 2011 IEEE Symposium on*. IEEE, 135–136.

[20] Xun Luo, Robert Kenyon, Derek Kamper, Daniel Sandin, and Thomas DeFanti. 2007. The effects of scene complexity, stereovision, and motion parallax on size constancy in a virtual environment. In *Virtual Reality Conference, 2007. VR'07. IEEE*. IEEE, 59–66.

[21] Kevin Ponto, Michael Gleicher, Robert G Radwin, and Hyun Joon Shin. 2013. Perceptual calibration for immersive display environments. *IEEE transactions on visualization and computer graphics* 19, 4 (2013), 691–700.

[22] Zachary D Siegel, Jonathan W Kelly, and Lucia A Cherep. 2017. Rescaling of perceived space transfers across virtual environments. *Journal of Experimental Psychology: Human Perception and Performance* 43, 10 (2017), 1805.

[23] Ian Stavness, Billy Lam, and Sidney Fels. 2010. pCubee: a perspective-corrected handheld cubic display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1381–1390.

[24] Ian Stavness, Florian Vogt, and Sidney Fels. 2006. Cubee: a cubic 3D display for physics-based interaction. In *ACM SIGGRAPH 2006 Sketches*. ACM, 165.

[25] Jeanine K Stefanucci, Sarah H Creem-Regehr, William B Thompson, David A Lessard, and Michael N Geuss. 2015. Evaluating the accuracy of size perception on screen-based displays: Displayed objects appear smaller than real objects. *Journal of Experimental Psychology: Applied* 21, 3 (2015), 215.

[26] Maurice HPH van Beurden, Andre Kuijsters, and Wijnand A IJsselsteijn. 2010. Performance of a path tracing task using stereoscopic and motion based depth cues. In *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 176–181.

[27] D Vishwanath. 2011. Information in surface and depth perception: Reconciling pictures and reality. *Perception beyond inference* (2011), 201–240.

[28] Andrew Wagemakers, Dylan Fafard, and Ian Stavness. 2017, to appear. Interactive visual calibration of volumetric head-tracked 3D displays. In *2017 SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM.

[29] Andrew John Wagemakers, Dylan Brodie Fafard, and Ian Stavness. 2017. Interactive visual calibration of volumetric head-tracked 3d displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3943–3953.

[30] Leonard C Wanger, James A Ferwerda, and Donald P Greenberg. 1992. Perceiving spatial relationships in computer-generated images. *IEEE Computer Graphics and Applications* 3 (1992), 44–51.

[31] Colin Ware. 2006. 3D contour perception for flow visualization. In *Proceedings of the 3rd symposium on Applied perception in graphics and visualization*. ACM, 101–106.

[32] Colin Ware. 2012. *Information visualization: perception for design*. Elsevier.

[33] Colin Ware and Glenn Franck. 1996. Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics (TOG)* 15, 2 (1996), 121–140.

[34] Qian Zhou, Georg Hagemann, Dylan Fafard, Ian Stavness, and Sidney Fels. 2019. An Evaluation of Depth and Size Perception on a Spherical Fish Tank Virtual Reality Display. *IEEE transactions on visualization and computer graphics* 25, 5 (2019), 2040–2049.

[35] Qian Zhou, Georg Hagemann, Sidney Fels, Dylan Fafard, Andrew Wagemakers, Chris Chamberlain, and Ian Stavness. 2018. Coglobe: a co-located multi-person FTVR experience. In *ACM SIGGRAPH 2018 Emerging Technologies*. ACM, 5.

[36] Qian Zhou, Gregor Miller, Kai Wu, Daniela Correa, and Sidney Fels. 2017. Automatic calibration of a multiple-projector spherical fish tank vr display. In *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*. IEEE, 1072–1081.