Effects of Depth Information on Visual Target Identification Task Performance in Shared Gaze Environments

Austin Erickson, Nahal Norouzi, Kangsoo Kim, Joseph J. LaViola Jr., Gerd Bruder, and Gregory F. Welch

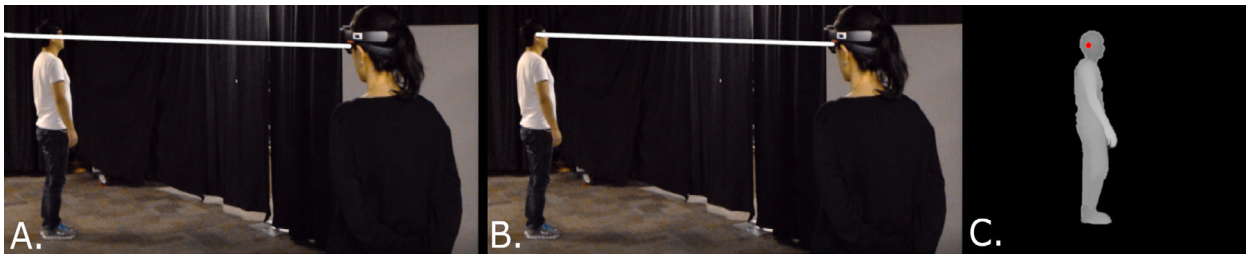


Fig. 1. An illustration depicting a shared gaze visualization in augmented reality without and with depth information: (A) in traditional eye tracked setups without real-time depth information, the user can tell the direction of their partner's gaze but it is ambiguous in terms of which object they are observing along the ray, while (B) in shared gaze setups with depth information, such ambiguities can be resolved by truncating the ray at the target distance, e.g., provided by (C) real-time depth sensor information, where the pixel queried in the depth map (the termination point of the ray) is highlighted in red.

Abstract—Human gaze awareness is important for social and collaborative interactions. Recent technological advances in augmented reality (AR) displays and sensors provide us with the means to extend collaborative spaces with real-time dynamic AR indicators of one's gaze, for example via three-dimensional cursors or rays emanating from a partner's head. However, such gaze cues are only as useful as the quality of the underlying gaze estimation and the accuracy of the display mechanism. Depending on the type of the visualization, and the characteristics of the errors, AR gaze cues could either enhance or interfere with collaborations. In this paper, we present two human-subject studies in which we investigate the influence of angular and depth errors, target distance, and the type of gaze visualization on participants' performance and subjective evaluation during a collaborative task with a virtual human partner, where participants identified targets within a dynamically walking crowd. First, our results show that there is a significant difference in performance for the two gaze visualizations *ray* and *cursor* in conditions with simulated angular and depth errors: the ray visualization provided significantly faster response times and fewer errors compared to the cursor visualization. Second, our results show that under optimal conditions, among four different gaze visualization methods, a ray without depth information provides the worst performance and is rated lowest, while a combination of a ray and cursor with depth information is rated highest. We discuss the subjective and objective performance thresholds and provide guidelines for practitioners in this field.

Index Terms—Shared Gaze, Augmented Reality, Depth Error, Gaze Visualization, Performance Measures

1 Introduction

Recent advances in augmented reality (AR) systems and component technologies, including tracking, processing power, and display quality, rekindled public interest in many long-envisioned interaction techniques and applications [44]. Among these, collaboration in AR was one of the most highly cited application areas for usability studies in AR from 2005 to 2014 [9] and has emerged as an active research topic in recent years [22]. Compared to conventional collaboration platforms, AR offers several benefits, including seamlessly blended real and virtual content, improved physical and spatial cues, and unconstrained view perspectives for both collocated and remote collaborators [2].

Human gaze awareness can improve performance of collaborative tasks [36]. It can be an indicator of one's focus of attention [24], and plays an important role during many collaborative efforts by facilitating coordination [6], disambiguation [17], and joint attention [30]. Researchers have investigated the use of shared gaze information in collaborative systems for different applications such as search tasks, assembly tasks, and teleconferencing [4, 13, 25, 37, 38], utilizing various visualization paradigms such as heat maps to indicate dynamic individual and shared attention, and virtual frusta, cursors, and rays[1] to indicate head and eye gaze information [7, 35].

Such visualizations inherently rely on real-time estimates of head pose, eye gaze, and scene information, typically obtained from head trackers, eye trackers, and depth cameras, respectively. For example, to represent a user's gaze ray towards an object in the scene, it is necessary to know the user's head pose and gaze direction. For a less ambiguous representation, it is further possible to leverage real-time depth sensors such that the gaze ray terminates at the intended target. While the tracking, depth, and display technologies necessary for such visualizations have improved considerably over time, they will never be free from error. Depending on the users, their tasks and environments, and the choices of visualization paradigms, the errors will affect the quality of the collaborations and their outcomes.

In this paper, we address the following research questions:

- **RQ1:** What is the impact of angular and depth gaze errors on users' task performance and evaluation of the collaboration?
- **RQ2:** What are the objective and subjective thresholds for the magnitude of errors up to which user performance is not affected?

- *The first two authors contributed equally to this work.*
- *Austin Erickson, Nahal Norouzi, Kangsoo Kim, Joseph J. LaViola Jr., Gerd Bruder, and Gregory F. Welch are with the University of Central Florida. E-mail: {ericksona, nahal.norouzi}@knights.ucf.edu, kangsoo.kim@ucf.edu, jjl@cs.ucf.edu, {bruder, welch}@ucf.edu.*

Manuscript received 10 Sept. 2019; accepted 5 Feb. 2020. Date of publication 18 Feb. 2020; date of current version 27 Mar. 2020. Digital Object Identifier no. 10.1109/TVCG.2020.2973054

[1] In keeping with prior literature we use the term "ray" to broadly represent both rays—3D lines with "infinite" extent that are defined by their starting point and orientation, and ray segments—3D rays that terminate at a point in space.

- **RQ3:** Under optimal conditions, which gaze visualization method results in the highest task performance and is preferred by users?

To answer these research questions, we designed a shared gaze setup and collaborative task where participants were asked to identify a target among a dynamic crowd using the gaze information of a (simulated) human partner. Based on this setup, we conducted two human-subject studies where we varied the amount and type of simulated errors introduced to the shared gaze, as well as the type of gaze visualization, and target distances. We measured each participant's task performance through their response time and error rate, and collected subjective feedback about their performance and perception of the collaborative shared gaze interaction. We present objective and subjective error thresholds to support practitioners in making informed decisions about the use of adequate sensor hardware and visualization methods.

The remainder of this paper is structured as follows. Section 2 presents an overview of related work. Section 3 presents our first experiment, focusing on an evaluation of two gaze visualization methods, two target distances, and two error types with seven levels each. Section 4 describes our second experiment, comparing four gaze visualization methods under optimal conditions for two target distances. Section 5 presents a general discussion. Section 6 concludes the paper.

2 Related Work

In this section, we describe previous work that studied gaze as an enhancement cue in collaborative AR or virtual reality (VR) setups, and discuss error sources, including eye trackers and depth cameras, that can influence gaze localization. We also review some of the previously studied approaches for visualizing shared gaze.

2.1 Shared Gaze in AR and VR

In a recent survey paper by de Belen et al., provision of non-verbal cues such as gaze and pointing cues was identified as one of the important factors in enhancing collaboration quality in mixed reality spaces [8]. Kiyokawa et al. introduced the idea of "enhanced awareness" to improve the quality of collaboration by visualizing a user's face direction using a ray, as users collaborating in the real world can often perceive where the other person is looking [23]. Billinghurst and Kato developed a mixed reality web browser that conveyed awareness of a user's gaze by highlighting the web page the user looked at, and assigning a gaze icon to each page that was viewed by the other user [1]. Grasset et al. utilized a ray to communicate one user's gaze direction, finding it a sufficient awareness signal in a maze navigation task [15].

In a helper/worker scenario for a puzzle assembly task, Gupta et al. investigated the impact of sharing a worker's gaze and a helper's pointing cues, finding improved performance and communication quality when the cues were available [16]. Masai et al. developed empathy glasses, sharing cues such as gaze and facial expression from the worker to the helper and augmenting the helper's pointing cues in the worker's field of view for a 2D puzzle assembly task [29]. Their findings from a pilot study suggest the benefit of shared gaze in establishing accurate spatial referencing and a shared understanding. Piumsomboon et al. introduced the CoVAR system for mixed space collaboration with the capability of visualizing different awareness cues such as head ray, gaze ray and hand gestures [35]. In a search and placement task, they found that the inclusion of awareness cues positively affected users' performance. Piumsomboon et al. further developed Mini-Me, an avatar facilitating mixed space remote collaboration where different awareness cues such as gaze and hand pointing information where shared through cursors and rays between users and supported their system's capability in conveying non-verbal cues for an object placement task [34].

Although past work has shown the benefits of shared cues such as gaze in the quality of users' collaborative experiences, there has been little work examining the effects of different types and magnitudes of (inevitable) gaze errors on the AR/VR tasks. Knowledge about the effects could help developers make design choices that result in more reliable gaze perception.

2.2 Gaze and Depth Accuracy

Researchers have identified multiple factors that could result in eye tracking error, and proposed solutions to mitigate their influence. For example, highly dynamic tasks might result in the movement of the camera positions, environmental lighting conditions can change, users have inherent physiological differences, and calibration and mapping results can be sub-optimal [3, 19, 20, 32]. Feit et al. investigated the influence of error sources such as lighting and target distance on eye tracking accuracy and proposed approaches to optimize parameter choices for factors such as outlier detection and target size [11]. Holmqvist et al. emphasized on the importance of eye tracking accuracy in relation to the task at hand, e.g., 0.5 degrees of error can be considered either poor or ideal quality depending on the target size [19]. Norouzi et al. investigated the effects of simulated eye tracking errors affecting the accuracy and precision of the gaze data, finding that accuracy offsets can significantly hamper user performance in a collaborative AR environment compared to low precision in a dynamic search task [31].

In previous papers, eye tracking errors were limited to the two dimensions corresponding to gaze direction, but correctly estimating gaze depth (distance to target) becomes all the more important for 3D interactions such as in AR and VR. Wang et al. proposed a computational approach, calculating the 3D gaze position from a monocular eye tracker with high depth accuracy rates of less than 2% for a 3D object at roughly 55 cm distance [42]. To estimate gaze depth for eye trackers used in AR and VR, Lee et al. implemented a multi-layer perceptron neural network with gaze normal vectors as input and evaluated its performance for distances of 1m to 5 m, finding an average depth error of 0.42 m to be acceptable for applications that do not require high precision [26]. Elmadjian et al. proposed geometric and regression-based approaches for 3D gaze estimation with the goal of facilitating 3D gaze interaction, reaching average depth errors of 0.53 m and 0.19 m respectively for the two approaches for distances of 0.75 m to 2.75 m [10]. Weier et al. created a feature set of various gaze-based measures including eye vergence to train a machine learning model for gaze depth estimation reaching average errors of 0.5 m for targets at a 6 m distance [43]. Mardanbegi et al. proposed an approach based on the vestibulo-ocular reflex for estimating a target's depth in 3D environments and compared their approach to a method relying on vergence [28]. Their proposed approach performed better than the vergence-based approach by 18% in a wide-range scene covering the three target distances of 0.5 m, 1.5 m, and 7 m.

Researchers also evaluated the depth estimates of devices such as the Microsoft Kinect and Microsoft HoloLens. Khoshelham et al. evaluated the accuracy and resolution of depth data from Kinect v1 and described elements such as lighting conditions and object surface properties as possible sources of error impacting the quality of the collected point cloud [21]. Yang et al. evaluated the depth accuracy and resolution for the Kinect v2 with a planar surface at different angles and distances from the Kinect [45]. For their setup, they found errors of more than 4 mm at a 4 m distance. Looking at hologram stability for the Microsoft HoloLens, Vassallo et al. identified different activities that can negatively affect the tracking performance of the device such as a user walking away from the hologram and coming back, motion with sudden acceleration, occlusion of the RGB-D cameras, and insertion of an external object into the hologram [40]. After multiple trials for each type of interference, they found a mean hologram displacement of 5.83 mm for their static hologram setup. Liu et al. designed experimental procedures to evaluate the HoloLens' performance in terms of aspects such as head localisation, spatial mapping, and environmental reconstruction [27]. Their findings suggest an average accuracy deviation of 73.8% for the spatial mapping task by calculating the distance between the placement of a hologram and its target location at distances of 0.5 m to 3.5 m.

2.3 Shared Gaze Visualization

In the field of human-computer interaction, mostly for 2D displays (e.g., computer screens) researchers have studied the effects of different gaze visualization types in shared gaze environments. In the context of driving, Trösterer et al. discussed how the gaze information of the

passenger can be helpful to the driver in communicating upcoming hazards and studied two types of gaze visualizations where in one the target is overlaid with a dot, while the other approach relies on a horizontal strip of LEDs where the corresponding LED would light up depending on the position of the passenger's gaze [39]. They found that the LED approach did not communicate a sufficient amount of information to the driver during a navigation task. To facilitate a visual search task between remote collaborators, D'Angelo and Gergle experimented with three types of gaze visualizations, (a) a *heatmap* approach that encapsulates where each person looked at in a range of time, (b) *shared area* where a circle would appear when collaborators looked at the same area at the same time, and (c) a *path* approach where a trail of gaze information would appear from the past three seconds [7]. Their findings suggest that gaze visualization affects certain factors such as task performance, and the amount of time required for one person to help their partner. Zhang et al. studied the influence of the shared gaze visualization type on the performance of collocated collaborators [46]. Their four approaches were (a) cursor, (b) trajectory, (c) highlight, and (d) spotlight. In their study, participants found the highlight and spotlight approaches to be less distracting.

Shared gaze environments in the AR/VR domain mostly rely on either a ray or a 3D cursor visualization to communicate a user's gaze point or the direction of their gaze (Section 2.1), and to our knowledge the influence of the type of gaze visualization has not been studied. Piumsomboon et al. is one of the few examples were two different visualization techniques of a ray and a frustum were introduced in their CoVAR mixed space collaborative platform [35]. Further investigation is warranted to understand the influence of errors in different shared gaze visualizations for collaborative AR environments.

3 Experiment I

We conducted two human-subjects studies, which were performed sequentially back to back with the same participants. In this section, we describe the first study assessing the influence of error type, error level, target distance, and gaze visualization in a shared gaze AR environment.

3.1 Participants

23 participants (5 female, 18 male, age $M = 24.60$, $SD = 8.74$) were recruited for this study. All the participants were students or employees of our university, and had normal or corrected vision and normal hearing. Using a 7-point Likert scale with 1 = Not familiar/Novice and 7 = Very Familiar/Expert, we assessed the participants' familiarity with AR ($M = 5.04$, $SD = 1.77$), VR ($M = 5.65$, $SD = 1.52$), virtual humans ($M = 4.43$, $SD = 2.17$), and their computer expertise ($M = 6.04$, $SD = 0.92$). The institutional review board of our university had approved the protocol for our human-subject study.

3.2 Material

The study scenario involved a crowd of simulated humans (virtual humans, VHs), which participants could see through a Microsoft HoloLens—an optical see-through AR head-mounted display (HMD)—in a shared AR space. Additionally, a simulated virtual *partner* was positioned adjacent to the area where the participants stood. Participants were tasked with identifying the VH target that the virtual partner was looking at among a dynamic crowd of walking VHs in front of them by calling out the number that was floating above the targets head (Figure 2). The virtual scene was implemented using the Unity game engine (version 2018.2.21). We conducted the experiment in an open 4.6 m × 10.4 m space in our laboratory. For ease of control over the study conditions, a client-server networked approach was implemented so that the client program on a HoloLens, which the participants wore, was connected to the server program on a laptop PC that the experimenter used to control the sequences of the study. This setup also allowed the experimenter to see exactly what the participants were seeing on the HoloLens. For the server laptop we used an Intel Core i7-7820HK CPU @ 2.9 GHz, 16Gb RAM, NVIDIA GeForce 1070, running Windows 10 Pro.

Fig. 2. Illustration of the four visualization methods: (A) truncated gaze ray that halts on an intersection with an object, (B) gaze ray that passes through all objects in the scene, (C) cursor (e.g., used by the HoloLens) that appears at the point of intersection between the gaze vector and object in the scene, and (D) combination of ray and cursor visualizations.

3.2.1 Shared Gaze Information

To ensure that each participant experienced the same shared gaze stimuli in the same experimental conditions, we utilized a virtual partner with simulated gaze information instead of a real human. Figure 3 shows the virtual partner, displayed as a 3D virtual character next to a participant. The virtual partner was visible to participants through the HoloLens and was programmed to stand facing forward a meter away to the left of the participants. The virtual partner's gaze information was calculated

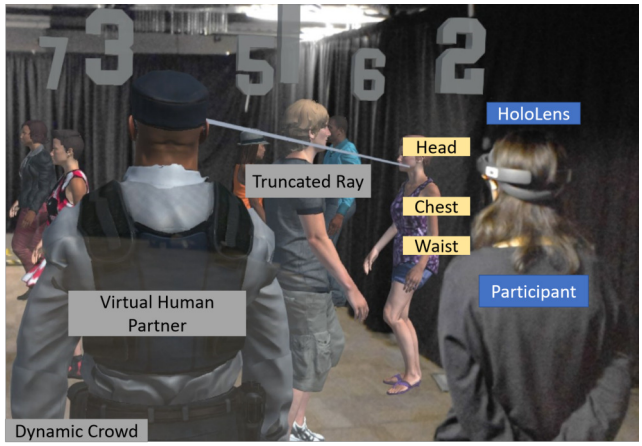


Fig. 3. Screenshot showing the participant next to the virtual human partner during the target identification task.

based on pre-recorded gaze data and visualized either in the form of a truncated gaze ray or a gaze cursor, which are common forms of gaze visualization used in AR/VR environments (see Figure 2 A and C).

We chose to use pre-recorded gaze data as our method of simulating the virtual partner's gaze after realizing that several heuristic-based approaches resulted in unrealistic gaze behavior in AR. To record the gaze data, a lab member was seated at a distance of one meter away from a stationary target point being displayed on a computer monitor, and fixated at the point for 30 seconds. Upon finishing the gaze recording we found that the angular accuracy error of the recorded data was 0.55° and the precision error was 0.08°. This data was then analyzed to find the average gaze position observed (with the given accuracy error), and was then normalized around the found position to yield data with no accuracy error.

We played back the recorded gaze data from the perspective of the virtual partner in Unity on the HoloLens. The gaze data described the virtual partner's gaze towards three points on the target VH's body in

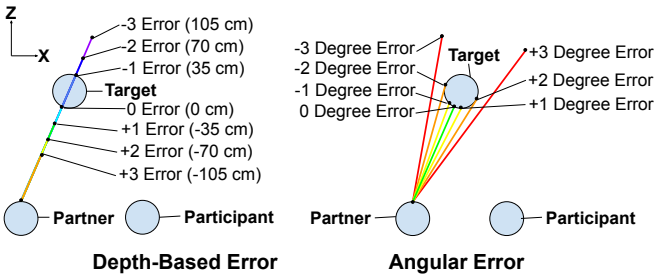


Fig. 4. Illustration depicting a top-down view of depth and angular errors in gaze data as they relate to the shared gaze scenario.

the dynamic virtual crowd: one on their head, one on their chest, and one near their waist (Figure 3). For the saccadic gaze movement, we implemented the gaze behavior randomly moving between the three points, every 750 ms, with a 50% chance of choosing the point on the head and equal chances of 25% of choosing either of the other two points on the target VH's body. We could also achieve smooth pursuit behavior by allowing the gaze to follow these points on the different VHs in the crowd.

3.2.2 Dynamic Virtual Crowd

We used eight VH characters (4 female, 4 male) from the Unity Asset Store and Mixamo for our dynamic crowd. The VHs were animated to pace between predefined points in our experimental space, covering a distance of six meters perpendicular to the participant's view direction. The VHs were 0.7 m apart from each other, with the closest one being 1.75 m away from the participant's position. 3D numbers (from 0 to 7) were placed on top of each VH's head and followed the VH's movement while always being correctly oriented towards the participant (Figure 3). The placements of the VHs along the forward direction and their walking speed were randomized for every trial in the study. The walking speed was randomly chosen from the range of 0.6 m/s to 1 m/s with increments of 0.05 m/s as it is close to average speed of walking humans [12].

3.2.3 Simulated Gaze Error

As discussed in Section 1, we considered two different types of errors that result from using a combination of depth sensors and eye trackers in a shared gaze AR scenario. Figure 4 illustrates each error type in comparison to a no-error example. Below, we describe each error, its possible source, and how it was implemented for our study. In an actual (non-simulated) shared gaze scenario, the gaze data is continuously streaming from the depth sensor and the eye trackers, which may occur at different frequencies, such that the final gaze visualization can be influenced by the interaction between those two types of errors. For our purposes, however, we assume that the target depth is calculated after receiving a gaze direction that includes any amount of error introduced by the eye tracker. In this manner, when a gaze point 'falls off' the VH target due to a high degree of angular error, the depth value is read from a pixel outside of the silhouette of that VH in the depth map (compare Figure 1 and angular error in Figure 4).

Angular Accuracy: We simulated typical angular errors in eye tracker performance via an angular offset that we added to the correct gaze direction. To implement this error, a gaze vector between the virtual partner and the target VH was calculated without any accuracy error, while applying the recorded gaze data. Following this calculation, the vector was rotated in horizontal direction either leftward or rightward of the target by a variable number between 0° and 3° at 1° intervals to achieve an angular accuracy offset along the horizontal axis. The gaze visualization (truncated ray or cursor) was then rendered using the direction of this rotated vector.

While physical eye tracking systems introduce errors in both the horizontal and vertical directions, here we are introducing them solely to the horizontal direction. For the purposes of our study scenario, we decided to pursue the most significant types of accuracy error that could occur for the given scenario consisting of horizontally-pacing VHs. In our scenario, vertical accuracy error is less significant than horizontal accuracy error in that vertical error is often still calculated to be drawn on target (albeit at a different point on the target's body). Such vertical errors lead to targets that are easier for users to identify than if the error occurs in the horizontal direction.

Depth Accuracy: We simulated depth errors as an offset that we added to the actual depth of the target that the gaze vector intersected with. Such errors commonly originate in inaccuracies in the depth maps provided by a depth sensor, e.g., in the HoloLens, but can also originate in an eye tracker's estimated focus distance (based on eye convergence). This error is implemented by using the direction of the virtual partner's gaze and a depth map produced by a secondary camera in the Unity environment at the virtual partner's point of view to calculate the depth of the gaze point along the gaze vector. If this vector intersects any VH in the scene, then the endpoint is calculated at the depth of the first object along this vector. Alternatively, if the vector does not intersect with any VH in the scene, then we chose to set it to a fixed distance and upper limit of ten meters, representing the back wall in our laboratory. The endpoints of the gaze visualizations (ray or cursor) were then at one of seven levels relative to the correct depth, symmetrically, at each three positive and negative steps of 35 cm. For the positive levels, the endpoint had a depth error that positioned the endpoint between the participant and the target VH at increasing distances from the target for higher error levels. At zero, no error was present in the depth direction and the endpoint appeared at the collision point with the target VH. At negative levels, the endpoint had a depth error that positioned the endpoint behind the target, at increasing distances for higher error levels.

When choosing the angular and depth error ranges described above, we looked at the literature summarized in Section 2 in addition to the nominal performance reported by manufacturers for commercial head-worn eye trackers, such as from Tobii and Pupil Labs, as well as commercial depth sensors, such as the StereoLabs ZED and the internal HoloLens tracking system. To make sure that participants were able to perceive differences in the ranges of angular and depth errors, we set our maximum errors to be above the ranges reported in the above-mentioned sources.

3.2.4 Shared Gaze Visualization

We implemented two types of shared gaze visualizations for the study. The first type is a *truncated ray* that is drawn from the virtual partner's head toward a target that is being observed, where its end point is always set to the depth of the first object it intersects with (Figure 2 A). In case of no such intersection, as mentioned above, we set an upper limit of ten meters on the ray length, which is far enough to be behind all other VHs in the scene and matches the distance between the participant and the back wall of the laboratory.

The second visualization type is a *cursor*, where the placement of the cursor is similarly set to the depth of the first object that the gaze vector intersects with, or is set to the same upper limit distance of ten meters when no such intersection occurs (Figure 2 C).

3.3 Methods

3.3.1 Experimental Design

We chose a $2 \times 7 \times 2 \times 2$ within-subjects design considering that individual differences can influence participants' task performance. Our independent variables are as follows:

- **Error Type** ($\times 2$), tested independently, which were either: (a) **x-error:** Angular eye tracking error in the horizontal (x) direction, and (b) **z-error:** Depth tracking error in the forward (z) direction.
- **Error Level** ($\times 7$), which were: (a) **x-error:** Seven levels in $\{-3°, -2°, -1°, 0°, 1°, 2°, 3°\}$, with zero indicating no error, positive numbers indicating error to the right, and negative numbers indicating error to the left, and (b) **z-error:** We defined a range of seven levels

in $\{-105\,\text{cm}, -70\,\text{cm}, -35\,\text{cm}, 0\,\text{cm}, 35\,\text{cm}, 70\,\text{cm}, 105\,\text{cm}\}$, with zero indicating no error, negative numbers indicating errors behind the target, and positive numbers indicating errors in front.

- **Target Depth** ($\times 2$), which were: **Close:** Target VH was set at 3.15 m from the virtual partner. **Far:** Target VH was set at 5.25 m from the virtual partner.

- **Gaze Visualization Type** ($\times 2$), which were: **Truncated Ray:** The end point of the ray was set to the target's depth and would be truncated at a new depth if intersected by an obstacle closer to the virtual partner. **Cursor:** The cursor was set to the target's depth and would translate to a new depth if the gaze vector was intersected by an obstacle closer to the virtual partner.

This study design results in a total of 56 trials per participant. The trials were divided into four blocks where within each block the error and visualization type remained constant and the error levels and target distance were varied resulting in 14 trials per block. The order with which participants were exposed to the four blocks and the 14 trials within them were randomized to account for learning effects.

3.3.2 Procedure

At the beginning, the participants provided their informed consent, and filled out a questionnaire about their familiarity with related technology. The experimenter then reviewed the procedure with the participants, guided them into position for the start of the experiment and instructed them on how to don the HoloLens. Participants took part in five practice trials to get familiarized with the system in which they were tasked with identifying which VH in the crowd was being observed by the VH partner. Participants were to identify this target by observing the partner's gaze within a time frame of up to 60 seconds to make a selection. Once they identified a target, they were instructed to verbally indicate the number that floated above its head, which was then recorded by the experimenter (Figure 3). The error blocks were presented in randomized order as described in Section 3.3.1. After the end of each block, participants were asked to remove the HMD and fill out questionnaires regarding their experience. Afterward, the participants moved on to Experiment II that is explained in Section 4.

3.3.3 Measures

In this section, we describe the objective and subjective measures used to assess participants' task performance and to collect their subjective evaluations with regards to the different error types and gaze visualizations.

Objective Measures: We used ***response time*** as the amount of time taken by each participant to identify the target, and ***accuracy rate*** as accuracy of participants' responses for each trial to assess participants' task performance throughout the study, where participants were asked to keep both speed and accuracy in mind as target identification factors for each trial.

Subjective Measures: We used the questionnaires described below to assess participants' perception of the error types, gaze visualizations, and subjective performance for each condition block.

- **Performance Evaluation (PE):** The questions described in Table 1 were used to assess participants' confidence in their responses, and their subjective threshold level for the type of error.

- **Task Load:** The short version of the NASA-TLX questionnaire [18] was used to assess the task load.

- **Usability:** The System Usability Scale (SUS) questionnaire [5] was used to assess the usability aspects of the shared gaze system.

Table 1. Questions assessing participants' performance evaluation.

Performance Evaluation Questions	
PE1	How confident were you on the correctness of your choices in this section of the experiment? (7-point Likert Scale)
PE2	What is the maximum amount of error that you think you could tolerate when using such a system? (numeric response)

3.3.4 Hypotheses

Based on pilot testings and the previous literature, we formulated the following hypotheses:

- **H1:** For both error types, an increase in error levels will result in an increased response time and decreased accuracy rate.

- **H2:** For higher error levels, the cursor visualization will increase participants' response time and decrease their accuracy rate when compared with the truncated ray visualization.

- **H3:** For blocks where the gaze is visualized as a truncated ray, participants will exhibit higher confidence in their responses, report a higher usability score, experience lower task load, and have a higher tolerance for error.

- **H4:** For blocks where simulated z-error is introduced, participants will exhibit higher confidence in their responses, report a higher usability score, and experience lower amounts of task load than the conditions where simulated x-error is introduced.

- **H5:** Participants' performance will decrease for targets at the far distance compared to targets at the close distance.

3.4 Results

In this section, we report our results for Experiment I. We excluded two of our participants from the analysis due to technical issues.

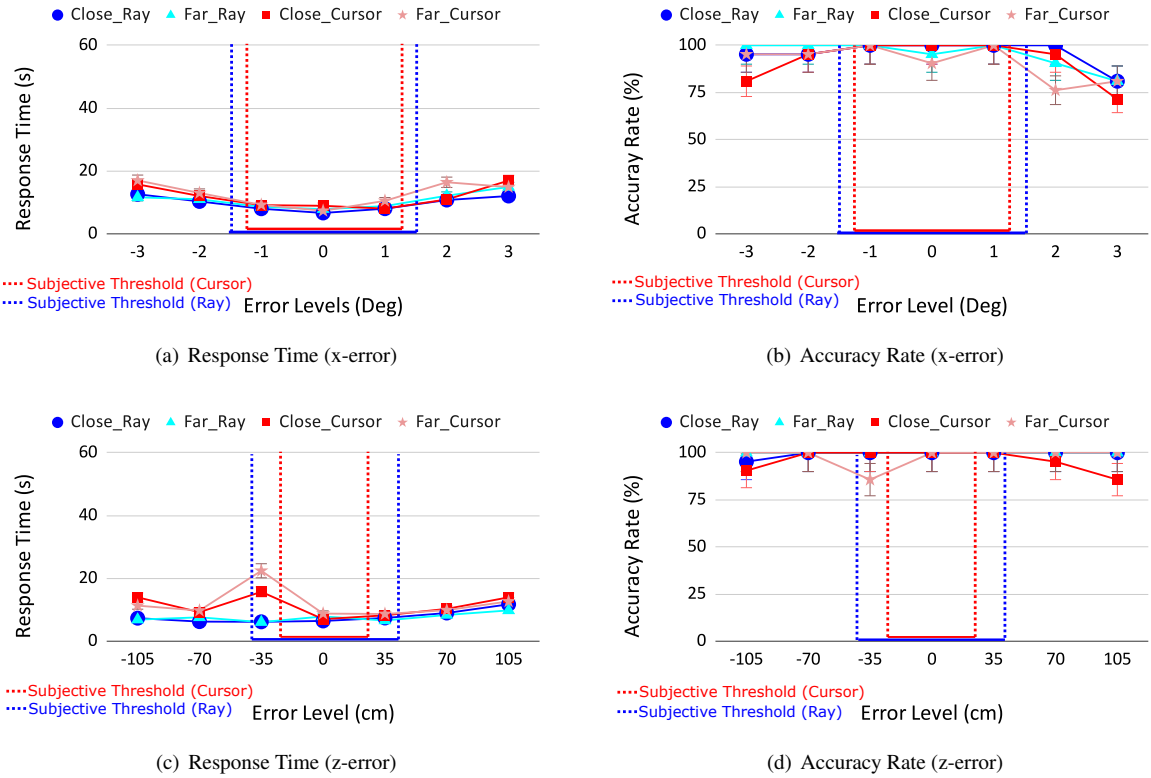
3.4.1 Objective Measures

We used repeated measures ANOVAs, and paired samples t-tests for the analysis of our results at the 5% significance level with Bonferroni correction. Shapiro-Wilk tests and Q-Q plots were used to test for normality. For cases where sphericity was not assumed through Mauchly's test, Greenhous-Geisser results were reported accordingly.

Response Time (X-Error): Figure 5 (a) shows the response times for the x-error levels at the two target depths for both visualization types. We found a significant main effect of *gaze visualization type* on response time, $F(1,20) = 13.20$, $\boldsymbol{p = 0.002}$, $\eta_p^2 = 0.39$, indicating higher response times for the cursor than the ray visualization. We also found a significant main effect of *error level*, $F(3.27, 65.40) = 34.06$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.63$. Pairwise comparisons showed a significant increase in response time for error levels with a magnitude above ± 1 (all $\boldsymbol{p < 0.05}$). We further observed a non-significant trend for a main effect of *target depth* on response time, $F(1,20) = 4.12$, $p = 0.06$, $\eta_p^2 = 0.17$.

Accuracy Rate (X-Error): Figure 5 (b) shows the results for the accuracy rates. We found a significant main effect of *error level* on accuracy rate, $F(6,120) = 11.69$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.36$. Post-hoc tests indicated that the accuracy rate decreased as the error levels increased (all $\boldsymbol{p < 0.05}$). We did not find significant main effects for *gaze visualization type*, $F(1,20) = 2.37$, $p = 0.13$, $\eta_p^2 = 0.10$, and *target depth*, $F(1,20) = 0.74$, $p = 0.39$, $\eta_p^2 = 0.03$, on accuracy rate.

Response Time (Z-Error): Figure 5 (c) shows the response times for the z-error levels at the two target depths for both visualization types. We found a significant main effect of *gaze visualization type* on response time, $F(1,20) = 76.00$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.79$, indicating higher response times for the cursor than the ray visualization. We also found a significant main effect of *error level*, $F(3.16, 63.27) = 17.95$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.47$. Post-hoc tests showed that the response time



(a) Response Time (x-error)

(b) Accuracy Rate (x-error)

(c) Response Time (z-error)

(d) Accuracy Rate (z-error)

Fig. 5. Comparison of (a) x-error response time, (b) x-error accuracy rate, (c) z-error response time, and (d) z-error accuracy rate for different error levels, gaze visualizations, and target depths.

increased as the error levels increased (all $\boldsymbol{p < 0.05}$). We found no main effect for *target depth*, $F(1,20) = 6.10$, $p = 0.44$, $\eta_p^2 = 0.03$.

In order to test for asymmetrical effects of positive and negative z-errors, we divided our results into groups for errors behind and in front of the target (zero error) and compared the means between the two sides using a paired t-test. We did not find a significant difference, $t(1) = 1.17$, $p = 0.25$, in terms of the side of the error. We looked more deeply into the interesting outlier at the error level of -35 cm and conducted a repeated measures ANOVA between the visualization types and target distances. Our results suggest that the cursor visualization for far targets significantly increased participants' response time compared to the other conditions, followed by cursor visualization for close targets, with similar significant effects (all $\boldsymbol{p < 0.05}$).

Accuracy Rate (Z-Error): Figure 5 (d) shows the results for the accuracy rates. We found a significant main effect of *visualization type* on accuracy rate, $F(1,20) = 5.65$, $\boldsymbol{p = 0.02}$, $\eta_p^2 = 0.21$, indicating lower accuracy for the cursor than the ray visualization. We found no significant main effects for *target depth*, $F(1,20) = 1.64$, $p = 0.21$, $\eta_p^2 = 0.07$, and *error level*, $F(6,120) = 1.77$, $p = 0.10$, $\eta_p^2 = 0.08$.

As for response time, we divided our results into errors behind and in front of the target and compared the means between the two sides using a paired t-test. We did not find a significant difference, $t = -0.62$, $p = 0.54$, in the side of the error. For the error level of -35 cm, a repeated measures ANOVA between the visualization types and target distances showed no significant difference between the accuracy rates.

3.4.2 Subjective Measures

We used non-parametric statistical tests for the analysis of our data.

Subjective Performance: Figure 6 (a) shows participants' confidence scores (PE1 from Table 1). We found significant differences between z-errors with ray visualization and the other three blocks, which are z-errors with cursor visualization, $W = 0.00$, $Z = -2.91$, $\boldsymbol{p = 0.005}$, x-errors with ray visualization, $W = 6.00$, $Z = -3.09$, $\boldsymbol{p = 0.002}$, and x-errors with cursor visualization, $W = 14.00$, $Z = -2.83$, $\boldsymbol{p = 0.005}$. This indicates that participants had more confidence in their answers when the ray visualization was used and z-errors were applied.

For PE2 (Table 1), we compared participants' maximum tolerated error for the blocks with x-errors and z-errors separately. Figures 6 (b) and (c) show the estimated error thresholds and tolerance regions for the experimental blocks. We found a significant difference between x-errors with ray visualization and x-errors with cursor visualization, $W = 14.00$, $Z = -2.00$, $\boldsymbol{p = 0.04}$, and a significant difference between z-errors with ray visualization and z-errors with cursor visualization, $W = 57.00$, $Z = -2.03$, $\boldsymbol{p = 0.04}$, suggesting that error thresholds were higher when gaze was visualized as a ray than as a cursor.

Task Load: Figure 6 (d) shows the task load scores for the experimental blocks. We found significant differences between ray visualization with z-errors and cursor visualization with z-errors, $W = 157.00$, $Z = -3.14$, $\boldsymbol{p = 0.002}$, and x-errors, $W = 217.00$, $Z = -3.53$, $\boldsymbol{p < 0.001}$. For x-errors, the ray and cursor visualizations were also significantly different, $W = 158.00$, $Z = -2.53$, $\boldsymbol{p = 0.01}$. Moreover, we observed a trend between x-errors and z-errors with ray visualization, $W = 143.00$, $Z = -1.93$, $p = 0.053$. These results suggest that participants estimated the ray visualization as less challenging than the cursor visualization and the z-error blocks induced less task load.

Usability: Figure 6 (e) shows the usability scores for the experimental blocks. We found significant differences between x-errors with cursor visualization and z-errors with ray visualization, $W = 5.00$, $Z = -3.73$, $\boldsymbol{p < 0.001}$, x-errors with cursor visualization and x-errors with ray visualization, $W = 40.00$, $Z = -2.21$, $\boldsymbol{p = 0.27}$, and x-errors and z-errors both with ray visualizations, $W = 14.00$, $Z = -3.27$, $\boldsymbol{p = 0.001}$. We did not find significant differences for the remaining comparisons. These results suggest an increased usability of the ray visualization compared to the cursor visualization and also the z-errors being estimated as easier compared to the x-errors.

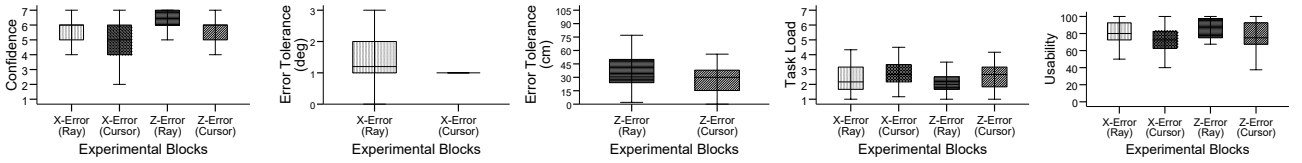


Fig. 6. Subjective comparisons of (a) confidence, (b) x-error tolerance, (c) z-error tolerance, (d) task load, and (e) usability in Experiment I.

3.5 Discussion

In Experiment I, we found significant effects of visualization type and error level on participants' objective performance and subjective evaluations. In this section, we discuss our findings in more detail.

3.5.1 Influence of Error Type and Error Level on Task Performance and Subjective Evaluation

Responding to our research questions **RQ1** and **RQ2**, we found significant effects of error levels on participants' performance for both error types, although this effect is more limited to the highest error level (i.e., +105) in the case of the z-error and more symmetric in the case of the x-error (Figures 5 a-d). These drops in performance affected both response time and accuracy rate in the case of x-errors but only response time in the case of z-errors. This may be explained by the fact that for z-errors the gaze is always oriented towards the target, providing the appropriate directional information, while this information is not maintained for x-errors, as the gaze orients to the sides of the target, resulting in a higher chance for error. This, in part, confirms our Hypothesis **H1**. For the context of our task and the utilized visualization types, we were able to observe a significant decrease in performance when the x-error levels were outside of the -1 to +1 degree range. Although, it is harder to define a clear boundary for z-error conditions, we can deduce that depth errors farther than +105 or those very close and behind the target will negatively affect users' performance.

For the case of z-errors with the ray visualization, our participants exhibited more confidence in their responses and experienced less task load compared to the experimental blocks involving other visualization modes and x-errors. These results partly confirm our Hypotheses **H3** and **H4**, which can be explained by the correct directional information provided by the z-errors.

3.5.2 Influence of Gaze Visualization on Task Performance and Subjective Evaluation

When we compare the effects of the two types of gaze visualization on the participants' performance, we can see a clear trend (Figure 5) in support of our Hypothesis **H2**. For both error types, the ray visualization type led to both significantly decreased response time and increased accuracy when compared to the cursor visualization. We believe that this is due to the directional information that can be gathered from observing the (truncated) ray that cannot be gained from the cursor alone. The inclusion of the directional information leads participants' eyes from the virtual partner's head toward the target, which is helpful both when an x-error causes the ray to 'fall off' the target, and when a z-error causes the endpoint of the (truncated) ray to stop short of the target. With respect to x-errors, on a target miss, the participant can then see which target is consistently closest to the ray and make their selection based on that information, which would be unavailable in the case of a cursor visualization. With respect to z-errors, if there is a positive error then the truncated ray acts as a pointer, pointing toward the target, and if there is a negative error, the truncated ray passes through the target as opposed to being displayed hidden behind the target as it would for the cursor visualization.

The ray visualization is further supported by the subjective analysis of the participants' confidence in their selections, task load, and usability scores, which all suggest better scores when using the ray visualization as opposed to the cursor visualization, which supports our Hypothesis **H3**.

3.5.3 Influence of Target Depth on Task Performance and Subjective Evaluation

We hypothesized in **H5** that an increased distance to the target would result in decreased performance by the user. However, no significant effects were found to support this claim. When examining the results on Figure 5, we see that the performance between close and far targets was comparable when using the ray, but when using the cursor, we see several differences between the target depths. In particular, Figures 5 (a) and (b) show that the performance when using the cursor visualization at the far target distance was worse at two degrees of positive error than it was for close target distances. One may think this was caused by the cursor 'falling off' the far target at positive two degrees, and extending to the theoretical infinity at ten meters, however if this were the case we would likely see the same effect occur on the negative two degree error level, but this could not be observed.

4 Experiment II

In this section, we describe our second human-subject study that we conducted to evaluate four types of gaze visualizations under optimal conditions, i.e., in an error-free shared gaze environment.

4.1 Participants

The same participants described in Experiment I (Section 3.1) took part in Experiment II.

4.2 Material

The same devices, physical space, shared gaze information (see Section 3.2.1), and dynamic virtual crowd (see Section 3.2.2) described in Experiment I were used in Experiment II.

Shared Gaze Visualization: Apart from the truncated ray and cursor visualization types described in Section 3.2.4, we implemented two additional types of gaze visualization shown in Figures 2 (b) and (d). The first type is the *ray + cursor* where the end point of the ray is visualized as a cursor and the length of the ray is set to the target's depth. If the ray + cursor intersects with an obstacle, the length of the ray + cursor will get adjusted to match the depth of the obstacle. The second type is the *ray* where the length of the ray is always constant and set to to the theoretical infinity which was set to ten meters away from the ray's starting position and only takes into account the target's direction but not its depth.

4.3 Methods

4.3.1 Experimental Design

We chose a 4×2 within-subject design to account for the possible individual differences affecting participants' task performance. Our independent variables were as follows:

- **Visualization Type** ($\times 4$) which were:

  Truncated Ray: The end point of the ray (i.e., ray's length) was set to the target's depth and would get truncated at a new depth if intersected by another obstacle.

  Cursor: The cursor was set to the target's depth and would get truncated at a new depth if intersected by another obstacle.

  Ray + Cursor: The end point of the ray visualized by the cursor was set to the target's depth and would get termianted at a new depth if intersected by another obstacle.

Infinite Ray: The end point of the ray was set to be at the theoretical infinity, ten meters away from its starting point without truncated at other obstacles.

- **Target Depth** ($\times 2$), which were: **Close:** Target was set at 3.15 meters from the virtual partner. **Far:** Target was set at 5.25 meters from the virtual partner.

In Experiment II, for each visualization type, each depth of target was tested twice resulting in a total of 16 trials. The trials were divided into four blocks based on each visualization type. The order with which participants were exposed to the four blocks and the four trials within them was randomized to account for learning effects.

4.3.2 Procedure

Participants were guided to the start position and asked to don the HMD. Then, the experimenter started one of the blocks for Experiment II in a randomized order. During each block, the participant was tasked with identifying four targets at two varying depth levels (note that no error was added to the gaze information as it was in Experiment I). At the end of each block, the participants were asked to fill out usability and cognitive load questionnaires. This procedure was repeated three more times. Then, the participants were asked to answer a demographics questionnaire, were interviewed for any additional feedback, and received a compensation for their participation in the study.

4.3.3 Measures

For Experiment II, we used the first performance evaluation question (see Table 1) used in Experiment I. We also measured participants' ***preference*** with regards to gaze visualization type, by asking them to order the types from their most to least preferred type.

4.3.4 Hypotheses:

Based on pilot testings we present the following hypothesis for the context of our experimental task:

- **H1:** Participants will perform better in the *ray + cursor* and *truncated ray* visualization types more than *cursor* and *infinite ray*, as the former provide both the direction of the target and also terminate at the target.
- **H2:** Participants will prefer the *ray + cursor* and *truncated ray* visualizations more than *cursor* and *ray* as the former provide both the direction of the target and also terminate at the target.
- **H3:** Participants will attribute a higher preference score to the *ray + cursor* visualization and exhibit a higher confidence in its performance compared with the other visualization types.

4.4 Results

In this section, we report our findings for Experiment II. We excluded two participants from the analysis due to technical issues.

4.4.1 Objective Measures

We used repeated measures ANOVAs for the analysis of our results at the 5% significance level with Bonferroni correction. Shapiro-Wilk tests at the 5% significance level and Q-Q plots were used to test for normality. For cases where sphericity was not assumed through Mauchly's test, Greenhous-Geisser results were reported accordingly.

Response Time: Figure 7 (a) shows the participants' response times for the four visualizations at the two target distances. We found a significant main effect of *visualization type*, $F(3,60) = 6.70$, $\boldsymbol{p = 0.001}$, $\eta_p^2 = 0.25$, and a significant interaction between visualization type and target depth, $F(2.23,44.70) = 4.36$, $\boldsymbol{p = 0.016}$, $\eta_p^2 = 0.17$, on response time. No significant main effect was found for target depth on response time, $F(1,20) = 0.11$, $p = 0.73$, $\eta_p^2 = 0.01$.

Post-hoc tests indicate a significant difference between response time for close and far targets when *ray + cursor* and *cursor* visualizations were used. Also, *ray* visualization showed a significantly higher response time for close targets compared to *ray + cursor* and *cursor*, and for far targets when compared with *truncated ray* (all $\boldsymbol{p < 0.05}$).

Accuracy Rate: We found no significant differences for accuracy rate. In fact, all participants had exactly *100% accuracy rate* in the error-free environment for all visualization types.

4.4.2 Subjective Measures

We used non-parametric statistical tests for our analysis.

Subjective Performance: Figure 7 (b) shows the confidence scores (PE1 from Table 1) for the four visualizations at the two target distances. We found significant differences between *ray + cursor* visualizations and the other three blocks *truncated ray* visualization, $W = 0.00$, $Z = -2.64$, $\boldsymbol{p = 0.008}$, *cursor* visualization, $W = 3.5$, $Z = -2.12$, $\boldsymbol{p = 0.03}$, and *infinite ray* visualization, $W = 31.5$, $Z = -2.12$, $\boldsymbol{p = 0.034}$. This suggests that participants had more confidence in their answers when the gaze was visualized through *ray + cursor*.

Preference: Figure 7 (c) shows participants' gaze visualization preference based on the above scoring system. We asked our participants to order the four visualization types from least preferred to most preferred. We assigned scores from 1 to 4 to their ranking choices, where 1 indicated the least preferred and 4 the most preferred. We found that the *ray + cursor* visualization was significantly different from *truncated ray*, $W = 45.5$, $Z = -2.71$, $\boldsymbol{p = 0.007}$, *cursor*, $W = 13.5$, $Z = -3.59$, $\boldsymbol{p < 0.001}$, and *infinite ray* visualizations, $W = 201.00$, $Z = -3.01$, $\boldsymbol{p = 0.003}$, suggesting participants' inclination towards the *ray + cursor* visualization. We also found a significant difference between *cursor* and *truncated ray* visualizations, $W = 201.00$, $Z = -3.09$, $\boldsymbol{p = 0.002}$, indicating a higher preference towards *truncated ray*.

4.5 Discussion

The results of Experiment II indicate that the shared gaze visualization approach has both subjective and objective impacts on the user.

4.5.1 Influence of Gaze Visualization Type on Task Performance

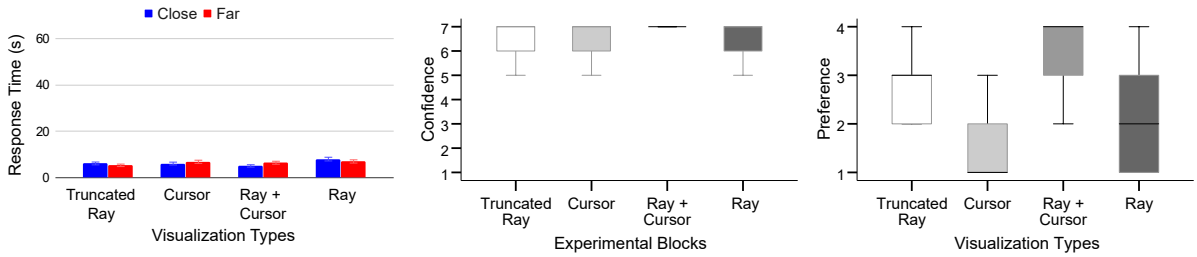
The results of Experiment II support Hypothesis **H1** in that participants performed better in their shared gaze task when utilizing the *ray + cursor* visualization type, than they do when using the other types. This suggests that the combination of the directional information from the *truncated ray* visualization and the depth information provided by the *cursor*, are better than either of the two visualization methods on their own. This is interesting, as *truncated ray* and the *ray + cursor* visualizations both include a directional component defined by the body of the ray and a depth component defined by the termination point of the ray body. The only difference between the two is that the *ray + cursor* type adds an additional 3D object at the termination point of the ray, which suggests that the 3D object itself influences the performance of the user. It is possible that the apparent size changes of this object with respect to depth can influence user performance by allowing them to better judge the distance between the object and other nearby VHs than they otherwise could using the *truncated ray* alone.

4.5.2 Influence of Gaze Visualization on Preference and Subjective Evaluation

Our results also support Hypothesis **H2** since the *ray + cursor* visualization type affected the objective performance of the participants in Experiment II, it is no surprise that it also had significant impacts on the participants' subjective preference of visualization type. *Ray + cursor* was found to be the most preferred visualization method of the tested four techniques, which is likely due to its unique combination of visual cues that were discussed in the previous paragraph.

The combination of visual cues found in the *ray + cursor* visualization type has further impacts on the confidence of the user when compared with the other visualizations. This follows the logic described in the previous paragraph, as it would be expected that having access to the most amount of information and visual cues would lead to the most amount of confidence in the performance of the user.

We additionally found that *truncated ray* was significantly preferred over the *cursor* visualization type. We believe this is because directional information has a larger impact on user preference than the inclusion of



(a) (b) (c)

Fig. 7. Comparison of (a) response times at the two target distances, (b) confidence scores, and (c) user preference for the four visualization types.

the 3D fixation object, since the *cursor* provides depth information and a 3D fixation object, but lacks the directional information inherent in the *truncated ray*, whereas the *truncated ray* provides directional and depth information but lacks the 3D fixation object.

5 General Discussion

As discussed in Sections 3.5 and 4.5, the results of our two experiments suggest that horizontal errors, depth-based errors, and the visualization types applied to shared gaze information all have significant objective and subjective impacts on the user. As usage of AR HMDs increases and the demand for shared gaze applications rises, these factors will need to be carefully examined by future developers.

5.1 Visualization Methods

The participants of both studies significantly preferred and performed better when observing shared gaze information using a combination of both truncated ray and cursor visualization types. This is not surprising, as this gaze visualization type combines two separate forms of information for the user to observe when making decisions in the shared gaze environment. The truncated ray gives directional information, leading the user from the virtual partner's head to the target, while the cursor provides a 3D object at the endpoint of the ray. This combination yields better results than either of the two techniques alone, and because of this, future gaze visualization techniques should continue to explore this and other similar visualizations that provide simultaneous direction and depth information while giving users a 3D object to fixate on. Also, as shared gaze in AR can facilitate various applications, it would be beneficial to investigate other visualizations not investigated in this work, such as arrows, cones, and spotlights, as well as the effect of the gaze ray's color based on the context of the task.

5.2 Error Thresholds

As discussed above, users experienced the best performance in Experiment I when using the truncated ray visualization type, while the results of Experiment II suggest that the ray + cursor visualization type may yield a performance that is equal to or better than that of truncated ray alone. While the ray + cursor visualization was not tested with error levels in this work, its subjective benefits suggest that future work in shared gaze scenarios should make use of this visualization method.

While user performance in angular error conditions decreased symmetrically at higher error levels, the same cannot be said for depth-based errors when using the truncated ray visualization type. The data gathered from Experiment I suggests that positive depth errors increase in difficulty more quickly than negative depth errors, likely because of negative depth errors passing through the target before termination while positive errors stop short. This is an interesting observation in that if positive depth errors are expected, a negative offset could be potentially introduced to gaze data without negatively impacting the performance of the user. Future shared gaze systems may adjust for depth errors and x-errors in accuracy automatically or manually with input from a user by visualizing one's own gaze ray. In the case of moving targets, x-errors in accuracy could potentially be adjusted for by comparing the speed of the user's gaze to the speed of nearby objects on the depth sensor, and then adjusting an angular offset so that the gaze indicator would be aligned with the target object. This approach is very similar to smooth pursuit based calibration methods [14, 33, 41]. For stationary targets, users could potentially account for x-errors in accuracy if they are able to see their own gaze ray by shifting their gaze so that the gaze indicator falls into the correct position. This offset between the user's actual gaze point and the position of the gaze indicator in either stationary or dynamic circumstances could then be used for self-calibration purposes.

5.3 Limitations

Both studies presented here involve the same scenario for the shared gaze environment, and as such, our findings must be interpreted with this specific scenario in mind. It is possible that by changing the position of the participant, or by changing the movement behavior of the virtual crowd, that the outcomes of the study may be different than what was gathered by our work. For example, if viewing the scene from another angle other than the front, it is possible that users may tolerate less or more error than what we observed for this particular scenario.

6 Conclusion

In this paper we presented a human-subject study investigating the influence of error type, error level, gaze visualization method and target depth on users' performance and subjective evaluation during a collaborative search task where participants used the gaze information of a virtual human partner to identify targets among a dynamic crowd. Our findings suggest that levels of angular and depth errors above certain thresholds decrease users' performance. Also, for environments susceptible to error, a truncated ray visualization results in a higher performance and less cognitive load for the user. Separately, we investigated the effects of shared gaze visualization, in a similar but error-free environment, finding that the addition of a 3D cursor to the end point of the truncated ray enhances users' performance for close targets, increases their confidence, and achieved the highest preference ranking compared to the other visualization methods.

In the future, we plan to investigate the influence of error in search tasks in static scenes under collaborative circumstances, to improve our understanding of the behavioral differences in dynamic scenes, and of search tasks in general. Also, we plan to investigate and compare how shared gaze influences the quality of collaboration in VR/AR when two real human users are taking part in the task.

Acknowledgments

This material includes work supported in part by the National Science Foundation under Award Number 1564065 (Dr. Ephraim P. Glinert, IIS) and Collaborative Award Numbers 1800961, 1800947, and 1800922 (Dr. Tonya Smith-Jackson, IIS) to the University of Central Florida, University of Florida, and Stanford University respectively; the Office of Naval Research under Award Number N00014-17-1-2927 (Dr. Peter Squire, Code 34); and the AdventHealth Endowed Chair in Healthcare Simulation (Prof. Welch). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the supporting institutions.

References

[1] M. Billinghurst and H. Kato. Collaborative mixed reality. In *Proceedings of the First International Symposium on Mixed Reality*, pp. 261–284, 1999.

[2] M. Billinghurst and H. Kato. Collaborative augmented reality. *Communications of the ACM*, 45(7):64–70, 2002.

[3] P. Blignaut and D. Wium. Eye-tracking data quality as affected by ethnicity and experimental design. *Behavior Research Methods*, 46(1):67–80, 2014.

[4] S. E. Brennan, X. Chen, C. A. Dickinson, M. B. Neider, and G. J. Zelinsky. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 106(3):1465–1477, 2008.

[5] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194):4–7, 1996.

[6] S. D'Angelo and D. Gergle. Gazed and confused: Understanding and designing shared gaze for remote collaboration. In *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 2492–2496, 2016.

[7] S. D'Angelo and D. Gergle. An eye for design: Gaze visualizations for remote collaborative work. In *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.

[8] R. A. J. de Belen, H. Nguyen, D. Filonik, D. Del Favero, and T. Bednarz. A systematic review of the current state of collaborative mixed reality technologies: 2013–2018. *AIMS Electronics and Electrical Engineering*, pp. 1–43, 2019.

[9] A. Dey, M. Billinghurst, R. W. Lindeman, and J. Swan. A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI*, 5:1–28, 2018.

[10] C. Elmadjian, P. Shukla, A. D. Tula, and C. H. Morimoto. 3d gaze estimation in the scene volume with a head-mounted eye tracker. In *ACM Workshop on Communication by Gaze Interaction*, pp. 1–9, 2018.

[11] A. M. Feit, S. Williams, A. Toledo, A. Paradiso, H. Kulkarni, S. Kane, and M. R. Morris. Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design. In *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 1118–1130, 2017.

[12] K. Fitzpatrick, M. A. Brewer, and S. Turner. Another look at pedestrian walking speed. *Transportation Research Record*, 1982(1):21–29, 2006.

[13] S. R. Fussell and L. D. Setlock. Using eye-tracking techniques to study collaboration on physical tasks: implications for medical research. *Unpublished manuscript, Carnegie Mellon University*, pp. 1–25, 2003.

[14] A. R. Gomez and H. Gellersen. Smooth-i: smart re-calibration using smooth pursuit eye movements. In *ACM Symposium on Eye Tracking Research & Applications*, pp. 1–5, 2018.

[15] R. Grasset, P. Lamb, and M. Billinghurst. Evaluation of mixed-space collaboration. In *IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 90–99, 2005.

[16] K. Gupta, G. A. Lee, and M. Billinghurst. Do you see what i see? the effect of gaze tracking on task space remote collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 22(11):2413–2422, 2016.

[17] J. E. Hanna and S. E. Brennan. Speakers eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, 57(4):596–615, 2007.

[18] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Human Factors and Ergonomics Society Annual Meeting*, pp. 904–908, 2006.

[19] K. Holmqvist, M. Nyström, and F. Mulvey. Eye tracker data quality: what it is and how to measure it. In *ACM Symposium on Eye Tracking Research and Applications*, pp. 45–52, 2012.

[20] M. Kassner, W. Patera, and A. Bulling. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pp. 1151–1160, 2014.

[21] K. Khoshelham and S. O. Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454, 2012.

[22] K. Kim, M. Billinghurst, G. Bruder, H. B.-L. Duh, and G. F. Welch. Revisiting trends in augmented reality research: A review of the 2nd decade of ismar (2008–2017). *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2947–2962, 2018.

[23] K. Kiyokawa, H. Takemura, and N. Yokoya. A collaboration support technique by integrating a shared virtual reality and a shared augmented reality. In *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 48–53, 1999.

[24] S. R. Langton, R. J. Watt, and V. Bruce. Do the eyes have it? cues to the direction of social attention. *Trends in Cognitive Sciences*, 4(2):50–59, 2000.

[25] G. Lee, S. Kim, Y. Lee, A. Dey, T. Piumsomboon, M. Norman, and M. Billinghurst. Improving collaboration in augmented video conference using mutually shared gaze. In *International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, pp. 197–204, 2017.

[26] Y. Lee, C. Shin, A. Plopski, Y. Itoh, T. Piumsomboon, A. Dey, G. Lee, S. Kim, and M. Billinghurst. Estimating gaze depth using multi-layer perceptron. In *International Symposium on Ubiquitous Virtual Reality*, pp. 26–29, 2017.

[27] Y. Liu, H. Dong, L. Zhang, and A. El Saddik. Technical evaluation of hololens for multimedia: a first look. *IEEE MultiMedia*, 25(4):8–18, 2018.

[28] D. Mardanbegi, T. Langlotz, and H. Gellersen. Resolving target ambiguity in 3d gaze interaction through vor depth estimation. In *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[29] K. Masai, K. Kunze, M. Billinghurst, et al. Empathy glasses. In *ACM SIGCHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1257–1263, 2016.

[30] M. B. Neider, X. Chen, C. A. Dickinson, S. E. Brennan, and G. J. Zelinsky. Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review*, 17(5):718–724, 2010.

[31] N. Norouzi, A. Erickson, K. Kim, R. Schubert, J. J. LaViola, G. Bruder, and G. F. Welch. Effects of shared gaze parameters on visual target identification task performance in augmented reality. In *ACM Symposium on Spatial User Interaction*, pp. 1–11, 2019.

[32] M. Nyström, R. Andersson, K. Holmqvist, and J. Van De Weijer. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, 45(1):272–288, 2013.

[33] K. Pfeuffer, M. Vidal, J. Turner, A. Bulling, and H. Gellersen. Pursuit calibration: Making gaze calibration less tedious and more flexible. In *ACM Symposium on User Interface Software and Technology*, pp. 261–270, 2013.

[34] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: an adaptive avatar for mixed reality remote collaboration. In *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2018.

[35] T. Piumsomboon, Y. Lee, G. A. Lee, A. Dey, and M. Billinghurst. Empathic mixed reality: Sharing what you feel and interacting with what you see. In *IEEE International Symposium on Ubiquitous Virtual Reality*, pp. 38–41, 2017.

[36] B. Schneider and R. Pea. Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, 8(4):375–397, 2013.

[37] R. Stein and S. E. Brennan. Another person's eye gaze as a cue in solving programming problems. In *ACM International Conference on Multimodal Interfaces*, pp. 9–15, 2004.

[38] W. Steptoe, R. Wolff, A. Murgia, E. Guimaraes, J. Rae, P. Sharkey, D. Roberts, and A. Steed. Eye-tracking for avatar eye-gaze and interactional analysis in immersive collaborative virtual environments. In *ACM Conference on Computer-Supported Cooperative Work*, pp. 197–200, 2008.

[39] S. Trösterer, M. Wuchse, C. Döttlinger, A. Meschtscherjakov, and M. Tscheligi. Light my way: Visualizing shared gaze in the car. In *ACM International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 196–203, 2015.

[40] R. Vassallo, A. Rankin, E. C. Chen, and T. M. Peters. Hologram stability evaluation for microsoft hololens. In *Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment*, vol. 10136, pp. 1–7. International Society for Optics and Photonics, 2017.

[41] M. Vidal, A. Bulling, and H. Gellersen. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 439–448, 2013.

[42] X. Wang, D. Lindlbauer, C. Lessig, and M. Alexa. Accuracy of monocular gaze tracking on 3d geometry. In *Workshop on Eye Tracking and Visualization*, pp. 169–184, 2015.

[43] M. Weier, T. Roth, A. Hinkenjann, and P. Slusallek. Predicting the gaze depth in head-mounted displays using multiple feature regression. In *ACM Symposium on Eye Tracking Research & Applications*, pp. 1–9, 2018.

[44] G. F. Welch, G. Bruder, P. Squire, and R. Schubert. Anticipating widespread augmented reality: Insights from the 2018 ar visioning workshop. Technical Report 786, University of Central Florida and Office of Naval Research, 2019.

[45] L. Yang, L. Zhang, H. Dong, A. Alelaiwi, and A. El Saddik. Evaluating and improving the depth accuracy of kinect for windows v2. *IEEE Sensors Journal*, 15(8):4275–4285, 2015.

[46] Y. Zhang, K. Pfeuffer, M. K. Chong, J. Alexander, A. Bulling, and H. Gellersen. Look together: Using gaze for assisting co-located col-

laborative search. *Personal and Ubiquitous Computing*, 21(1):173–186, 2017.