IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Gaze Analysis in Early Warning Visual Feedback System for Hand Tracking Failures in Virtual Reality

MUCAHIT GEMICI[1], AMAL HATIRA[2], VRUSHANK PHADNIS[3] and Anil Ufuk Batmaz [4], (Member, IEEE)

[1]Computer Science and Software Engineering Department, Concordia University, Montreal, QC, Canada (e-mail: mucahit.gemici@mail.concordia.ca)
[2]Computer Science and Software Engineering Department, Concordia University, Montreal, QC, Canada (e-mail: amal.hatira@mail.concordia.ca)
[3]Google, Mountain View, California, United States, (e-mail: vrushank@google.com)
[4]Computer Science and Software Engineering Department, Concordia University, Montreal, QC, Canada (e-mail: ufuk.batmaz@concordia.ca)

Corresponding author: Anil Ufuk Batmaz (e-mail: ufuk.batmaz@concordia.ca).

This work was supported and funded by Google Inc.

ABSTRACT Hand-tracking failures significantly affect user performance and experience in VR; thus, minimizing the negative effects of such failures remains a critical area of research. Although prior studies have introduced early-warning feedback systems to mitigate hand-tracking failures, it remains unclear *why* and *how* such feedback enhances user performance and reduces frustration. To address this gap, we investigated user gaze behaviors in a visual early-warning feedback system with three common hand-tracking failure scenarios: Low-Intensity Light Level, Out-of-Vision Hands, and Self-Occlusion. Our findings indicate that user attention toward feedback notifications varies depending on the type of simulated tracking failure. Providing early-warning feedback did not disrupt users' attention to their primary tasks. Instead, users maintained their focus on object placement, resulting in fewer placement errors during pick-and-place tasks. These results contribute to our understanding of early-warning visual feedback systems and provide valuable insights for designing feedback mechanisms that effectively improve performance without creating distractions in VR interactions.

INDEX TERMS Early Warning, Feedback, Hand Tracking, Hand Tracking Failures, Virtual Reality

I. INTRODUCTION

Hand interactions in Virtual Reality (VR) have gained widespread popularity because they enable users to interact intuitively and naturally with virtual objects and immersive environments [34]. Unlike traditional controller-based interactions, hand interactions provide a more direct and realistic experience, which significantly enhances precision-demanding tasks, such as carefully positioning and rotating virtual surgical instruments or assembling complex virtual components [39]. For instance, surgeons in VR can intuitively grasp and manipulate tools, replicating real-world movements closely and thus reducing cognitive overload that arises when translating physical actions into virtual movements [21]–[23]. Consequently, these natural interactions not only foster increased user engagement but also improve overall task efficiency by reducing cognitive load [33].

To be able to translate users' hand movements to a virtual environment, current low-cost VR head mounted display (HMD) systems use hand tracking systems. These systems can include depth sensors (e.g., [63]), trackers (e.g., [64]), RGB cameras (e.g., [65]), hand-tracking algorithms (e.g., [66]), and image processing (e.g., [67]). The main purpose of such a hand tracking system is to accurately map real-world hand movements to the virtual world. Yet, existing VR hand-tracking technologies still face various technical challenges that can compromise their effectiveness. Common issues include poor lighting conditions that disrupt camera-based tracking, hand occlusions when fingers overlap or block the view of each other, or situations in which users inadvertently move their hands beyond the tracking area of the headset's cameras [22], [24], [25]. For example, during a complex virtual assembly task, a user's hand might momentarily go out of the headset's camera view, causing immediate tracking inaccuracies, such as erroneous hand position. Such scenarios frequently result in unsuccessful task completion and increased frustration, as users need to repeatedly correct their errors.

To mitigate these detrimental effects, recent research has explored the implementation of early-warning feedback systems [32], [35]. These systems provide notifications just *before* a hand tracking failure occurs, offering users the opportunity to adjust their hand positions or movements or prepare themselves for a critical hand tracking error. Typically, this feedback consists of subtle yet noticeable visual cues, such as notifications appearing within the user's peripheral vision [11], [32], [35]. For instance, if a user's hand approaches a known occlusion area, a visual notification might appear as an alert in the scene to prevent imminent tracking failure [35].

Previous studies demonstrated that such early-warning feedback significantly improves usability, user satisfaction, and task efficiency while reducing cognitive load [32], [35]. However, these prior works mainly evaluated task outcomes, such as completion time or accuracy, without exploring *why* these systems effectively reduce cognitive load and improve usability. Understanding the underlying mechanisms might help designers create feedback systems that align better with human cognitive processing and make it easier to generalize findings to other VR contexts and tasks.

Therefore, this study shifts the focus from performance outcomes to the behavioral dynamics of gaze and attention when early-warning feedback is presented, and aims to fill that gap by examining how early-warning feedback affects users' visual attention and gaze behaviors during VR interactions. Specifically, we aim to investigate whether participants notice and respond to the early-warning visual feedback, how their visual attention shifts between the feedback and the task, and whether the feedback introduces any unintended distractions. To our knowledge, none of the attentional or (gaze) behavioral evaluation metrics have been previously examined in the context of early-warning visual feedback for VR hand tracking failures.

To investigate the gaze movement of the users when early-warning feedback is shown, we replicated the previous work [3] and designed a simple pick-and-place task where the participants had to place objects in the correct space in the task environment. While the participants performed the experiment, we collected their gaze data and investigated how their gaze moved during the task. Our contributions are:

- We showed that early-warning feedback reduces hand-tracking errors without disrupting user focus or task time, which might contribute to reduced cognitive workload
- We identified how attention to feedback varies by tracking error type, informing effective feedback design.
- We demonstrated that repeated feedback exposure leads to habituation, which might be used for future feedback system improvements.

These insights collectively contribute to a deeper understanding of how visual feedback systems influence user performance and attentional behaviors in early warning systems in VR, guiding future designs of feedback mechanisms to enhance user experience without introducing unintended distractions.

II. RELATED WORK

A. VISUAL NOTIFICATION AS FEEDBACK

Visual notifications in VR serve as crucial feedback mechanisms that increase user experience by providing timely information without breaking immersion [52]. These notifications can be designed using different metaphors and strategically placed to reduce distraction while maximizing effectiveness [55]. Their integration into VR environments is essential for maintaining user awareness of both the virtual and physical surroundings, thereby improving safety and task performance [53].

By keeping users informed of real-world events—such as incoming calls or obstacles in their physical space—visual notifications can prevent accidents and promote seamless interaction [55]. Additionally, they offer contextual feedback that helps users better understand their environment, enhancing decision-making and task execution [68].

Research shows that the placement and design of visual notifications significantly influence user response times and perceived intrusiveness, both of which impact overall task performance [53]. For example, effective visual feedback has been found to improve hand movement accuracy in confined and occluded spaces, highlighting its role in facilitating more precise interactions with virtual objects [54].

Previous work on early-warning systems in virtual environments employed visual notifications as a feedback mechanism, presenting warnings as visual pop-ups [35]. These notifications were triggered before hand tracking failures occurred, enabling participants to anticipate and respond to potential failures. In this paper, we adopt a similar approach by implementing visual notifications as feedback to replicate and build upon the findings of prior research [35].

B. USING GAZE AS AN INPUT AND OUTPUT CHANNEL

Gaze-aware feedback systems have been an important research area within HCI and the XR community. Most of this work treats gaze as an input modality, i.e., using eye-tracking data to identify user attention or intent and trigger system adaptations, such as including UI personalization (e.g., dynamic re-arrangement of elements [70]), attention guidance [73], combining with gesture to control a system (e.g., [71]) and gaze-only interaction to support accessibility and hands-free control [75]. These systems aim to enhance efficiency, reduce cognitive overhead, and increase adaptability by responding dynamically to the user's point of focus.

In contrast to using gaze as an input channel, our work shifts the role of gaze from input to output, which is also frequently studied in HCI, e.g., [68], [72]. Many gaze-based workload studies involve static or observational tasks, rather than dynamic, error-prone manipulation tasks that require moment-to-moment correction. Rather than using gaze to control the interface, here, we use it to investigate *how* users respond to visual early-warning feedback. Prior work on early-warning feedback [32], [35] has shown performance and usability improvements through early warning visual feedback, while other research has used gaze to estimate workload or attention [43], [51]. Yet, to the best of our knowledge, none of the studies combine early-warning visual feedback with gaze behavior analysis in a controlled, interactive VR task.

C. MEASURING ATTENTION FROM GAZE

The measurement of attention in VR environments has emerged as a critical area of research, leveraging gaze tracking to understand cognitive processes. Previous literature used the gaze information to measure user attention [4], [5], [16], [30], [31]. Movement of the eyes reveals which visual elements users pay attention to, how long time engage with them, or which other directions they focus on [12]. Saccades and Fixations are well known concepts in eye-tracking systems [16].

Fixation is the relatively stationary behaviour of the gaze direction [5], [13]. While the fixation duration provides information about a specific area's content or complexity [12], the sum of fixation durations can be used to determine the allocation of attention across multiple targets [5]. A longer fixation duration may indicate difficulty in information extraction [15]. Moreover, if a gaze falls on an area of interest prior to a possible event, it can be an indicator of anticipation [14].

Besides fixations, saccades are rapid eye movements that occur between fixations [5]. Even though Saccades are related to searching a sequence of particular areas of interest, in the context of visual attention- and this paper-, saccadic metrics are not as relevant as fixations due to visual information not being processed [16].

Separately, gaze metrics have been used to infer user cognitive workload or attention in immersive tasks [43], [45]. Studies show that metrics such as fixation duration, frequency, and saccade patterns correlate with mental effort, particularly under varying task complexity. For example, [43] used fixation density to approximate workload across interface types. However, such studies rarely include system-generated visual cues or explore dynamic interaction tasks where error anticipation and correction play a role.

D. WARNING SYSTEMS FOR ERRORS

Hand tracking performance in VR systems is affected by errors [27]–[29]. There are various error detection approaches and corresponding correction systems to reduce errors and improve performance. Sendhilnathan et al. [8] showed gaze dynamics can effectively distinguish between accurately recognized inputs, input recognition errors, and user errors across different tasks. In another example, the error-related potential (ErrP) detection ability of EEG signals was explored by Wimmer et al. [9]. While auditory feedback has been found to reduce error rates without affecting task time [10], it is possible to predict hand movements to detect erroneous actions to show feedback [11]. However, all these studies focus on increasing user performance *after* the errors happened. Moreover, these error feedback mechanisms are mostly related to errors related to the task itself to prevent confusion and unconscious errors during task execution. (Most) gaze-based feedback systems are designed for adaptive interface control or attention monitoring (e.g., for redirecting focus or modifying display content), whereas our focus is on understanding attentional responses to early-warning cues, not using gaze as an input modality.

However, early-warning feedback systems warns users *before* the tracking failure happens. Moreover, it is not designed as a preemptive method, i.e., it does not force users to change their behavior or movement to avoid hand tracking failures.

III. MOTIVATION AND RESEARCH QUESTIONS

Recent studies have demonstrated that early-warning feedback can effectively reduce negative impacts of hand-tracking failures by alerting users just before errors occur, thus providing a critical opportunity for users to mentally prepare and adjust their movements [32], [35]. However, prior research has mainly focused on evaluating performance outcomes—such as task completion time and accuracy—without exploring the underlying mechanisms responsible for these improvements, particularly how such feedback influences visual attention and gaze behaviors.

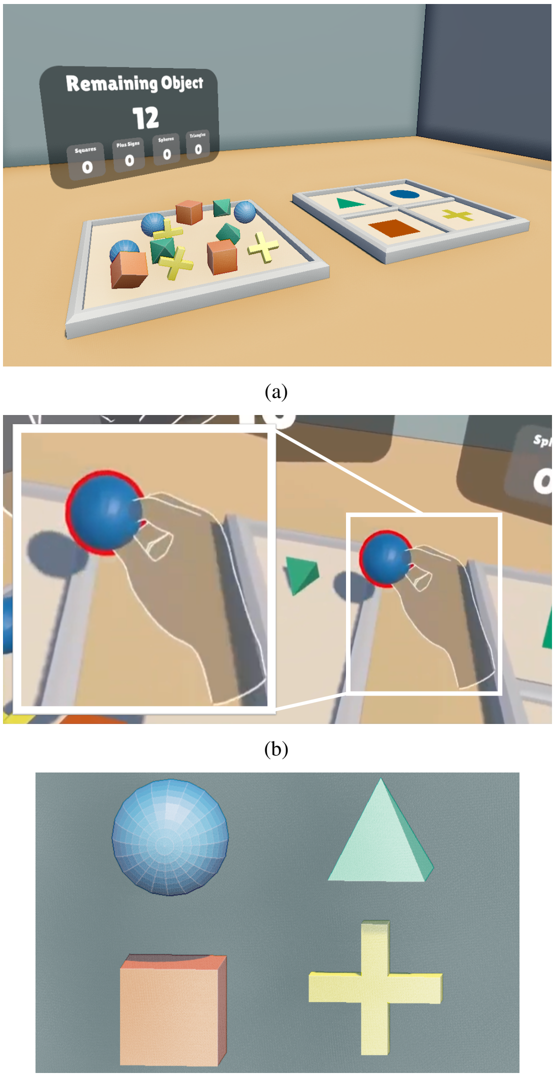
Gaze behavior is closely linked to cognitive load, providing insights into how users allocate attention and process visual information [47], [48]. When users engage in tasks that impose high cognitive demands, their gaze patterns typically become more erratic, featuring frequent gaze shifts and shorter fixation durations [49], [50]. Such rapid shifts in visual attention often indicate active, effortful cognitive processing. For instance, longer fixation durations generally reflect deeper cognitive engagement with specific visual elements, whereas shorter, frequent fixations may signal confusion or inefficiency due to information overload [51].

By analyzing gaze movements during the activation of early-warning feedback, we can explore precisely how and why such feedback may help reduce users' cognitive load. Understanding these gaze behaviors can clarify whether the feedback allows users to allocate attention more efficiently, minimizing unnecessary visual search and reducing overall cognitive effort. While previous studies have explored early visual warning systems (e.g., [32], [35]) and gaze-based workload estimation in immersive environments (e.g., [43], [48]), to our knowledge, no prior work combines these two perspectives in a single, controlled experimental framework focused on hand-tracking failures in VR.

Therefore, this study is motivated by the need to investigate the attentional and cognitive mechanisms behind the effectiveness of early-warning systems. By exploring the relationship between early-warning feedback, gaze dynamics, and cognitive load, we aim to inform the design of more intuitive, less cognitively demanding VR notification systems. Thus, in this paper, we investigated the following research questions (RQs):

(RQ1) How do users' gaze movements change when early-warning visual feedback systems are activated? **(RQ2)** In which context do early-warning systems help reduce users' cognitive load?

IV. METHODOLOGY



(a)

(b)

(c)

FIGURE 1. (a) Task environment. Objects are initialized to the left. Their designated target areas are located at the right and color-coded by category. A counter shows the number of object placements remaining and the error rate for each category. (b) A red outline appears around an object when the user grabs it. (c) Task objects. From left to right: sphere, cube, square-base pyramid, and plus-sign.

A. PARTICIPANTS

We recruited 12 participants (6 Female and 6 Male) from a local university with ages ranging between 21 and 31 (M = 24.75, SD = 3.10). This study was approved by Concordia University Human Research Ethics Committee, with certification number of 30019273. All participants had normal or corrected-to-normal vision. One participant reported no VR HMD experience, three reported having used VR 1-3 times, and eight reported more than 5 times. We asked participants how familiar they were with hand interactions in VR.

B. APPARATUS

The user study was carried out on a desktop PC equipped with an Intel(R) i7-12700F processor running at 2.1GHz, 16GB of RAM, and an NVIDIA GeForce RTX 3060 Ti graphics card. For the VR HMD, we used the Meta Quest Pro, which has an eye tracker with an average accuracy of $1.652°$, as well as a precision of $0.699°$ (standard deviation) and $0.849°$ (root mean square) for a visual field spanning $15°$ [17]. There were also two 2D 27''monitors attached to the computer where the experimenter could observe the participants' movements. The 3D models for the virtual environments were created in Blender 4.0, while user interface components were designed using Figma. The VR application was developed in Unity 2022.3.34f1, with the Meta XR All-In-One SDK 66.0.0.

C. PICK AND PLACE TASK

The experiment involves a pick-and-place sorting task where participants are instructed to sort and organize virtual objects according to their colors and shapes. This task was also used in the previous work to investigate the efficiency of the early-warning systems [35]. The previous work chose this task to ensure participants could focus on the experience rather than the task mechanics.

The pick-and-place task requires participants to pick objects from a platform on their left and place them onto a specific zone matching their shape-and-color on a platform on their right as shown in Fig. 1(a). We also placed a scoreboard as a user interface that displays the number of remaining objects. A total of 16 objects are used in the task: 4 spheres, 4 plus signs, 4 cubes, and 4 square-base pyramids (Fig. 1(c)). Even though we did not have any color blind participants, we selected colors accessible to color-blind users for this study.

Upon grabbing an object, a red outline appears around the object (Fig. 1(b)), accompanied by a sound cue, signaling the start of the task. Participants then placed the object in its corresponding target area on the right platform. If the object is placed correctly, a success sound is triggered, and the counters for both manipulated and remaining objects are updated. No feedback is provided for incorrect placements. Participants had to correct their wrong placements by moving the object to the appropriate target area.

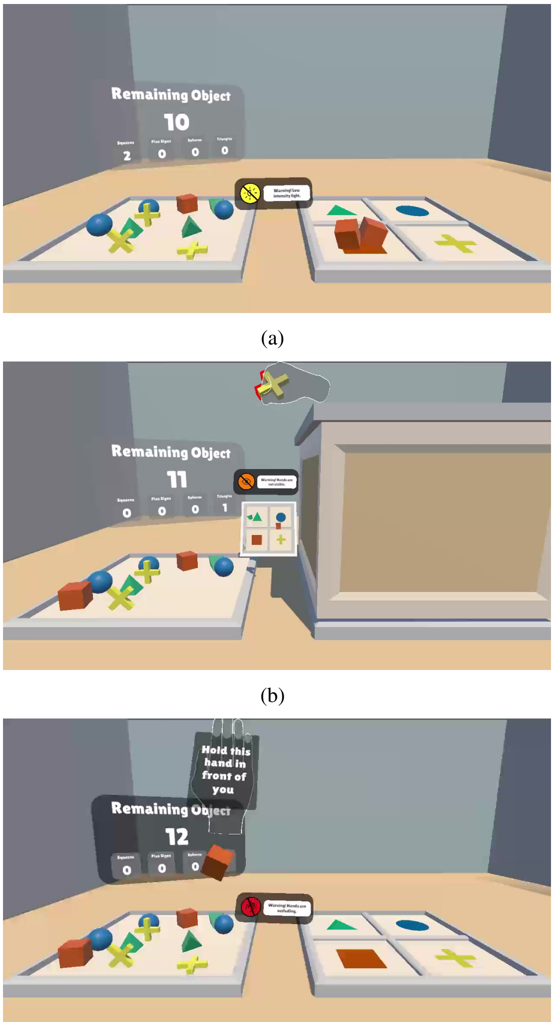
D. HAND-TRACKING ERRORS

As in previous work [35], we designed three different hand tracking error conditions.

1) Low Intensity Light

In the Low-Intensity Light condition, we changed the lighting conditions of the experiment room using a Wizard-of-Oz method. We performed the experiments in a room without any daylight so we could control the light conditions.

In this condition, the experimenter dimmed the lights in the room according to prompts displayed on a computer screen.



(a) (b) (c)

FIGURE 2. Visualization of the virtual environment and the three simulated hand-tracking errors: (a) Low-Intensity Light; (b) Out of Vision Hands – virtual walls requiring hand movements outside the headset's camera field of view; (c) Self-Occlusion – where one hand obstructs the tracking of the other

A Unity script guided the procedure by displaying a countdown that indicated when to turn the lights on and off. The experimenter followed these prompts and dimmed the room lighting accordingly. To prepare for each simulated tracking failure, the system signaled the experimenter 6 seconds in advance. During this time, the experimenter began dimming the lights, and visual feedback was simultaneously presented to the participant for 3 seconds after the countdown started (Fig. 2(a)). By the end of the 6-second window, the room was completely dark, triggering a deliberate tracking failure. The lights remained off for 2 seconds, after which the experimenter immediately turned them on. To prevent participants from anticipating the timing of these simulated errors, the algorithm randomized the initiation of each lighting event throughout the user study.

2) Out of Vision Hands

For the Out-of-Vision Hands condition, we designed a virtual environment that included an opaque wall surrounding the platform on the right side. These walls blocked the participants' direct line of sight to the interior of the platform, requiring them to lift their hands above the wall to correctly place the objects. Additionally, we placed a 2D screen between the two platforms, which displayed a top-down view of the right platform. This setup forced participants to rely on indirect vision — they had to look at the screen to accurately position the object on the target.

Because the VR HMD's hand-tracking system relies on cameras embedded in the headset, its vertical FoV is limited and shifted relative to the user's perspective. As participants lifted their hands above the level of the opaque wall to perform the task, their hands frequently moved outside the tracking FoV, as shown in Fig. 2(b). This caused hand-tracking failures. To correctly place the object with hand tracking, the participant had to move their hand and move their head slowly forward.

To predict and warn users before a hand-tracking failure occurred, the system continuously tracked three metrics: Forward Angle, Right Angle, and Relative Position-Y. These were calculated based on the vector distance between the user's eyes and their hands. Specifically, the Forward Angle was computed using the equation shown in Eq. (1), where $\overrightarrow{A}$ represents the vector from the eyes to the hand, and $\overrightarrow{B}$ denotes the forward-facing direction of the user's gaze.

$$angle = \arccos\left(\frac{(\overrightarrow{A} \cdot \overrightarrow{B})}{(\|\overrightarrow{A}\| * \|\overrightarrow{B}\|)}\right) \tag{1}$$

The same formula shown in Eq. (1) was also applied to compute the Right Angle, where $\overrightarrow{A}$ represented the vector from the eyes to the hand, and $\overrightarrow{B}$ corresponded to the rightward direction of the user's gaze. For Relative Position-Y, we used the Y-axis component of the hand's distance vector. This approach allowed us to account for differences in participants' arm lengths, ensuring more consistent detection across users.

A warning was triggered under two predefined conditions:

- When Relative Position-Y was 17 cm or higher, and either the Forward Angle or Right Angle exceeded 0.7 degrees.
- When Relative Position-Y was 0 cm or higher, and either angle exceeded 0.6 degrees.

The first condition was designed to detect when users raised their hands above head level, while the second targeted hand movements occurring in the upper-ipsilateral region.

3) Self Occlusion

In the Self-Occlusion condition, participants were asked to hold their non-dominant hand in front of their face. A visual cue labeled "Hold this hand in front of your face" was displayed on the hand itself to guide placement (see Fig. 2(c)). This positioning caused the dominant hand to move behind the non-dominant one, resulting in reduced hand-tracking quality. When the participants moved behind their right hand behind their left hand, the hand tracking system failed.

In this experiment, we build upon previous work [35] by focusing on gaze behavior analysis during the same pick-and-place sorting task under the simulated hand-tracking error conditions. Specifically, we aim to investigate how participants' visual attention shifts when interacting with virtual objects and attempting to correct hand-tracking errors, especially after the early-warning feedback is displayed.

E. EXPERIMENTAL DESIGN

We designed a two-factor within-subjects study with 3 **Simulated error** conditions (3_{SE}: Low Intensity Light, Out of Vision Hands, and Self-Occlusion) and 2 Visual Feedback conditions (2_{VF}: ON and OFF). Combining these factors resulted in (3_{SE} X2_{VF} = 6_{EC}) 6 experiment conditions.

Using G*Power software [20], with a significance criterion of α = .05 and power = .80, we calculated the minimum sample size needed with large effect size ($\eta^2 = 0.14$) as N = 12 for RM ANOVA. Eventually, we recruited 12 participants for the study and counterbalanced the order of conditions using a Latin Square design to eliminate any potential bias effects.

Each participant completed all conditions three times (6_{EC} X $3_{repetitions}$ = 18_{trials}), resulting a total of 18_{trials}X $12_{participants}$ = 216 data points.

F. PROCEDURE

Upon arrival, participants were first asked to provide their consent and complete a pre-experiment questionnaire regarding their demographic information. The participants were then instructed to wear the headset and the eye tracking was calibrated for each participant.

Before starting the experiment, participants were given a training session to familiarize themselves with the virtual environment, tasks, and the three simulated hand-tracking errors. During the training phase, we did not record any data. The training session took around 5 minutes and once participants were comfortable with the system, the experiment began. The participants performed the pick-and-place sorting task with 6 different experiment conditions. The experiment took approximately 25 minutes.

G. MEASUREMENT OF FIXATION AND SACCADES

To capture the Fixations and Saccades, we classified each object in the virtual environment under Feedback, TaskUI, Hand, Target, Object, Object UI, and Environment.

- **Feedback** represents the visual notifications that appear in the scene, which are placed between two platforms.
- **TaskUI** represents the UI element showing the remaining objects in the task, which is placed behind the platfroms.
- **Hand** represents the virtual hand avatar of the participants.
- **Target** represents the target area where participants must drop off for each manipulated object class.
- **Object** represents the 4 manipulated objects: Squares, Spheres, Square-Based Pyramids, and Plus Signs.
- **ObjectUI** represents 4 objects' counters, which are placed on the UI behind the platform.
- **Environment** represents the rest of the virtual objects, such as the wall behind table.

After grouping every object in the virtual environment, we used the Raycasting method to determine which objects the participant looked at. However, due to eye tremors, microsaccades, and noise, filtering must be applied to capture Fixation and Saccades correctly [50]. Thus, we first applied a One-Euro Filter [3] with a 120 Hz frequency that is selected by conducting a pilot study, to the simulated eyes' rotation to reduce jittery movement, then calculated the average position and direction of the eyes and cast a ray from the eye position to the eye forward direction.

We implemented an IDT (Dispersion-Threshold Identification) [6], [7] approach to capture Fixations and Saccades. During the ray casting, we recorded a fixation if the participant looked at the same object for more than 250 ms, and if the participant looked at an object less than 250 ms and more than 20 ms, we recorded it as a Saccade. These Fixation and Saccade thresholds were selected according to the ranges defined by Poole and Ball [5].

H. EVALUATION METRICS

1) Performance

As in previous work on early-warning systems [35], we recorded two dependent variables. **Task Time** represents the time between acquiring the first object and the correct placement of the last object. The other metric, **Wrong Placements**, is the number of attempts when the participant puts the object into the wrong target area.

2) Fixation-Based

We used Fixation-Based metrics to measure user attention, **Number of Fixation** was used and calculated by dividing the number of fixations by the number of feedback spawns in that condition to normalize the metric. The **Fixation Duration** is the time between the start and end of the fixation. Additionally, **Time to First Fixation** is the start time of the first fixation within a feedback period. We also calculated the **Attention Percentage**, which has a similar concept to Faleel et al. [4], that is, the ratio of the fixation duration of an object class to the sum of fixation durations of all object classes. We used the Attention Percentage to observe the user attention distribution of object classes.

3) Saccade-Based

We recorded Saccade-Based metrics for further data analysis regarding particiapants' attention. The first metric was the **Saccade Duration**, which is the time between the start and end of the saccade. Another metric was the **Time to First**

Saccade, which was similar to the Time to First Fixation, but this time, we used saccade instead of fixation.

V. RESULTS

The data were analyzed using SPSS 24. We considered the data to be distributed if Skewness (S) and Kurtosis (K) were within ± 1 [1], [2]. If not, we performed a log-transform to yield a normal distribution. We first evaluated the performance metrics and then proceeded with the assessment of the gaze data.

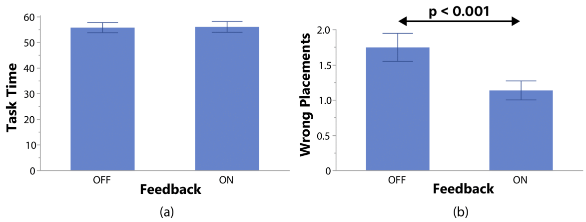


FIGURE 3. (a) The mean task time for each feedback condition, (b) the mean number of wrong placements

A. PERFORMANCE RESULTS

Fig. 3(a) shows the average time spent to finish the task for the two feedback conditions, while Fig.3 (b) shows the mean number of wrong placements for the feedback conditions.

Data for the task time had a normal distribution (S=0.695, K=0.661). RM ANOVA showed no significant difference in the task completion time between the two feedback conditions (F (1,35) = 0.25, p = 0.876, η^2 = 0.001). This means that the feedback did not have an effect on the time the participants took to finish the task (Fig. 3(a)).

Data for the number of wrong placements did not have a normal distribution after log-transform, so we performed an ART [26]. We found a significant difference in the number of wrong placements between the feedback conditions (F (1,35) = 14.674, p $<$ 0.001, η^2= 0.295). This significant difference shows that the early-warning feedback decreases the number of wrong placements (Fig. 3 (b)).

B. GAZE RESULTS

Data for Number of Fixations (S = 0.543, K = -0.844) and Time to First Fixation (S = 0.5, K = -1) and the were normally distributed. Fixation Duration (S = 0.83, K = -0.15) was normally distributed after log-transform.

We found a significant difference in the Number of Fixations (F (2,70) = 32.127, p $<$ 0.001, η^2 = 0.479) as shown in Fig. 4(a). The number of fixations on Low Intensity Light Level was significantly more than Out of Vision Hands (M = 0.669, SE = 0.085, p $<$ 0.001) and Self Occlusion (M = 0.213, SE = 0.096, p $<$ 0.05).

Moreover, the Number of Fixations for Self Occlusion Feedback was significantly higher than the Out of Vision Hands Feedback (M = 0.456, SE = 0.073, p $<$ 0.001). We found participants fixated the least on the Out of Vision Hands Feedback. Moreover, we did not observe significant

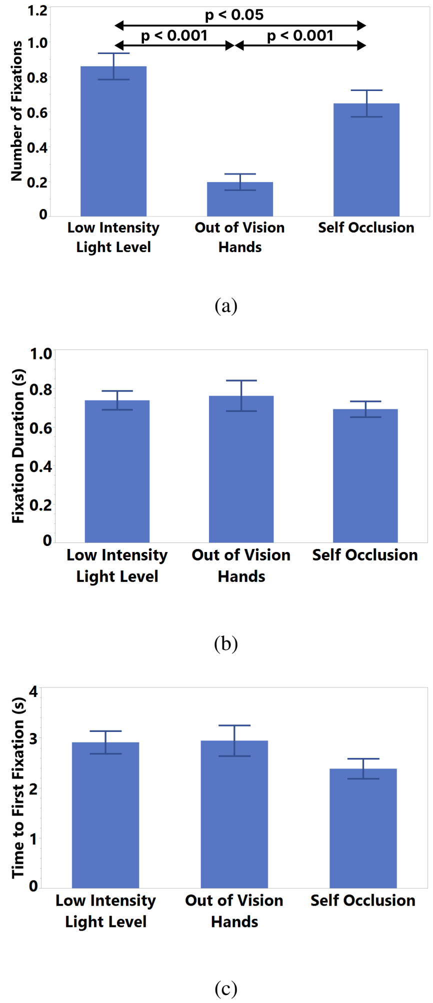


FIGURE 4. The comparison of (a) Number of Fixations, (b) Fixation Duration, and (c) Time to First Fixation for each simulated error. We found significant differences in the Number of Fixations between all simulated errors.

differences between simulated errors for Fixation Duration, shown in Fig. 4(b), and Time to First Fixation, shown in Fig. 4(c).

We also visualized participants' gaze data by calculating time periods of feedback, fixations, and saccades to better understand the cognitive change. An example gaze data visualization is provided in Fig. 6. In this analysis, we observed a decrease in the number of fixations as the condition repetition increased. Due to this pattern, we ordered each feedback in a single run and called it Feedback Repetition. The Fig. 5(a) demonstrates that the number of fixations is reduced with increasing feedback repetitions. Initially, the number of fixations was high, but it gradually declined and leveled off after around the 10th or 11th repetition, following an exponential function pattern ($y = x_0 e^{-\lambda x}$, R^2 = 0.961, and AIC = 25.27) as shown in Fig. 5 (a). Moreover, we calculated the Attention Percentage by dividing fixation durations for each object class by total fixation durations and compared

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3621645

them between Feedback ON and OFF conditions as shown in Fig. 5(b), and found a reduction in Attention Percentages by the following amounts; %4 from Environment, and %4 from Objects, leading to %8 Feedback Fixation Duration.

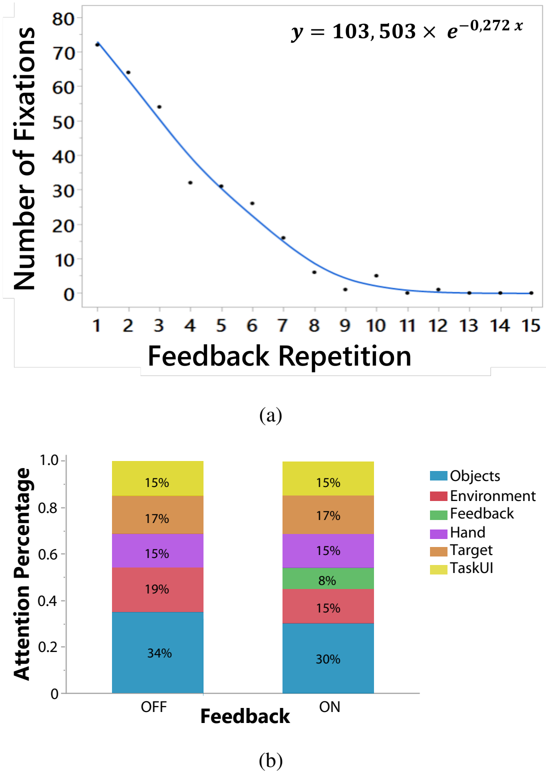


FIGURE 5. (a) The change in the number of fixations with successive feedback repetitions. Black dots represent data points, which are the number of fixations at a particular feedback repetition. (b) Attention Percentages between Feedback ON and OFF conditions. Each color represents a different class of objects.

We also calculated the average Fixation and Saccade on Feedback with Time to First Fixation/Saccade and Fixation/Saccade Duration. For the average Fixation on Feedback, we measured a Time to First Fixation of 2.695 s with a Fixation Duration of 0.72 s, as shown in Fig. 7(a). On the other hand, we found the average Time to First Saccade was 2.45 s with a Saccade Duration of 0.105 s, as shown in Fig. 7(b). We observed that the Time to First Saccade is earlier than the Time to First Fixation on the Feedback and has a shorter duration than Fixation.

VI. DISCUSSION

In this study, we investigated how early-warning visual feedback for common hand-tracking failures—Out of Vision Hands, Self-Occlusion, and Low-Intensity Light Level—influences user attention and contributes to improved interaction in VR environments. Our primary goal was to understand not only whether these visual feedback systems reduce errors, but also how they affect user attention and gaze behavior.

We developed a Fixation and Saccade recorder to capture gaze data and conducted a user study with 12 participants, who performed a pick-and-place task under different feedback conditions. The results confirmed previous findings [32], [35], showing that early-warning feedback significantly reduced the number of hand-tracking errors, without increasing task completion time. This supports the argument that early-warning systems enhance task performance without compromising efficiency or adding cognitive overload.

To understand the behavioral dynamics behind this improvement, we analyzed gaze patterns across the three error conditions and to answer our **RQ1** How do users' gaze movements change when early- warning visual feedback systems are activated? Our findings revealed clear differences in the number of fixations based on the type of error and the corresponding feedback.

Among the three hand tracking failure conditions, the Low-Intensity Light Level feedback received the highest number of fixations, suggesting that it was the most noticeable or attention-grabbing. In contrast, the Out of Vision Hands feedback had the fewest fixations, indicating that users either overlooked it more often or processed it more quickly. According to prior work [18], a high number of fixations generally reflects that a visual element is perceived as more important or demanding.

We also observed a decrease in fixations over time with repeated exposure to the same early-warning feedback, as shown in Fig. 5(a). This suggests a possible habituation effect, where users become familiar with the visual cue and require less visual engagement to interpret it. While this may lead to faster task execution, future research is needed to determine the extent and implications of this exponential decay in attention over longer sessions or more complex tasks.

We also assessed the accuracy and timing of the simulated failure conditions. Our analysis showed that the early-warning visual feedback system consistently triggered before hand-tracking failures occurred, effectively functioning as a notification mechanism. Moreover, the system occasionally issued feedback in response to potential tracking failures that did not result in hand tracking failures. This design aligns with the goal of warning without restriction, allowing users to maintain agency while still benefiting from system guidance.

Importantly, users were free to ignore these early warnings visual cues, and in some trials—particularly under Out-of-Vision and Self-Occlusion conditions—participants chose to continue their actions, which sometimes led to tracking loss. Nevertheless, we observed an overall reduction in wrong object placements in the feedback conditions, suggesting that participants adjusted their hand posture when they noticed a warning. These adaptations likely contributed to improved tracking reliability.

In cases where the hand tracking failure could not be avoided, such as under low-light conditions, the feedback was still beneficial. Participants appeared to mentally prepare for the loss of tracking. Observations by the experimenter confirmed that users often paused their movement or assumed a neutral position when feedback was triggered—signaling a shift in behavior to accommodate anticipated failure.

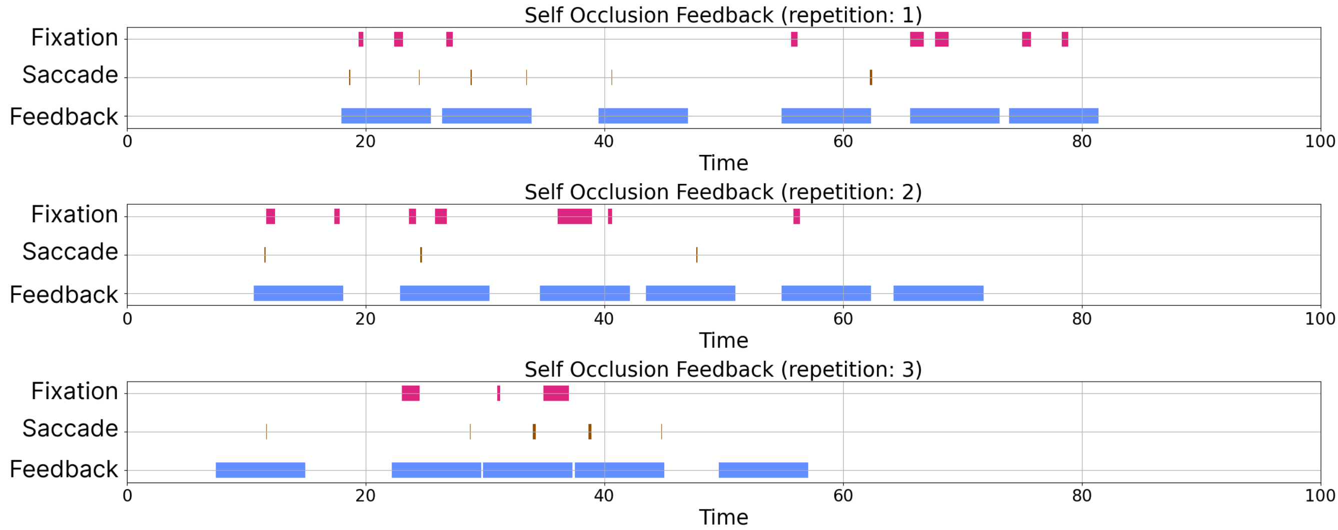
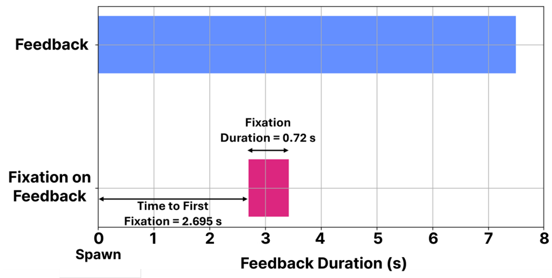
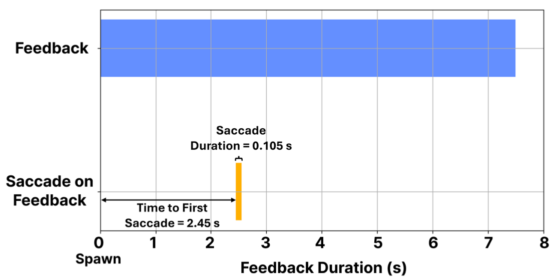


FIGURE 6. **Gaze data visualization of one of the participants. We observed a reduction in fixations as the number of repetitions increased.**



(a)



(b)

FIGURE 7. **(a) Average Fixation on Feedback. The blue bar represents the amount of time feedback was shown to the participant. The red bar represents the fixation. It took 2.695 s for users to fixate on the feedback on average, and once they fixated on it, the average fixation duration was 0.72 s. (b) Average Saccade on Feedback. The yellow bar represents the saccade. The time to first saccade was 2.45 s, and the saccade duration was 0.105 s on average.**

Fixation duration can reflect the information content or complexity of an area [12]. Additionally, difficulty on information extraction or being a more engaging object can result in longer fixation durations [19]. In terms of complexity and information extraction, we can say the early warning visual feedback for each simulated error is considered to be the same, even though the positioning of the Out of Vision Hands is different than the other two. Moreover, we found that the Attention Percentage for Target and TaskUI stayed the same after enabling the feedback, which indicates the user attention on the task was not affected by the feedback.

This finding answers our first **RQ1** and suggests that the early warning visual feedback mechanism was well-integrated into the interface, allowing users to seamlessly interpret the information without compromising their primary focus. By keeping their attention percentages consistent for the Target and TaskUI, users demonstrated that the early warning visual feedback did not impose additional cognitive load or distract from the core task. This potentially explains the reduced cognitive load and higher usability score results in early-warning feedback systems [32].

Under high cognitive load, users selectively direct their visual attention toward task-relevant information, potentially missing peripheral or secondary visual cues [36]. This selective attention can explain why certain visual feedback cues might be overlooked when cognitive demands are high [37]. As it can be seen in Fig. 6, Time to First Saccade is shorter than the Time to First Fixation in average, meaning that Saccades occurred earlier than the Fixations in general. Short saccade duration is expected due to the nature of the Saccades, which is 0.105 s for our feedback. The shorter Time to First Saccade suggests that users promptly directed their attention toward the feedback elements after they appeared. This rapid response indicates that the feedback cues were both noticeable and effective at capturing initial attention [38]. Frequent gaze shifts or interruptions can increase the cognitive load by causing users to mentally reset their focus each time their attention is diverted. This relationship suggests that notifications or feedback systems must be carefully designed to minimize unnecessary gaze shifts.

In contrast, the longer Time to First Fixation highlights the additional time required for users to stabilize their gaze and process the visual information provided by the visual feed-

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3621645

back. Fixations represent the moments when users actively focus and extract information, which inherently takes longer than the swift redirection of gaze achieved by saccades. This also answers our second **RQ2** In which context do early-warning systems help reduce users' cognitive load?

The thresholds used in this work were specifically tailored to the task environment and the capabilities of the VR HMD. For example, the values used to detect potential hand-tracking failures — particularly in the out-of-vision condition —–were designed based on the interaction area in our task. These values were selected by considering the height of the virtual box and how far participants needed to move their hands upward, which in turn was constrained by the tracking FoV of the headset. Because of this, the exact threshold values we used may not directly apply to other VR systems. For instance, a headset with a wider tracking FoV might not trigger the same warning conditions and would require recalibrated thresholds.

However, the main goal of this study was not to fine-tune the system for a specific hardware configuration or task, but rather to demonstrate the general effectiveness of early warning visual feedback systems and explain the result of previous work on early warning systems [35]. Future work can adapt and refine these threshold values to suit different HMDs or application scenarios, depending on the interaction space and system limitations.

In this study, we did not include SUS (System Usability Scale) and NASA-TLX (Task Load Index) measurements. The decision was intentional. Previous work [32], [35] has already consistently shown that early-warning visual feedback reduces user frustration and improves task performance — for example, by lowering the number of incorrectly placed objects. Therefore, rather than replicating those usability outcomes, our aim in this paper was to take a step further and focus specifically on gaze movements to better understand how and why usability and cognitive load improve when early-warning feedback is provided. Yet, we also acknowledge that the analyzed gaze metrics here are *indirect* indicators of attentional allocation, which correlate with cognitive effort in VR and HCI studies [60]–[62], but we do not claim they offer conclusive evidence of cognitive load in isolation.

Nonetheless, we did collect data on task errors, and the reduction in the number of wrong placements in our results aligns with previous findings [32], [35]. This further supports that our implementation of early-warning visual feedback is consistent in terms of performance outcomes, while also providing new insights into user behavior through gaze data.

We conducted this study with 12 participants, and this number was determined using G*Power analysis to ensure that we could detect a high effect size with sufficient statistical power. Since the goal of the study was to analyze why early-warning feedback helps reduce cognitive load—through gaze behavior, we prioritized collecting detailed, high-quality gaze data over a large sample size. Once we began to observe stable and consistent behavioral patterns in the data, additional participants were not necessary. Our results showed clear trends that allowed us to draw meaningful conclusions about how early-warning feedback supports attention management and reduces mental effort during interaction.

The fixation and saccade thresholds used in this study were selected based on prior research and insights from pilot studies. As noted in the literature, fixation and saccade thresholds often vary depending on the task type and experimental environment [46]. For example, Arslan et al. [41] used a 0.5-second threshold for fixations, while Król and Król [42] applied a threshold of 0.085 seconds. Jansen et al. [44] used 0.3 seconds, Manor and Gordon [43] used 0.2 seconds, and Llanes et al. [45] tested a range between 0.2 and 0.4 seconds, ultimately recommending a 0.25-second threshold.

In line with Llanes et al. [45], we also tested a range of fixation thresholds in our pilot phase and selected the one most suitable for our specific task. We recommend that future studies similarly evaluate a range of thresholds to determine the most appropriate values for their task and environment.

Our findings highlight the potential of early-warning visual feedback to improve user experience in VR without disrupting task performance. Specifically, we show that visual feedback can:

- Effectively warn users before failures occur
- Guide attention without imposing cognitive overload
- Adapt to user behavior over time (habituation)
- Be integrated seamlessly into visually demanding tasks

In this paper, we did not aim to generalize gaze patterns across all VR tasks or environments, but rather to explore how gaze behavior changes within a controlled, repeated, pick-and-place task in response to early-warning feedback for hand-tracking failures. Unlike studies on general vision science theory, such as [56], [58], [59], [74], our findings here are contextualized to the task structure and feedback modality. We recommend future studies to explore the gaze behavior in early warning systems with different tasks and environments, in AR and MR settings, and different hand tracking failures, and different feedback modalities, before generalizing our findings to the vision science theory.

While this paper is built upon prior early-warning feedback systems [32], [35], our work represents a substantive extension that addresses a critical and previously unexplored gap in the literature: understanding **why** these systems improve user performance, not just **whether** they do. This distinction is important for several reasons in the XR community. First, by analyzing how visual feedback interacts with gaze behavior, we uncover attentional mechanisms (e.g., time to first fixation, attention allocation, and habituation) that allow early-warning systems to reduce cognitive effort. Understanding these mechanisms can enable researchers and developers to design task-specific early warning systems across different contexts and applications, such as surgical VR, training simulators, and collaborative environments. Second, rather than treating visual feedback as a black box, our study reveals **how** users adapt to notifications over time, reduce unnecessary fixations, and selectively attend to relevant stimuli under error-prone conditions. This deeper analysis advances the theoretical understanding of visual attention and adaptation in an immersive environment. Also, by modeling how and when users respond to early-warning cues, our findings offer practical design guidance for future notification systems. Novel empirical data refines how we understand user behavior in immersive systems is a foundational goal in HCI and XR research.

Compared to recent gaze-aware feedback systems, which often rely on real-time gaze input to dynamically alter the interface or control information flow, our work does not treat gaze as a control signal [71], [75]. Prior work has mainly focused on gaze-adaptive interfaces (e.g., for information filtering or attention redirection), where user gaze actively drives system behavior [70], [73]. These systems are typically designed to improve task efficiency or support adaptive personalization. Instead, we treat gaze as an observational metric to analyze users' attentional engagement with system-triggered cues.

These results have practical implications for VR developers aiming to design early-warning notification systems that are both non-intrusive and highly functional. Future research should explore the integration of other modalities, such as haptic or auditory feedback, and compare their effectiveness across different user groups and task types. Additionally, investigating long-term habituation effects and personalization of feedback intensity could further enhance system adaptability.

While our findings offer novel insights into attentional dynamics under early-warning visual feedback, we recognize that gaze metrics alone cannot definitively distinguish cognitive load from visual salience. Future work should incorporate multi-modal workload assessment frameworks, combining gaze analysis with physiological signals such as EEG or fNIRS, as well as subjective measures like NASA-TLX. These additional signals would allow researchers to triangulate the user's mental state, helping determine whether gaze shifts reflect attentional prioritization, salience-driven responses, or effortful processing under cognitive strain. By integrating these modalities in a unified experimental framework, future studies could not only validate our gaze-based findings but also enable the development of adaptive, real-time workload-aware feedback systems for immersive environments. This would also explain some of the limitations of our study here. Moreover, such strategies would help researchers and developers to design applications with higher usability.

VII. CONCLUSION

In this paper, we examined participants' perception of our early-warning visual feedback that was shown for common hand tracking failures by analyzing the gaze data based on fixation-based metrics. We found that user attention to feedback changes based on the simulated error, in a way that feedback for errors that are likely to happen gets more attention from the user. Additionally, showing feedback before a hand tracking error occurs does not significantly lead to a disadvantage in task execution.

We hope that this work can lead to more efficient visual feedback systems, not limited to hand tracking failures, but can even be used in generic feedback systems where user perception is important. The study is limited to gaze analysis of visual feedback only, therefore future work should focus on the effect of haptic, auditory, and a combination of different types of feedback systems on user gaze behavior.

REFERENCES

[1] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis: Pearson New International Edition PDF eBook*. London, U.K.: Pearson Higher Ed, 2013.

[2] P. Mallery and D. George, *SPSS for Windows Step by Step*. Boston, MA, USA: Allyn & Bacon, Inc., 2000.

[3] G. Casiez, N. Roussel, and D. Vogel, "1 € filter: A simple speed-based low-pass filter for noisy input in interactive systems," in *Proc. SIGCHI Conf. Human Factors Comput. Syst. (CHI '12)*, Austin, TX, USA, 2012, pp. 2527–2530, doi: 10.1145/2207676.2208639.

[4] S. A. M. Faleel, S. Kwon, D. Ahlström, and P. Irani, "Validating eyes-free affordance of on-finger hand proximate user interfaces in in-situ scenarios," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Singapore, 2024, pp. 1283–1292, doi: 10.1109/ISMAR62088.2024.00145.

[5] A. Poole and L. Ball, "Eye tracking in human-computer interaction and usability research: Current status and future prospects," in *Encyclopedia of Human Computer Interaction*, 2006, pp. 211–219.

[6] R. Andersson, L. Larsson, K. Holmqvist, M. Stridh, and M. Nyström, "One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms," *Behavior Res. Methods*, vol. 49, pp. 616–637, 2017.

[7] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proc. Symp. Eye Tracking Res. Appl. (ETRA)*, Palm Beach Gardens, FL, USA, 2000, pp. 71–78.

[8] N. Sendhilnathan, T. Zhang, B. Lafreniere, T. Grossman, and T. R. Jonker, "Detecting input recognition errors and user errors using gaze dynamics in virtual reality," in *Proc. ACM Symp. User Interface Softw. Technol. (UIST '22)*, Bend, OR, USA, 2022, Art. no. 38, 19 pp., doi: 10.1145/3526113.3545628.

[9] M. Wimmer, N. Weidinger, N. ElSayed, G. R. Müller-Putz, and E. Veas, "EEG-based error detection can challenge human reaction time in a VR navigation task," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Singapore, 2023, pp. 970–979, doi: 10.1109/ISMAR59233.2023.00113.

[10] A. U. Batmaz and W. Stuerzlinger, "The effect of pitch in auditory error feedback for Fitts' tasks in virtual reality training systems," in *Proc. IEEE Virtual Reality 3D User Interfaces (VR)*, Lisbon, Portugal, 2021, pp. 85–94, doi: 10.1109/VR50410.2021.00029.

[11] J. Wolf, Q. Lohmeyer, C. Holz, and M. Meboldt, "Gaze comes in handy: Predicting and preventing erroneous hand actions in AR-supported manual tasks," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Bari, Italy, 2021, pp. 166–175, doi: 10.1109/ISMAR52148.2021.00031.

[12] Z. Bylinskii and M. Borkin, "Eye fixation metrics for large scale analysis of information visualizations. ETVIS Work," *Eye Track. Vis.*, 2015.

[13] P. Blignaut, "Fixation identification: The optimum threshold for a dispersion algorithm," *Attention, Perception, & Psychophys.*, vol. 71, pp. 881–895, 2009.

[14] G. Hauland, "Measuring team situation awareness by means of eye movement data," in *Human-Centered Computing*, Boca Raton, FL, USA: CRC Press, 2019, pp. 230–234.

[15] W. Graf and H. Krueger, "Ergonomic evaluation of user-interfaces by means of eye-movement data," in *Proc. 3rd Int. Conf. Human-Computer Interaction, Vol. 1 on Work with Computers: Organizational, Management, Stress and Health Aspects*, Boston, MA, USA, 1989, pp. 659–665.

[16] M. Borys and M. Plechawska-Wójcik, "Eye-tracking metrics in perception and visual attention research," *EJMT*, vol. 3, pp. 11–23, 2017.

[17] S. Wei, D. Bloemers, and A. Rovira, "A preliminary study of the eye tracker in the Meta Quest Pro," in *Proc. ACM Int. Conf. Interactive Media Experiences*, 2023, pp. 216–221.

[18] A. Poole, L. J. Ball, and P. Phillips, "In search of salience: A response-time and eye-movement analysis of bookmark recognition," in *People and*

Computers XVIII—Design for Life, London, U.K.: Springer, 2005, pp. 363–378.

[19] M. A. Just and P. A. Carpenter, "Eye fixations and cognitive processes," *Cognitive Psychol.*, vol. 8, no. 4, pp. 441–480, 1976, doi: 10.1016/0010-0285(76)90015-3.

[20] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, "G*Power 3: A flexible statistical power analysis program for the social, behavior, and biomedical sciences," *Behavior Res. Methods*, vol. 39, pp. 175–191, May 2007, doi: 10.3758/BF03193146.

[21] G. Buckingham, "Hand tracking for immersive virtual reality: Opportunities and challenges," *Frontiers in Virtual Reality*, vol. 2, Oct. 2021, doi: 10.3389/frvir.2021.728461.

[22] N. Jamalian, M. Gillies, F. F. Leymarie, and X. Pan, "The effects of hand tracking on user performance: An experimental study of an object selection based memory game," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Singapore, 2022, pp. 768–776, doi: 10.1109/ISMAR55827.2022.00095.

[23] A. Schäfer, G. Reis, and D. Stricker, "Controlling teleportation-based locomotion in virtual reality with hand gestures: A comparative evaluation of two-handed and one-handed techniques," *Electronics*, vol. 10, no. 6, p. 715, Mar. 2021, doi: 10.3390/electronics10060715.

[24] Y. Ferstl, R. McDonnell, and M. Neff, "Evaluating study design and strategies for mitigating the impact of hand tracking loss," in *Proc. ACM Symp. Applied Perception (SAP '21)*, Virtual Event, France, 2021, Art. no. 3, 12 pp., doi: 10.1145/3474451.3476235.

[25] R. McGloin, K. Farrar, and M. Krcmar, "Video games, immersion, and cognitive aggression: Does the controller matter?," *Media Psychol.*, vol. 16, no. 1, pp. 65–87, 2013, doi: 10.1080/15213269.2012.752428.

[26] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures," in *Proc. SIGCHI Conf. Human Factors Comput. Syst. (CHI '11)*, Vancouver, BC, Canada, 2011, pp. 143–146, doi: 10.1145/1978942.1978963.

[27] B. Stenger, A. Thayananthan, P. H. S. Torr, and R. Cipolla, "Model-based hand tracking using a hierarchical Bayesian filter," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1372–1384, 2006, doi: 10.1109/TPAMI.2006.189.

[28] F. Mueller, D. Mehta, O. Sotnychenko, S. Sridhar, D. Casas, and C. Theobalt, "Real-time hand tracking under occlusion from an egocentric RGB-D sensor," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017.

[29] A. Kowdle *et al.*, "The need 4 speed in real-time dense visual tracking," *ACM Trans. Graph.*, vol. 37, no. 6, Art. no. 220, Dec. 2018, doi: 10.1145/3272127.3275062.

[30] W. Lu, H. He, A. Urban, and J. Griffin, "What the eyes can tell: Analyzing visual attention with an educational video game," in *Proc. ACM Symp. Eye Tracking Res. Appl. (ETRA)*, 2021, pp. 1–7.

[31] L. Sidenmark, C. Clarke, J. Newn, M. N. Lystbæk, K. Pfeuffer, and H. Gellersen, "Vergence matching: Inferring attention to objects in 3D environments for gaze-assisted selection," in *Proc. CHI Conf. Human Factors Comput. Syst. (CHI '23)*, Hamburg, Germany, 2023, Art. no. 257, 15 pp., doi: 10.1145/3544548.3580685.

[32] M. R. Bashar and A. U. Batmaz, "An early warning system based on visual feedback for light-based hand tracking failures in VR head-mounted displays," *IEEE Trans. Vis. Comput. Graph.*, 2025.

[33] T. Galais, A. Delmas, and R. Alonso, "Natural interaction in virtual reality: Impact on the cognitive load," in *Adjunct Proc. 31st Conf. l'Interaction Homme-Machine*, 2019, pp. 1–9.

[34] P. Monteiro, G. Gonçalves, H. Coelho, M. Melo, and M. Bessa, "Hands-free interaction in immersive virtual reality: A systematic review," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 5, pp. 2702–2713, 2021.

[35] M. Gemici, V. Phantakis, A. U. Batmaz, "Before hands disappear: Effect of early warning visual feedback method for hand tracking failures in virtual reality," *PlosOne*, 20(6): e0323796. https://doi.org/10.1371/journal.pone.0323796.

[36] N. Lavie, A. Hirst, J. W. De Fockert, and E. Viding, "Load theory of selective attention and cognitive control," *J. Exp. Psychol.: General*, vol. 133, no. 3, p. 339, 2004.

[37] C. Stevens and D. Bavelier, "The role of selective attention on academic foundations: A cognitive neuroscience perspective," *Develop. Cogn. Neurosci.*, vol. 2, pp. S30–S48, 2012.

[38] D. Chen and R. Vertegaal, "Using mental load for managing interruptions in physiologically attentive user interfaces," in *CHI'04 Extended Abstr. Human Factors Comput. Syst.*, 2004, pp. 1513–1516.

[39] G. Molina, J. Gimeno, C. Portalés, and S. Casas, "A comparative analysis of two immersive virtual reality systems in the integration and visualization of natural hand interaction," *Multimedia Tools Appl.*, vol. 81, no. 6, pp. 7733–7758, 2022.

[40] B. Birawo and P. Kasprowski, "Review and evaluation of eye movement event detection algorithms," *Sensors*, vol. 22, no. 22, p. 8810, Nov. 2022, doi: 10.3390/s22228810.

[41] D. B. Arslan, M. Sükuti, and A. D. Duru, "The identification of individualized eye tracking metrics in VR using data driven iterative-adaptive algorithm," *AJIT-e: Academic J. Inf. Technol.*, vol. 14, no. 52, pp. 8–21, 2023.

[42] M. Król and M. E. Król, "Simple eye movement metrics can predict future decision making performance: The case of financial choices," *Judgm. Decis. Making*, vol. 14, no. 3, pp. 223–233, 2019.

[43] B. R. Manor and E. Gordon, "Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks," *J. Neurosci. Methods*, vol. 128, nos. 1–2, pp. 85–93, 2003.

[44] A. Jansen, C. Nederkoorn, and S. Mulkens, "Selective visual attention for ugly and beautiful body parts in eating disorders," *Behav. Res. Therapy*, vol. 43, no. 2, pp. 183–196, 2005.

[45] J. Llanes-Jurado, J. Marín-Morales, J. Guixeres, and M. Alcañiz, "Development and calibration of an eye-tracking fixation identification algorithm for immersive virtual reality," *Sensors*, vol. 20, no. 17, p. 4956, 2020.

[46] J. L. Orquin and K. Holmqvist, "Threats to the validity of eye-movement research in psychology," *Behavior Res. Methods*, vol. 50, pp. 1645–1656, 2018.

[47] L. Perkhofer and O. Lehner, "Using gaze behavior to measure cognitive load," in *Information Systems and Neuroscience: NeuroIS Retreat 2018*, Cham, Switzerland: Springer, 2018, pp. 73–83.

[48] J. Moon and J. Ryu, "The effects of social and cognitive cues on learning comprehension, eye-gaze pattern, and cognitive load in video instruction," *J. Comput. Higher Educ.*, vol. 33, no. 1, pp. 39–63, 2021.

[49] B. Steichen, C. Conati, and G. Carenini, "Inferring visualization task properties, user performance, and user cognitive abilities from eye gaze data," *ACM Trans. Interact. Intell. Syst.*, vol. 4, no. 2, pp. 1–29, 2014.

[50] 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 *Proc. CHI Conf. Human Factors Comput. Syst.*, 2017, pp. 1118–1130.

[51] J.-C. Liu, K.-A. Li, S.-L. Yeh, and S.-Y. Chien, "Assessing perceptual load and cognitive load by fixation-related information of eye movements," *Sensors*, vol. 22, no. 3, p. 1187, 2022.

[52] F. Pointecker, D. Oberögger, and C. Anthes, "Visual metaphors for notification into virtual environments," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct (ISMAR-Adjunct)*, 2023, pp. 60–64.

[53] L. Plabst, S. Oberdörfe, F. Ortega, and F. Niebling, "Press the red button: A user study comparing notification placement with augmented and non-augmented tasks in AR," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct (ISMAR-Adjunct)*, 2022, pp. 483–484.

[54] Y. Wang, Z. Hu, S. Yao, and H. Liu, "Using visual feedback to improve hand movement accuracy in confined-occluded spaces in virtual reality," *Vis. Comput.*, vol. 39, pp. 1485–1501, Feb. 2022, doi: 10.1007/s00371-022-02424-2.

[55] S. Ghosh *et al.*, "Notifivr: Exploring interruptions and notifications in virtual reality," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 4, pp. 1447–1456, 2018.

[56] C. A. Rothkopf, D. H. Ballard, and M. M. Hayhoe, "Task and context determine where you look," *J. Vis.*, vol. 7, no. 14, pp. 16–16, 2007.

[57] E. Kowler, "Eye movements: The past 25 years," *Vis. Res.*, vol. 51, no. 13, pp. 1457–1483, 2011.

[58] W. W. Sprague, E. A. Cooper, I. Tošić, and M. S. Banks, "Stereopsis is adaptive for the natural environment," *Sci. Adv.*, vol. 1, no. 4, p. e1400254, 2015.

[59] A. M. Aizenman, G. A. Koulieris, A. Gibaldi, V. Sehgal, D. M. Levi, and M. S. Banks, "The statistics of eye movements and binocular disparities during VR gaming: Implications for headset design," *ACM Trans. Graph.*, vol. 42, no. 1, pp. 1–15, 2023.

[60] A. W. Joseph and R. Murugesh, "Potential eye tracking metrics and indicators to measure cognitive load in human-computer interaction research," *J. Sci. Res.*, vol. 64, no. 1, pp. 168–175, 2020.

[61] A. Kar and P. Corcoran, "Performance evaluation strategies for eye gaze estimation systems with quantitative metrics and visualizations," *Sensors*, vol. 18, no. 9, p. 3151, 2018.

[62] V. Skaramagkas *et al.*, "Review of eye tracking metrics involved in emotional and cognitive processes," *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 260–277, 2021.

[63] B. Lange, A. Rizzo, C.-Y. Chang, E. A. Suma, and M. Bolas, "Markerless full body tracking: Depth-sensing technology within virtual environments," in *Proc. Interservice/Industry Training, Simulation, and Education Conf. (I/ITSEC)*, 2011.

[64] E. Foxlin and M. Harrington, "Weartrack: A self-referenced head and hand tracker for wearable computers and portable VR," in *Proc. 4th Int. Symp. Wearable Comput.*, 2000, pp. 155–162.

[65] F. Mueller, F. Bernard, O. Sotnychenko, D. Mehta, S. Sridhar, D. Casas, and C. Theobalt, "Ganerated hands for real-time 3D hand tracking from monocular RGB," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 49–59.

[66] Y. Lei, Y. Deng, L. Dong, X. Li, X. Li, and Z. Su, "A novel sensor fusion approach for precise hand tracking in virtual reality-based human–computer interaction," *Biomimetics*, vol. 8, no. 3, p. 326, 2023.

[67] Z. Pan, Y. Li, M. Zhang, C. Sun, K. Guo, X. Tang, and S. Z. Zhou, "A real-time multi-cue hand tracking algorithm based on computer vision," in *Proc. IEEE Virtual Reality Conf. (VR)*, 2010, pp. 219–222.

[68] Y. Zhang, A. Nowak, Y. Xuan, A. Romanowski, and M. Fjeld, "See or hear? Exploring the effect of visual/audio hints and gaze-assisted instant post-task feedback for visual search tasks in AR," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, 2023, pp. 1113–1122, doi: 10.1109/ISMAR59233.2023.00128.

[69] R. Jianu, N. Silva, N. Rodrigues, T. Blascheck, T. Schreck, and D. Weiskopf, "Gaze-aware visualisation: Design considerations and research agenda," in *Comput. Graph. Forum*, 2025, p. e70097.

[70] A. L. Gardony, K. Okano, G. I. Hughes, A. J. Kim, K. T. Renshaw, A. Sipolins, A. B. Whitig, F. Lu, and D. A. Bowman, "Characterizing information access needs in gaze-adaptive augmented reality interfaces: Implications for fast-paced and dynamic usage contexts," *Hum.–Comput. Interact.*, vol. 39, nos. 5–6, pp. 553–583, 2024.

[71] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen, "Gaze + pinch interaction in virtual reality," in *Proc. ACM Symp. User Interface Softw. Technol.*, 2017.

[72] A. K. Mutasim, W. Stuerzlinger, and A. U. Batmaz, "Gaze tracking for eye-hand coordination training systems in virtual reality," in *Proc. ACM Symp. User Interface Softw. Technol.*, 2020.

[73] P. Renner and T. Pfeiffer, "Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*, 2017, pp. 186–194.

[74] E. Kowler, "Eye movements: The past 25 years," *Vision Res.*, vol. 51, no. 13, pp. 1457–1483, 2011.

[75] A. K. Mutasim, A. U. Batmaz, M. H. Mughrabi, and W. Stuerzlinger, "The influence of eye gaze interaction technique expertise and the guided evaluation method on text entry performance evaluations," *ACM Trans. Comput.-Hum. Interact.*, vol. 9, no. 3, May 2025.

MUCAHIT GEMICI is a Ph.D. student in Computer Science at Concordia University, specializing in Virtual Reality, Human-Computer Interaction (HCI), and immersive systems. His research explores precision, feedback, and interaction techniques in VR, with publications at leading venues such as IEEE ISMAR, IEEE VR, and PLOS ONE. He holds a Bachelor's degree in Electrical and Electronics Engineering from Kadir Has University, where he graduated as the top-ranked student. Beyond academia, Mücahit has several years of experience in game development, including mobile and brain-computer interface games, combining his technical expertise with a strong focus on user experience and interaction design.

AMAL HATIRA received her B.Sc. degree in electrical and electronics engineering from the European University of Lefke, in 2022, and her M.Sc. degree in electronics engineering from Kadir Has University, Turkey, in 2024. She is currently pursuing the Ph.D. degree in computer science at Concordia University, Montreal, QC, Canada. She is a Research Assistant at EXIT lab, Concordia University, focusing on human-computer interaction, haptic feedback, and training systems in virtual reality environments. She is a recipient of the FRQNT Doctoral Research Award.

VRUSHANK PHADNIS is a UX Researcher at Google Inc., specializing in human-computer interaction (HCI) for augmented and virtual reality (AR/VR). He earned his Ph.D. from the Massachusetts Institute of Technology (MIT). His research focuses on input interactions, visual comfort, perception performance, and hardware concept testing. He has authored and refereed papers for top conferences and journals, including ACM CHI, IEEE VR, IEEE ISMAR, ACM UIST, and ASME JMD.

ANIL UFUK BATMAZ received his PhD from Strasbourg University. He is currently working as an assistant professor at Concordia University, Gina Cody School of Engineering and Computer Science, in Montreal, Quebec, Canada. Dr. Batmaz leads a research lab titled "EXtended reality and Interaction Technologies" (EXIT) lab. His research is focused on Virtual Reality, Augmented Reality, Training Systems and Simulators, and Human Computer Interaction.