Computers & Graphics 117 (2023) 105–113

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Technical Section

Perceptual thresholds of visual size discrimination in augmented and virtual reality

Liwen Wang [a], Shaoyu Cai [b,*], Christian Sandor [c]

[a] School of Creative Media, City University of Hong Kong, 18 Tat Hong Ave, Kowloon Tong, Hong Kong
[b] Engineering Design & Innovation Centre, College of Design and Engineering, National University of Singapore, Singapore
[c] Laboratoire Interdisciplinaire des Sciences du Numérique (LISN), Université Paris-Saclay/CNRS, Orsay cedex, France

A R T I C L E I N F O

Article history:
Received 2 May 2023
Received in revised form 27 September 2023
Accepted 2 October 2023
Available online 5 October 2023

Keywords:
Augmented reality
Virtual reality
Size perception
Psychophysical experiment

A B S T R A C T

The perception of size in virtual objects in Augmented Reality (AR) and Virtual Reality (VR) is a not trivial issue, as the effectiveness of manipulating and interacting with virtual content depends on the accuracy of size perception. However, there are missing straightforward comparisons between VR and AR in terms of size perception for the deep understanding of size perceptual differences. Understanding these perceptual differences can inform designers on how to adapt content when transitioning between these two spatial computing platforms. In this paper, we conducted two psychophysical experiments to measure the perceptual thresholds of size discrimination for virtual objects. Our results indicated that users are more sensitive to size changes in VR than in video see-through AR, suggesting that size differences are easier to be perceived in VR than in AR. Additionally, for increase or decrease of sizes, the accuracy of judgments showed an asymmetric trend in video see-through AR.

1. Introduction

Recently, Augmented Reality (AR) and Virtual Reality (VR) have gained increasing attention in various fields [1]. AR and VR, as well as their social [2], educational [3,4], and entertainment-related [5,6] applications, have attracted a large amount of research interest. For these applications, understanding and quantifying human perception is essential, as it enables us to design more effective interaction and visualization techniques. Visual perception has received considerable attention in psychology and cognitive science for a long time [7]. In our visual experience of AR and VR, accurately sized visual content can effectively and correctly convey information to users [8] and even evoke connections between multiple senses [9]. This paper examines how changes in the actual size of visual content affect people's perception and responses.

While many works have focused on the relationship between size and spatial properties (i.e., depth) in AR and VR [10], only a few studies have addressed size perception and its perceptual thresholds. Stefanucci et al. [11] used affordance judgments to compare the accuracy of size perception between real objects and screen-based displays. They concluded that the apparent size of virtual objects on the screen was underestimated. Additionally, Thomas [12] conducted several studies on the size perception of virtual cylinders in a VR system and calculated the precise thresholds of size perception. His results revealed that the just-noticeable difference (JND) of virtual objects is a very tiny value close to the reference object. Although these studies have investigated the size perception of virtual objects, no current research compares size perception between AR and VR. Our study aims to fill this gap by investigating AR vs. VR through a video see-through head-mounted display. Understanding this difference can help stakeholders (e.g., 3D modelers) create better AR and VR content or adapt AR content to VR and vice versa. For example, when a 3D modeler wants to indicate a different status of a switch (e.g., on/off) through a size change in AR, the same 3D model should only be used in AR and VR if they induce the same perceptual sensitivity of size discrimination.

Inspired by previous studies that have shown variations in perceptual sensitivities for visual perception [13] or haptic perception [14] in VR/AR environments, highlighting the distinctions between virtual and real environments, we formulate the following research question:

- Does human visual perceptual sensitivity for size differ between AR and VR environments?

To address this research question, we conducted two experiments based on the psychophysical method [15] to evaluate the detection thresholds for size discrimination in AR vs.

* Corresponding author.
E-mail addresses: liwen.wang@my.cityu.edu.hk (L. Wang), shaoyucai.aa@gmail.com (S. Cai), christian.sandor@universite-paris-saclay.fr (C. Sandor).

https://doi.org/10.1016/j.cag.2023.10.001

VR. Specifically, we adopted the two-alternative force choice (2AFC) method, allowing users to compare different increased volumes (from 850 to 1150 cm^3) of virtual cubes with the size of 10 × 10 × 10 cm^3 as the reference. Using the collected data, we derived psychometric functions for both AR and VR, enabling us to determine perceptual thresholds.

In this study, we found that the detection thresholds for size change are 952.5–1081.6cm^3 in AR and 949.8–1062.2 cm^3 in VR, under the same experimental conditions and reference (10 × 10 × 10 cm^3 cube). The point of subject equality (PSE) values for AR and VR are 1022.6 (2.3%) cm^3 and 1010.9 (1.1%) cm^3, respectively, and the just-noticeable difference (JND) values are 17.05 and 6 for AR and VR, respectively.

Our experimental results have revealed that participants are more sensitive to size discrimination in the VR environment. In other words, objects are perceived as larger in VR compared to AR, with a cube that is larger than the reference by 1020 (2%) cm^3 being perceived as larger than the reference in VR, but in AR, it is perceived as no different from the reference and as small as the reference. Moreover, for size increase or decrease, the accuracy of judgments showed an asymmetric trend in AR.

2. Related work

AR can directly link physical reality and virtual information about the real world [16]. Correspondingly, VR can create an imaginative virtual world by using an approach for creating concept shape designs [17]. Furthermore, measuring perception levels will become an essential indicator of how mixed reality integrates into real-world life. Based on our research goal, we discuss our related work in the following three parts: size perception for virtual content, psychophysical experiments of visual contents in AR/VR, and AR versus VR comparison.

2.1. Size perception for virtual content

Size perception is basic in human interaction, and many studies have explored its effects. Considering size effects is important when examining the effectiveness of visualization or physicalization variables in real environments [18–22]. Size discrimination is based on three factors: size constancy, perspective, and visual view on the retina [23–25]. Size constancy [23] refers to objects of a known size appearing the same regardless of their position from the viewer. To minimize size constancy impact, we did not inform users of object size beforehand or provide a constant reference frame around stimuli. Perspective [24], such as converging lines, can create an illusion of size changes. Visual view subtended by objects on the retina [25] also influences perceived size. An object that subtends a larger visual view on the retina is perceived as larger. Understanding those factors that influence size perception in general can help better comprehend their impact in AR/VR displays.

Several studies have investigated size perception in AR and VR displays, including Ahn et al. [26], who verified the correct size perception of augmented objects among three types of augmented reality devices (hand-held mobile device, video see-through HMD, and optical see-through HMD). Their results showed that users had an estimating bias in size perception between different AR displays. One interesting conclusion was that augmented objects from video see-through HMDs performed most near 1–1 scale matching to the actual reference object. Additionally, Thomas [12] conducted a series of psychophysical experiments to measure the perceptual size of virtual cylinders in VR. He found that there existed a difference between stimuli and references, but the differences were tiny. Zhou et al. [27] investigated the duality of size perception using a spherical fish tank VR display. The experimental results showed that depth cues (3D) and non-stereo cues (2D) could affect size perception and even lead to under/overestimating virtual objects. Kim et al. [28] explored the influence of interpupillary distance and eye height on the size perception of a virtual white cube in VR. Their results showed that eye height could not evoke different size perceptions, but virtual eye separation could.

In summary, the studies mentioned above demonstrate the feasibility of conducting experiments on size perception in various AR and VR displays. However, our study builds upon these previous studies by conducting experiments about size perception in AR/VR using a psychophysical approach.

2.2. Psychophysical experiments of visual contents in AR/VR

Psychophysical methods like the method of limits, the method of constant stimuli, and the method of adjustment are commonly used to evaluate visual content [29]. Since the human eye is sensitive to size changes, the method of constant stimuli is preferred in psychophysical experiments involving visual stimuli [15,30]. This method randomly presents multiple stimuli to the observer for judgment and is highly regarded in visual psychophysics [31]. Rolland et al. [31] found that this method reduces bias compared to other methods, but with higher variability. They used it to render virtual objects in a binocular HMD with a depth-aware accuracy of 2 mm and precision of 8 mm. Overall, we chose the method of constant stimuli as the most appropriate choice for our research goals.

Our study employed psychophysical methods to investigate levels of perception, focusing on two computed metrics: the Point of Subjective Equality (PSE) and the Just Noticeable Difference (JND). Previous research has shown that JNDs for size perception in VR are less than 1.5 mm in height and less than 2.3 mm in width when referencing objects below 90 mm [12], indicating that size discrimination in AR and VR should be measured on a small scale. In summary, the above literature suggests that it is feasible to use psychophysical methods to investigate size discrimination in AR and VR, with a focus on the apparent size of objects.

2.3. Augmented reality & virtual reality comparison

Augmented reality (AR) and virtual reality (VR) are popular platforms for virtual content, but they differ in human perception. For instance, Gaffary et al. [14] compared haptic perception of stiffness using virtual pistons and found that AR felt softer than VR with similar setups. Ping et al. [32] compared depth perception between AR and VR using virtual balls and found higher depth estimation accuracy in AR than in VR, with increasing errors over longer distances. Jones et al. [13] measured egocentric depth perception in VR and AR environments and found that VR compressed virtual space, leading to underestimated depth perception. While visual scenes in AR and VR can influence human perception in various ways, including haptic and depth perception, there is currently no research that specifically explores differences in size discrimination between AR and VR.

3. User perception experiment: Size discrimination in augmented reality and virtual reality

3.1. Hypotheses

Previous research on visual perception has indicated that distance/depth underestimation is more prevalent in virtual reality (VR) compared to augmented reality (AR) environments [13].

However, it remains unclear whether such differences can influence size perception performance. Similarly, Gaffary et al. [14] compared haptic perception of stiffness in VR and AR and observed significant differences between the two conditions. Based on these findings, we hypothesize that humans may exhibit distinct performance in size perception between VR and AR conditions. Therefore, we propose the following hypotheses:

H1. The threshold of size perception in VR is bigger than that in AR.

Jones et al.'s study [13] demonstrated that users in VR perceive virtual space as compressed compared to the real world, leading to consistent underestimations of egocentric depth. Their results suggest that the VR background contributes to the observed underestimation effect in VR. Consequently, users in VR environments might perceive virtual objects as closer and, therefore, larger than their actual size. Based on these findings, we hypothesize that a similar underestimation effect may occur in our research, where the perceptual threshold for size perception in VR is larger than that in AR.

H2. The accuracy of judgments is symmetric for increases and decreases of sizes, both in AR and VR.

As we take 1000 cm^3 as the center and chose the symmetrical values on the increase and decrease sides, with the same step size on both sides, we expect that the judgments on both sides would be symmetric under both the AR and VR scenes.

3.2. Purpose

This experiment aims to compare size discrimination between augmented reality (AR) and virtual reality (VR). We examined perceptual levels of virtual cubes and their perceived size. Participants selected the larger cube in AR and VR scenarios in a counterbalanced order. We evaluated performance by calculating each user's answers.

3.3. Pilot study: Interval determination

We conducted a pilot study to determine the optimal interval for our size discrimination experiments. To avoid recognition memory issues and reduce time consumption in psychophysical experiments [33], we limited the interval within a reasonable range. While we aimed to cover a perceived range of size changes, we narrowed it down as much as possible.

We invited eight participants (five females and three males, aged 21 to 28, mean = 25.13, SD = 3.04) for our pilot study, assigning four to AR and four to VR experiments. Using the constant stimuli method, we set the upper and lower boundaries of the stimuli cube 30% larger and smaller than the reference cube, respectively. We created 12 cubes with volumes ranging from 700 to 1300 cm^3 and a 5% change step. Each participant discriminated 72 pairs of cubes by choosing the larger one. Pilot study results are shown in Fig. 1, with dark blue indicating correct answers. We defined a "correct answer" as correctly selecting the larger volume cube from two stimuli.

Results showed that users easily detected changes within ±10% to ±30%, with one participant stating, "I can almost tell these two objects had different sizes. The first one was bigger, obviously". Even at 1250 cm^3, there was a 4% probability of making wrong choices, which we consider acceptable. People had a 92% probability of choosing the larger cube for changes of +10%, and 90% for changes of −10%. Changes of ±5% had a 79% probability of choosing the larger cube in AR and VR. Based on these results, we selected the stimuli range of 850 cm^3 to 1150 cm^3 for size discrimination in our experiments.

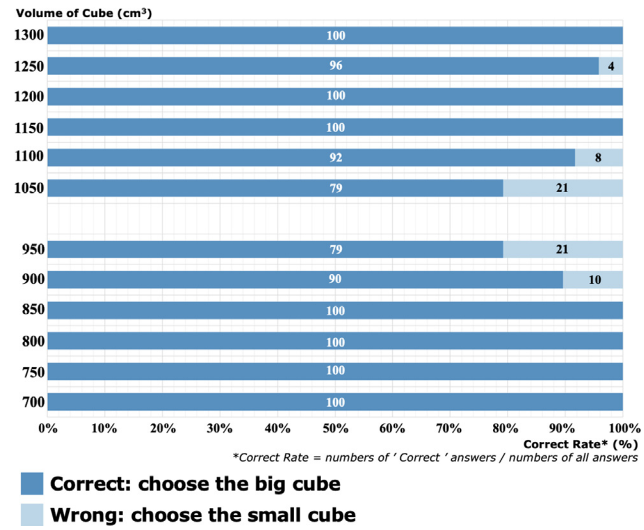

Fig. 1. The results of the pilot study. The X-axis represents the correct rate of users' responses in each condition. The dark blue bars indicate that the users made the correct answers under these stimuli, which means that they chose the larger cube in the pair. The light blue bars indicate that the users made the wrong choice and perceived the larger cube as the smaller one. The Y-axis represents the volume of virtual cubes.

3.4. User perception experiment: Size discrimination

3.4.1. Participants

22 participants took part in the experiment. Their ages ranged from 19 to 28 (Mean = 23.82, SD = 2.27). All participants were right-handed and had normal or corrected-to-normal vision. We defined the level of experience with AR or VR as follows: 0 - never, 1 - seldom (less than once a month), 2 - often (more than once a month), and 3 - expert (developer or related engineering). Regarding AR or VR experience, seven participants had no experience, 14 had used either device once or twice, and one participant had extensive knowledge of AR or VR. The average level of user experience was 0.77 (SD: 0.69). Participants received rewards for their efforts.

3.4.2. Apparatus and scenes

We used an HTC Vive Pro 2 device for AR and VR content creation. The HTC Vive Pro 2 HMD is a PC-powered display that renders computer graphics on a part of the field of view. It needs external base stations for complete tracking. We relied on Vive controllers and base stations for stable head and hand-tracking in our hardware tracking system.

Before using the HTC Vive Pro for our experiments, we conducted tests for technical details. Sauer et al. [34] successfully used the HTC Vive Pro to detect a checkerboard pattern, while Gil et al. [35] tested the device for high-color constancy performance. However, their tests did not include virtual objects, which were the focus of our experiment. Therefore, we conducted a specific test for virtual objects using pictures of a checkerboard pattern with squares in both AR and VR for calibration. Our HTC Vive Pro successfully detected the checkerboard patterns in both scenes, with straight and parallel lines. We also tested for image artifacts and the device passed the color fidelity test.

We utilized the built-in cameras of the HTC Vive Pro 2 for our AR system. By activating the dual camera above the HMD, we achieved our AR goal. The HTC Vive Pro 2 is a video-see-through HMD that minimizes lost visibility of virtual content in bright environments and maximizes the immersive experience. We used the HTC Vive Pro 2 for our VR system to ensure consistency and

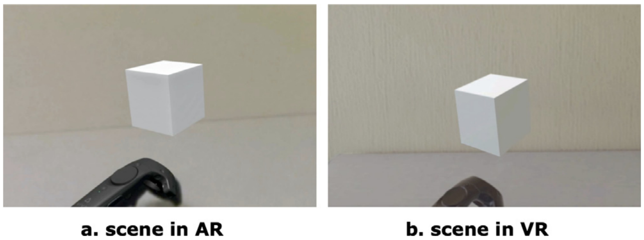

Fig. 2. AR and VR scenes in our experiments. Image (a) on the left shows a virtual cube above a hand controller in the AR scene. Image (b) on the right shows a virtual cube above a hand controller in the VR scene.

avoid external factors. The experimental setups were the same for AR and VR. Our system fulfilled the requirements for size discrimination between AR and VR.

In the AR scene, users explored virtual objects on a table while sitting in a clear corner to avoid referencing the surrounding environment for size perception. We replicated this arrangement in the VR scene by simulating the same colored walls and table as the real-world background. Consistency was maintained between the AR and VR scenes, as shown in Fig. 2.

Scenes were created in Unity (2018.4.30f) and run on a desktop with an AMD Ryzen9 3950X 16-core 3.49 GHz processor, an NVIDIA GTX 2080 graphics card, and Windows 10 Pro. The HTC Vive Pro 2 HMD displayed a resolution of 2448 × 2448 pixels and a field of view up to 120 degrees horizontally. The dual camera default setting in Stereo Pass-through mode with 720P captured the surroundings. The field of view (FOV) was checked to ensure that the size was the same in Augmented Reality (AR) and Virtual Reality (VR).

Participants used a Vive Controller as a grasping proxy, a hardware-based 6-DoF inside-out gamepad. The virtual cube was only a single front-side facing the participant when the controller faced them at eye level. Users' actions in both scenes matched what they saw without noticeable delay, and the virtual cube moved with the rotation of the controller accordingly. Users can view all sides of the virtual cube if desired. The height between the center of the virtual cubes and the controller remained stable, preventing users from judging size changes through displacement, regardless of stimuli changes. We did not closely regulate the movement of the participants towards the object, as our primary focus was on the metric dimensions of the object. In interactive AR or VR systems where users can adjust their viewpoint, they can freely manipulate the virtual content by looking from all sides and changing their position relative to the object. However, we did ask users to control the distance between the grasping proxy and their torso, ensuring it was not over their lower arm. The average length of lower arm for participants is 40 cm. Further information is provided in Fig. 3.

3.4.3. Stimuli and collection data

Two virtual cubes, consistent in texture, color, and shape, appeared one by one in the user's view. The reference stimulus had a volume of 10 × 10 × 10 cm^3. The primary visual conditions in this experiment are:

- **S** (stimuli) refers to the volume of the stimuli cubes. Ten values were chosen after the pilot study, corresponding to the sizes: 850, 880, 910, 940, 970, 1030, 1060, 1090, 1120, and 1150 cm^3.
- **R** (repeat times) refers to the number of times each stimulus was repeated, following the approach used in the classic work by Steincike et al. [15] and repeated each stimulus six times in our experiments.

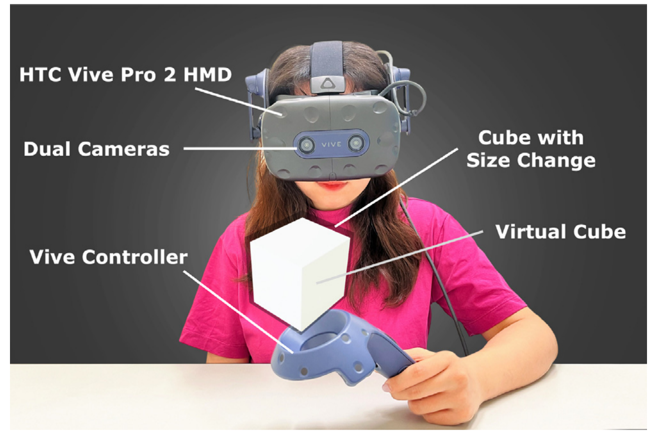

Fig. 3. A participant is holding a Vive hand controller to observe an increased cube. She is wearing the HTC Vive Pro 2 HMD to explore the AR scene.

Thus, each participant experienced 10 stimuli (**S**) × 6 repeat times for each stimulus (**R**) = 60 trials in one condition (AR or VR). Then, they repeated the same stimuli in the other scene. Therefore, at the end of the experiment, we collected 60 trials × 2 conditions = 120 pairs of responses from each participant.

3.4.4. Procedure

We conducted a within-subject user study with a counter-balanced AR and VR order. The procedure was kept consistent for both AR and VR conditions. To simplify, we only emphasize the experiment process in this section and omit descriptions of AR/VR.

During the experiment, each participant experienced a paperwork-block and an experiment-block. Paperwork-block introduced some background information about the experiment and collected some basic information about the participants. The experiment-block included a training session to familiarize participants with the task and a testing session to collect their responses.

Participants began with the paperwork-block which included providing information about the experiment and obtaining signed consent forms. Demographic information was also collected, followed by the completion of the Pre-SSQ questionnaire to assess physical and vision conditions. Participants who reported moderate or severe eye strain, difficulty focusing, blurred vision, or dizziness were advised to end the experiment. Once the paperwork-block was complete, participants moved on to the experiment-block, where they were assisted in putting on the HMD.

Participants interacted with the virtual scene using the controller while wearing the HMD. To calibrate consistently for each participant, we fixed the distance between the human eyes and the HMD screen in all experiments to minimize the effect of irrelevant variables. We asked them to adjust the IPD button according to HTC Vive guidance for the clearest vision during our experiments. To prevent HMD sickness, we suggested rotating the controller instead of their heads during free exploration. The training session included ten pairs of virtual cubes, with the first four pairs demonstrating possible changes and the last six pairs helping participants become familiar with the process. Additional pairs were provided if requested by participants. Answers from the training session were not recorded. The procedure for the testing session was the same as the training session. Since the training session and testing session share the same process, we explained the procedure together in the following.

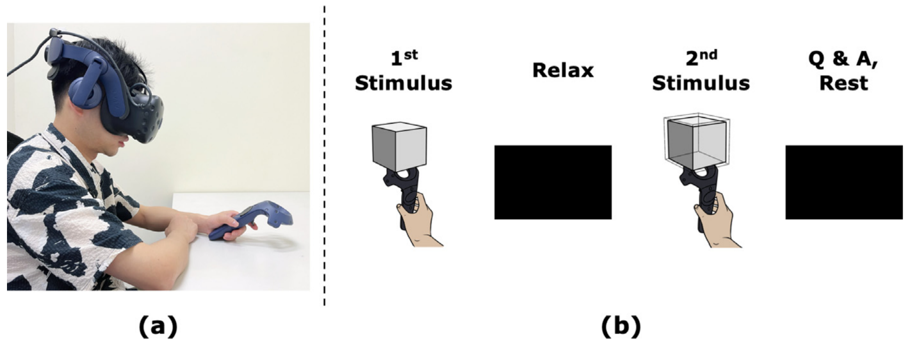

Fig. 4. (a): participant held a controller. (b): experiment procedure in the participant view.

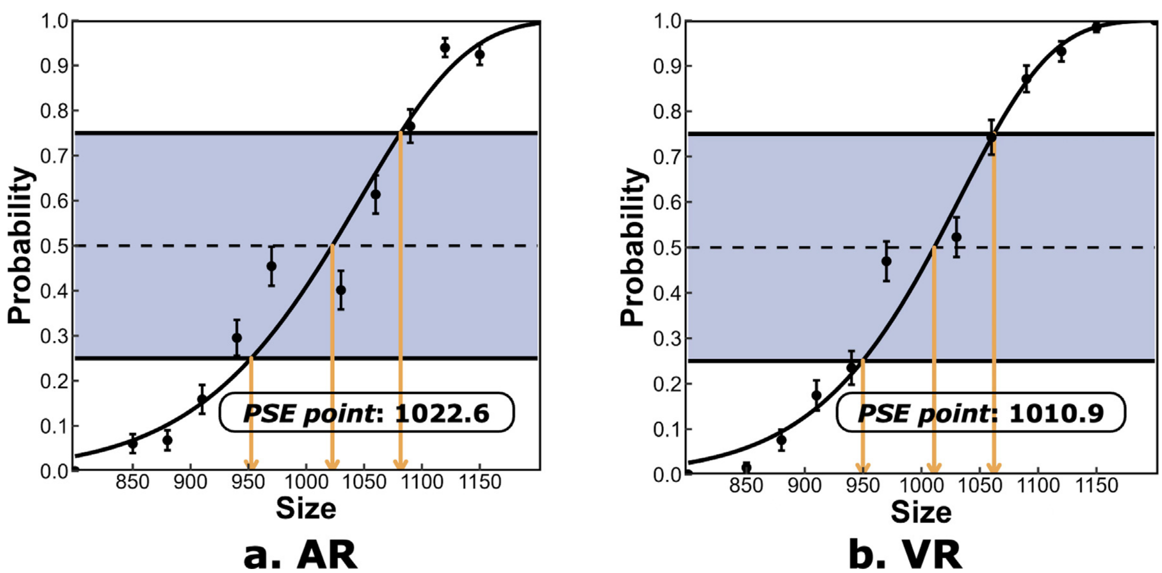

Fig. 5. Psychometric functions of size discrimination in AR and VR. (a) shows the range of 850 to 1150 cm^3 in AR. (b) shows the range of 850 to 1150 cm^3 in VR. X-axis represents the different volumes of virtual cubes. Y-axis shows the probability that users chose "stimulus".

During the training session, there were 60 trials, each divided into four parts: first stimulus, relax, second stimulus, and end-trial rest (Fig. 4). One of the first or second stimuli was the reference, randomly positioned as 1st or 2nd. Participants were asked to memorize the size of the virtual cube in the first stimulus part, with no time limit. After feeling confident in their memory, the view turned all-black for 1 s as a visual buffer before the second stimulus was presented. A *relax* part was necessary to reduce bias in the results and prevent sickness, with 1 s as a balanced duration. Participants were then asked, "Which one is bigger?" with no time limit to answer, but most of them answered directly. The *end-trial rest* followed with 1 s of black. Participants were informed that answers such as "I do not know"/"I can not answer" were not allowed.

After completing the four parts, one trial ended, and the next trial began. Once all trials were completed, participants filled out a Post-SSQ questionnaire to record any discomfort experienced during the experiment. The entire procedure took approximately 1 h per participant.

4. Results and analysis

4.1. Psychometric curves results

In our experiments, participants were asked to indicate if the presented stimuli (i.e., 850, 880, 910, etc.) were perceived to be larger than the reference stimuli (i.e., 1000), and we counted the number of times participants chose "stimulus". The probabilities under different stimuli are shown in Fig. 5, and we fitted this data into psychometric functions using the quickpsy [36] toolkit. Fig. 5 (a and b) shows the psychometric curves of AR and VR for all stimuli, and the meanings of the x-axis and y-axis are shown in the appendix. Additionally, we included the Interval of Uncertainty (IU) part, which is the 25%–75% range of response probabilities (shown in purple in Fig. 5), similar to Steinicke et al. [15]. Participants could not reliably detect size changes between two stimuli in this range. Using the sigmoidal function, we calculated the upper and lower boundaries of the IU, where 75% corresponds to the upper boundary of stimuli and 25% corresponds to the lower boundary. We also calculated the point of subjective equality (PSE) values, which represent the stimulus where participants have a 50% probability of choosing one choice from the reference and stimuli even if they are not the same. As shown in Fig. 5, the PSE value of size discrimination is 1022.6 (2.3%) cm^3 for AR and 1010.9 (1.1%) cm^3 for VR when the reference stimulus is 10 cm × 10 cm × 10 cm.

In addition to PSE values, we also report the just-noticeable difference (JND) value for our perceptual thresholds. JND is the smallest or least perceptible difference that can be perceived at least half the time[1]. We followed the method outlined in the book [37] to calculate the JND value. In this book, JND is defined as half of the 'interval of uncertainty' (purple part in

[1] https://en.wikipedia.org/wiki/Just-noticeable_difference

Table 1
The results of Interval of Uncertainty (25%–75%) and JND & PSE values.

Environments	Interval of Uncertainty (IU)			JND
	25%	50% (PSE)	75%	
AR	952.5	1022.6	1081.6	17.05
VR	949.8	1010.9	1062.2	6.00

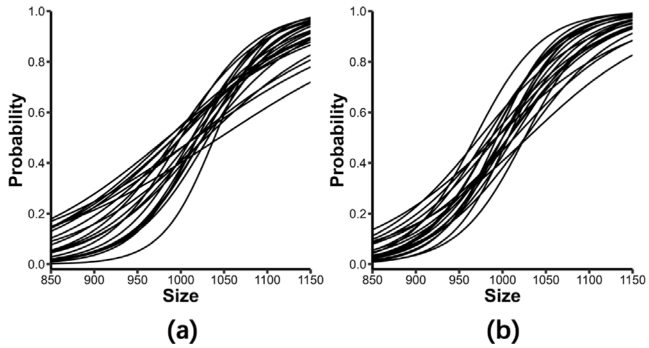

Fig. 6. The psychometric function curves of size discrimination for each participant in AR and VR are depicted below. Figure (a) illustrates the curves in AR, and Figure (b) displays the curves in VR. The *x*-axis represents the different volumes of cubes, while the *y*-axis represents the probability that the user chose the "stimulus".

Fig. 5). Therefore, the JND value is calculated using the following equation:

$$JND = \frac{P_{upper} + P_{lower}}{2} - Reference \quad (1)$$

Here, P_{upper} and P_{lower} represent the 75% and 25%, respectively. JND reflects precision, while PSE is a measure of bias [38]. Based on our experimental results, the JND values for AR and VR are JND_{AR} = 17.05 and JND_{VR} = 6, respectively.

All the related values are summarized together in Table 1.

4.2. ANOVA analysis

We first calculated PSE values for each participant in VR and AR conditions and performed one-way repeated measure ANOVA for two groups of PSE values. The results showed that there was a significant difference between the two conditions in terms of PSE values for our participant group ($F(1, 21) = 15.02$, $p = 0.001$, $\eta_p^2 = 0.417$). In particular, the participants' PSE values in the AR environment were significantly larger than those in the VR environment. Fig. 6 represents PSE for each participant in AR and VR.

In addition, we also analyzed the answers for all stimuli and computed the data groups that fit the psychophysical curve. We then performed a two-way repeated measures ANOVA for these data groups, considering the environments (i.e., AR vs VR) and presented stimuli (i.e., 850, 900, 950, etc.) as the independent factors. The results showed that the experimental environments had a statistically significant effect on the percentage of answers ($F(1, 21) = 6.216$, $p = 0.021$, $\eta_p^2 = 0.228$). In particular, post-hoc tests showed that the users' responses under the stimuli 1030 ($p = 0.046$), 1060 ($p = 0.026$), 1090 ($p = 0.016$), and 1150 ($p = 0.008$) in VR had a significant difference compared to AR (see Table 2).

Furthermore, we observed a significant effect for presented stimuli compared to the reference stimuli on the participants' answers ($F(9, 189) = 197.586$, $p < 0.0001$, $\eta_p^2 = 0.904$). As expected, as the size of the virtual cube increased, the number of trials where the presented stimuli were considered larger than the reference stimuli also increased (see Fig. 5). Specifically, pairwise comparison results showed that there was no significant difference among the following groups: 850 vs. 880, 970 vs. 1030, and 1120 vs. 1150 in AR, and 880 vs. 910, 910 vs. 940, 970 vs. 1030, and 1120 vs. 1150 in VR. However, all other pairwise comparisons showed significant effects (see Table 2). We can see that for both AR and VR scenes, the size changes in 970 vs. 1030 and 1120 vs. 1150 did not show significant differences. This means that judgments near the PSE and those that increase in ease of detection show similarity.

Table 2
The details of pairwise comparison among all groups of stimuli in AR and VR. The "$>$" represents the significant difference with $p < 0.05$, and the "$\sim$" represents no-significant difference.

Conditions	Stimuli & pairwise comparisons
AR	850 $\sim$ 880 > 910 > 940> 970 $\sim$ 1030 > 1060 > 1090 >1120 $\sim$ 1150
VR	850 > 880 $\sim$ 910 $\sim$ 940 > 970 $\sim$ 1030 > 1060 > 1090 > 1120 $\sim$1150

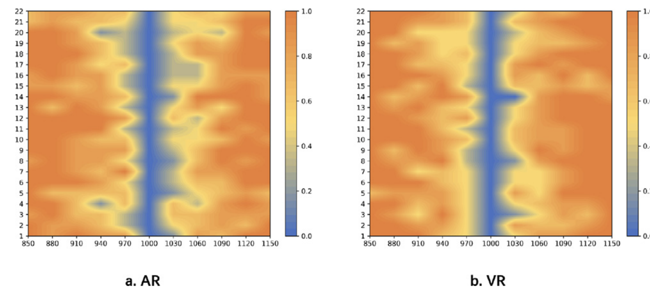

Fig. 7. Heat map shows the responses in size discrimination between AR and VR. The *x*-axis represents the volume of virtual cubes, while the *y*-axis shows the user ID. The color represents the distribution of the correct rate. We defined that when users choose a correct bigger volume, it is given greater weight. The orange color indicates that the user made all correct answers at that volume, while the blue color indicates that the user made all wrong answers.

4.3. Correct rates

To explore the effect of the direction of size changes (i.e., increase or decrease compared to the reference stimuli), we also calculated the correct rates of each stimulus in AR and VR conditions. Here, correct answer means that they chose the bigger one. Fig. 7 shows the heatmap of users' responses indicating the correct rates for AR and VR, respectively. We ran a two-way repeated measures ANOVA on the correct rates for AR and VR environments separately, considering the direction of size change and presented stimuli as the independent factors. The results showed that there was a significant effect for the direction of size change on the percentage of users' answers for AR ($F(1, 21) = 13.599$, $p = 0.01$, $\eta_p^2 = 0.393$). Post-hoc pairwise comparison showed that the correct rate of decreasing the presented stimuli of size change (M = 0.792; SD = 0.018) was higher than that of increasing the presented stimuli of size change (M = 0.729; SD = 0.014) in AR. However, we did not find any significant difference in the users' correct rates in VR in terms of the direction of size change ($F(1, 21) = 0.067$, $p = 0.799$, $\eta_p^2 = 0.003$), as shown in Fig. 8. In other words, in AR, people find it easier to distinguish the decrease rather than increase.

4.4. Questionnaire results

We assessed simulator sickness using the Simulator Sickness Questionnaire (SSQ) score and collected data in this section. Participants were asked to fill out the SSQ form at the beginning

Table 3
Means, standard deviations, percentage of participants with symptoms for all SSQ indicators.

	AR-pre			AR-post			VR-pre			VR-post		
	Mean	SD	>0%	Mean	SD	>0%	Mean	SD	>0%	Mean	SD	>0%
General discomfort	0.09	0.29	9.09	0.05	0.21	4.55	0.05	0.21	4.55	0.27	0.46	27.27
Fatigue	0.27	0.46	27.27	0.32	0.57	27.27	0.18	0.39	18.18	0.27	0.46	27.27
Headache	0.09	0.29	9.09	0.09	0.29	9.09	0.09	0.29	9.09	0.18	0.50	13.64
Eyestrain	0.45	0.51	45.45	0.5	0.60	45.45	0.45	0.60	40.91	0.5	0.67	40.91
Difficulty focusing	0.045	0.21	4.55	0.14	0.35	13.64	0.05	0.21	4.55	0.09	0.29	9.09
Sweating	0.23	0.53	22.73	0.09	0.29	9.09	0.09	0.29	9.09	0.05	0.21	4.55
Nausea	0.05	0.21	4.55	0.09	0.29	9.09	0.05	0.22	4.76	0.09	0.29	9.09
Difficulty concentrating	0.09	0.29	9.09	0.14	0.35	13.64	0.09	0.29	9.09	0.09	0.29	9.09
Blurred vision	0.09	0.29	9.09	0.23	0.53	18.18	0.14	0.35	13.64	0.18	0.50	13.64
Dizziness (eyes open)	0.09	0.29	9.09	0.14	0.35	13.64	0.05	0.21	4.55	0.09	0.29	9.09
Dizziness (eyes closed)	0.05	0.21	4.55	0.05	0.21	4.55	0.05	0.21	4.55	0.09	0.29	9.09
Vertigo	0.05	0.21	4.55	0.09	0.29	9.09	0.05	0.21	4.55	0.05	0.21	4.55

Table 4
Mean computation of SSQ score.

	Nausea(N)	Oculomotor(O)	Disorientation(D)
AR-pre	4.34	8.61	5.06
AR-post	3.9	11.03	10.12
VR-pre	3.04	7.92	5.06
VR-post	4.77	12.06	8.23

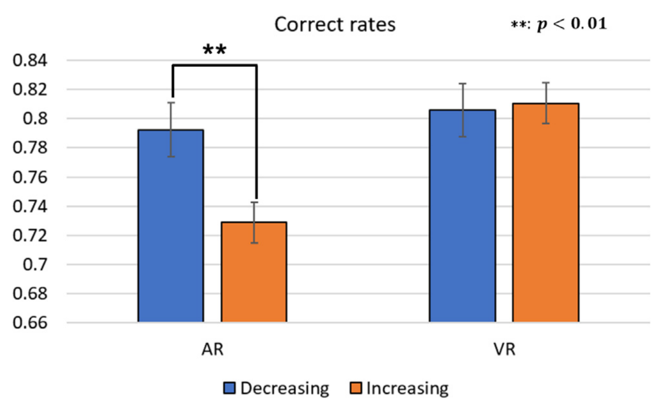

Fig. 8. The correct rates of increasing and decreasing stimuli in AR and VR conditions.

and end of each condition, defined as the pre-SSQ and post-SSQ, respectively. We calculated the Nausea (N), Oculomotor (O), and Disorientation (D) factors for all 16 symptoms using the weighted values from Kennedy et al.'s work [39], as shown in Table 4. Oculomotor symptoms contributed the most to simulator sickness, followed by disorientation. We also conducted t-tests to compare each symptom and found that only "General discomfort" in VR showed a significant difference between pre and post ($p = 0.04$).

We used a 4-point scale, as described by Vovk et al. [40], to evaluate the 16 symptoms. The scale is shown in Table 3, with eyestrain being the most commonly reported symptom. In AR, we excluded participants who reported slight eyestrain in the pre-experiment questionnaire, and found that 3 cases reported an increase in eyestrain symptoms. In AR-post, eyestrain was reported in 9 cases as slight and in 1 case as moderate. No severe symptoms were reported in either AR or VR. Fatigue was the second most commonly reported symptom, with 3 participants reporting increased fatigue in both AR and VR. In AR-post, fatigue was reported in 5 cases as slight and in 1 case as moderate. In VR-post, slight fatigue was reported in 5 cases. No severe symptoms of fatigue were reported.

5. Discussion

Our user-perception experiment found that participants had different size perception thresholds for the virtual cube in the $10 \times 10 \times 10$ cm^3 discrimination between VR and AR. The PSE values for size discrimination were 1010.9 (1.1%) cm^3 in the VR condition and 1022.6 (2.3%) cm^3 in the AR condition, indicating an average perceptual offset of 1.16%. Due to the smaller PSE values, the virtual object was significantly more often perceived as larger in the VR condition than in the AR condition. Participants were more likely to report the perceived size of the virtual object as larger than that of the reference object, with a probability of over 50% in the VR condition when presented with two virtual cubes of equal size in AR and VR environments. Overall, Our results suggest a psychological effect where the size differences of virtual cubes in the VR environment are easier to perceive than in the AR environment under our experimental settings. This finding refutes our Hypothesis 1. The JND values for size discrimination in our experiment were approximately 17.05 cm^3 (1.7%) for AR and 6 cm^3 (0.6%) for VR. These values are smaller than those reported in previous work on size perception in VR [12]. However, since we not only changed the heights but also fully altered the size of the virtual cubes, it is not surprising that the detection thresholds were different. Our results demonstrate that humans are highly sensitive to size changes, and even tiny changes can be noticed explicitly. These findings provide design guidance for AR/VR modeling, suggesting that the noticeable size change of virtual objects in VR/AR should be larger than 1081.6 cm^3 in AR and 1062.2 cm^3 in VR, respectively, where participants yielded a 75% probability of virtual size discrimination performance.

Furthermore, given the different visual perceptual thresholds of size discrimination in VR and AR, designers cannot simply duplicate 3D models from one condition to another for reuse. Specifically, for virtual objects with sizes ranging from 949.8 cm^3 to 1081.6 cm^3, designers should increase the size change by over 1.7% compared to the reference virtual object to enable users to perceive a significant size change in both VR and AR conditions. This finding has important implications for designing virtual objects with different sizes, such as virtual buttons or architectural models, where ensuring a consistent and accurate perception of size is crucial for user experience. Therefore, designers should consider the perceptual differences between VR and AR when creating 3D models and ensure that they are appropriately adjusted for each condition to achieve optimal user experience.

Finally, our experimental analysis revealed a significant difference in participants' responses to the correct rates in AR based on the direction of size change (i.e., increasing or decreasing size), with a higher correct rate for size decreasing than for size increasing. This finding refutes our Hypothesis 2. Hoba et al. [41] found that the brain has a preference for large objects, and larger objects activate early visual and ventral visual areas, as measured by scanning parameters using a 3 T TRIO MRI system. However, it is still unclear whether exists different areas of the brain are activated when comparing AR and VR through fMRI. Understanding this mechanism could help people from a basic level, and it has important implications for the design of AR/VR applications that involve size manipulation, such as virtual product customization or architectural design.

6. Limitations and future work

In terms of visual perception, Jones et al.'s study on depth perception in VR and AR [13] reveals that depth perception in VR is underestimated due to the compressed virtual space. However, our study yielded contrasting results. We found that the perceptual threshold in VR was smaller than that in AR (1010.0 vs. 1022.6), suggesting that the detection of size changes in virtual cubes was easier in VR compared to AR in our experimental setup. This may be due to the fact that depth perception is largely dependent on backgrounds [42], while size perception is influenced by the consistency of objects and backgrounds [43]. In our experimental settings, the resolution of the background in VR was significantly better than that in AR (1440p vs. 480p). However, it is worth noting that the resolution of virtual objects in both VR and AR scenes was consistent at 1440p. This discrepancy in background resolution may have made it easier to detect size changes in virtual objects. We acknowledge that this discrepancy in background resolution is a limitation of our study. As discussed in Section 3.4.2, although we ensured consistent rendering quality for virtual cubes in both AR and VR modes, there remains a difference in background resolution.

The reason for the relatively lower resolution of the background in AR is that we chose to use the same HMD for both AR and VR conditions. Such an approach has the drawback of poor display of the real environment with the pass-through mode, even if virtual contents are in the same resolution. Whether this difference in resolution between the physical surroundings and virtual content can affect our results remains uncertain. We asked the users whether an unclear background would affect their judgment of virtual objects during the AR condition. They responded that they were more focused on the virtual objects rather than the background. Therefore, we believe that the low-resolution influence does not have a big impact on our experiments, but its impact is still worth investigating. Furthermore, even if we fix the distance between the external cameras and the human eyes, there is still an offset between them as a camera that has extrinsic and intrinsic properties. This geometrical distortion of the perceived space problem is present in all video-see-through AR-HMDs and may impact distance and size perception. Another limitation is that the movement of objects can lead to changes in visual angles. The implications of these visual angle changes on size perception remain an open question.

Additionally, we acknowledge that existing devices, such as the Varjo XR-3 headset, Zed mini camera with HTC Vive, and Apple Vision Pro, can also switch between AR and VR modes. As our study was conducted with only one device, we cannot determine whether the perceptual statements we are making carry over to other devices and whether there would be changes to the magnitude of the values. We also observed that some participants reported experiencing eyestrain and fatigue during the experiment. These symptoms are consistent with the findings of [40]. Our experiment aimed to detect the visual perceptual threshold by discriminating small differences between provided stimuli and reference. This can be a mentally exhausting task, especially when the provided stimulus falls within the interval of uncertainty (IU). Additionally, we repeated each stimulus six times, following the methodology of Steinicke et al. [15], to ensure the credibility of our data. However, this repetition may have added mental demands and frustrations for participants during the experiment and could have led to over-skilling. Therefore, we aim to identify more efficient psychophysical methods to detect the threshold for human perception while minimizing these negative effects. Future work should explore and compare different psychophysical methods to determine which ones are most valuable.

Finally, we plan to extend our study to make it more rigorous by considering other related factors that may influence size discrimination, such as scene background. Moreover, it would be worthwhile to investigate whether our findings still hold if we express changes in linear dimensions instead of volume. Additionally, we are interested in exploring cross-modal effects between different perceptual channels, such as touch [44,45] and taste [46]. Understanding these effects could have important implications for designing HMDs that can seamlessly adapt content between AR and VR modes. Therefore, future research should build on our findings to further investigate the perceptual differences between AR and VR modes and explore how they can be applied to the design of HMDs that provide a seamless and optimal user experience.

7. Conclusion and future work

This paper presents psychophysical experiments on the perceptual thresholds of size discrimination in AR and VR. We explore the differences in human perceptual sensitivity between AR and VR and further calculate the finer perceptual thresholds of size changes in our experiments. We implemented our experiments based on an HTC Vive Pro HMD to explore the impact of virtual cubes between virtual and physical environments. The experimental results and analysis show that users are less sensitive to size changes in AR than in VR. In other words, people perceive the same content in VR as larger than in AR. Additionally, for size changes in AR, users are more sensitive to decreases rather than increases. Our experimental results on size sensitivity can be a valuable reference for 3D designers when designing virtual content in AR and VR. Our experimental protocol can also be used to study the detection thresholds of other senses in augmented reality and virtual reality.

CRediT authorship contribution statement

Liwen Wang: Conceptualization, Investigation, Methodology, Development, Validation, Data curation, Writing – original draft, Revised draft. **Shaoyu Cai:** Conceptualization, Investigation, Methodology, Data curation, Writing – review & editing. **Christian Sandor:** Supervision, Writing – review.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] Dionisio JDN, III WGB, Gilbert R. 3D virtual worlds and the metaverse: Current status and future possibilities. ACM Comput Surv 2013;45(3):1–38.
[2] Blascovich J, Loomis J, Beall AC, Swinth KR, Hoyt CL, Bailenson JN. Immersive virtual environment technology as a methodological tool for social psychology. Psychol Inq 2002;13(2):103–24.

[3] Yuen SC-Y, Yaoyuneyong G, Johnson E. Augmented reality: An overview and five directions for AR in education. J Educ Technol Dev Exchange (JETDE) 2011;4(1):11.
[4] Ke P, Keng K-N, Jiang S, Cai S, Rong Z, Zhu K. Embodied weather: Promoting public understanding of extreme weather through immersive multi-sensory virtual reality. In: The 17th international conference on virtual-reality continuum and its applications in industry. 2019, p. 1–2.
[5] Miller CH. Digital storytelling: A creator's guide to interactive entertainment. CRC Press; 2019.
[6] Cai S, Ke P, Narumi T, Zhu K. Thermairglove: A pneumatic glove for thermal perception and material identification in virtual reality. In: 2020 IEEE conference on virtual reality and 3D user interfaces. IEEE; 2020, p. 248–57.
[7] Garro V, Sundstedt V, Navarro D. A review of current trends on visual perception studies in virtual and augmented reality. In: SIGGRAPH Asia 2020 courses. 2020, p. 1–97.
[8] Moere AV, Purchase H. On the role of design in information visualization. Inf Vis 2011;10(4):356–71.
[9] Narumi T, Nishizaka S, Kajinami T, Tanikawa T, Hirose M. Augmented reality flavors: gustatory display based on edible marker and cross-modal interaction. In: Proceedings of the SIGCHI conference on human factors in computing systems. 2011, p. 93–102.
[10] El Jamiy F, Marsh R. Survey on depth perception in head mounted displays: Distance estimation in virtual reality, augmented reality, and mixed reality. IET Image Process 2019;13(5):707–12.
[11] Stefanucci JK, Creem-Regehr SH, Thompson WB, Lessard DA, Geuss MN. Evaluating the accuracy of size perception on screen-based displays: Displayed objects appear smaller than real objects. J Exp Psychol Appl 2015;21(3):215.
[12] Thomas BH. Examining user perception of the size of multiple objects in virtual reality. Appl Sci 2020;10(11):4049.
[13] Jones JA, Swan JE, Singh G, Kolstad E, Ellis SR. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In: Proceedings of the 5th symposium on applied perception in graphics and visualization. 2008, p. 9–14.
[14] Gaffary Y, Le Gouis B, Marchal M, Argelaguet F, Arnaldi B, Lécuyer A. AR feels "softer" than VR: Haptic perception of stiffness in augmented versus virtual reality. IEEE Trans Vis Comput Graph 2017;23(11):2372–7.
[15] Steinicke F, Bruder G, Kuhl S. Realistic perspective projections for virtual objects and environments. ACM Trans Graph 2011;30(5):1–10.
[16] Azuma RT. A survey of augmented reality. Presence: Teleoperat Virt Environ 1997;6(4):355–85.
[17] Dani TH, Gadh R. Creation of concept shape designs via a virtual reality interface. Comput Aided Des 1997;29(8):555–63.
[18] Baird JC. Psychophysical analysis of visual space: International series of monographs in experimental psychology. Vol. 9. Elsevier; 2013.
[19] Ekman G, Junge K. Psychophysical relations in visual perception of length, area and volume. Scandinavian J Psychol 1961;2(1):1–10.
[20] Stanek RJ. A parametric study of volume and surface judgments. Percept Psychophys 1969;6:16–8.
[21] Jansen Y, Hornbæk K. A psychophysical investigation of size as a physical variable. IEEE Trans Vis Comput Graph 2015;22(1):479–88.
[22] Teghtsoonian M. The judgment of size. Am J Psychol 1965;78(3):392–402.
[23] Carlson V. Overestimation in size-constancy judgments. Am J Psychol 1960;73(2):199–213.
[24] Gilinsky AS. The effect of attitude upon the perception of size. Am J Psychol 1955;68(2):173–92.
[25] Murray SO, Boyaci H, Kersten D. The representation of perceived angular size in human primary visual cortex. Nature Neurosci 2006;9(3):429–34.
[26] Ahn J-g, Ahn E, Min S, Choi H, Kim H, Kim GJ. Size perception of augmented objects by different AR displays. In: International conference on human-computer interaction. Springer; 2019, p. 337–44.
[27] Zhou Q, Wu F, Fels S, Stavness I. Closer object looks smaller: Investigating the duality of size perception in a spherical fish tank VR display. In: Proceedings of the 2020 CHI conference on human factors in computing systems. 2020, p. 1–9.
[28] Kim J, Interrante V. Dwarf or giant: The influence of interpupillary distance and eye height on size perception in virtual environments. In: 27th International conference on artificial reality and telexistence, ICAT 2017 and the 22nd Eurographics symposium on virtual environments. Eurographics Association; 2017, p. 153–60.
[29] Blackwell HR. Studies of psychophysical methods for measuring visual thresholds. JOSA 1952;42(9):606–16.
[30] Wang L, Sandor C. Can you perceive the size change? Discrimination thresholds for size changes in augmented reality. In: International conference on virtual reality and mixed reality. Springer; 2021, p. 25–36.
[31] Rolland JP, Meyer C, Arthur K, Rinalducci E. Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. Presence 2002;11(6):610–25.
[32] Ping J, Liu Y, Weng D. Comparison in depth perception between virtual reality and augmented reality systems. In: 2019 IEEE conference on virtual reality and 3d user interfaces. IEEE; 2019, p. 1124–5.
[33] Jang Y, Wixted JT, Huber DE. Testing signal-detection models of yes/no and two-alternative forced-choice recognition memory. J Exp Psychol [Gen] 2009;138(2):291.
[34] Sauer Y, Sipatchin A, Wahl S, García García M. Assessment of consumer VR-headsets' objective and subjective field of view (FoV) and its feasibility for visual field testing. Virt Real 2022;26(3):1089–101.
[35] Gil Rodríguez R, Bayer F, Toscani M, Guarnera D, Guarnera GC, Gegenfurtner KR. Colour calibration of a head mounted display for colour vision research using virtual reality. SN Comput Sci 2022;3:1–10.
[36] Linares D, López-Moliner J. Quickpsy: An R package to fit psychometric functions for multiple groups. R J 2016;8(1):122–31.
[37] Macmillan NA, Creelman CD. Detection theory: A user's guide. Psychology Press; 2004.
[38] Zeni S, Holmes NP. The effect of a regular auditory context on perceived interval duration. Front Psychol 2018;9:1567.
[39] Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993;3(3):203–20.
[40] Vovk A, Wild F, Guest W, Kuula T. Simulator sickness in augmented reality training using the Microsoft HoloLens. In: Proceedings of the 2018 CHI conference on human factors in computing systems. 2018, p. 1–9.
[41] Hoba S, Fink GR, Zeng H, Weidner R. View normalization of object size in the right parietal cortex. Vision 2022;6(3):41.
[42] Grove PM, Gillam B, Ono H. Content and context of monocular regions determine perceived depth in random dot, unpaired background and phantom stereograms. Vis Res 2002;42(15):1859–70.
[43] Aks DJ, Enns JT. Visual search for size is influenced by a background texture gradient. J Exp Psychol Hum Percept Perform 1996;22(6):1467.
[44] Cai S, Zhu K, Ban Y, Narumi T. Visual-tactile cross-modal data generation using residue-fusion gan with feature-matching and perceptual losses. IEEE Robot Autom Lett 2021;6(4):7525–32.
[45] Cai S, Zhao L, Ban Y, Narumi T, Liu Y, Zhu K. GAN-based image-to-friction generation for tactile simulation of fabric material. Comput Graph 2022;102:460–73.
[46] Narumi T, Ban Y, Kajinami T, Tanikawa T, Hirose M. Augmented perception of satiety: controlling food consumption by changing apparent size of food with augmented reality. In: Proceedings of the SIGCHI conference on human factors in computing systems. 2012, p. 109–18.