Optimizing Moving Target Selection in VR by Integrating Proximity-Based Feedback Types and Modalities

Xuning Hu
Xi'an Jiaotong-Liverpool University
Suzhou, China
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China

Wenxuan Xu
Dartmouth College
Hanover, USA
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China

Yushi Wei
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China

Hao Zhang
Institute of Software Chinese Academy of Sciences
Beijing, China

Jin Huang
Institute of Software Chinese Academy of Sciences
Beijing, China

Hai-Ning Liang *
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China

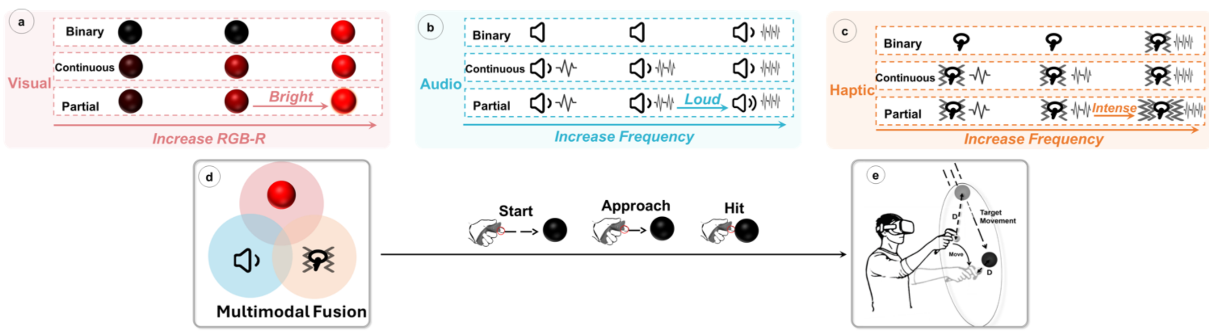

Figure 1: (a-c) depict the developed auxiliary feedback mechanisms, categorized into visual, auditory, and haptic feedback modalities, each incorporating three types: Binary, Continuous, and Partial, resulting in nine feedback mechanisms. For example, in the visual feedback context, the Binary type indicates that the target turns red when the controller hits it. The Continuous type shows a gradual color change to red as the controller approaches the target, while the Partial type combines this gradual color change with a brightening effect upon hitting the target (see Fig. 2 and Sec. 3). (d) shows the modality fusion used in Study 2, and (e) demonstrates our experimental setup.

Abstract

Proximity-based feedback provides users with real-time guidance as they approach an interaction goal. This type of feedback is particularly useful for tasks that require guidance during the interaction process, such as selecting moving targets. This work explores proximity-based feedback types and modalities to improve the selection of moving targets in VR by leveraging three feedback types that combine visual, auditory, and haptic modalities. We evaluated the performance of these mechanisms through two user studies, analyzing both objective data (e.g., selection time, error rate) and subjective data (e.g., user experience, preferences) to explore the characteristics of feedback types across different modalities and to examine the roles of various modalities within multimodal combinations. Our findings suggest optimal selection mechanisms for developers and should be tailored to different goals: achieving user precision, enabling quick movement to a target, considering task duration, and enhancing entertainment value. We also discuss applications that correspond to these different perspectives.

Index Terms: Virtual Reality, Moving target selection, Multimodal interaction and perception, Feedback Mechanism

*Corresponding author (e-mail: hainingliang@hkust-gz.edu.cn).
This work was supported in part by the National Natural Science Foundation of China (#62332017, #62172397). This research was reviewed and approved by the University Ethics Committee of Xi'an Jiaotong-Liverpool University (ER-LRR-1288940720240528160151).

1 Introduction

As virtual reality (VR) technology advances, there is a need for interfaces that can support the efficient performance of complex tasks. One such task is selecting moving targets, which is a key component in various interactive systems [57, 51]. It is relevant to a wide range of domains, such as interactive videos [39] and game development [17], as well as emerging applications like virtual training [49] and data visualization [24]. However, selecting moving targets presents unique challenges to users, especially in VR. Users must continuously track the target's motion. This involves accurately analyzing its trajectory, speed, and, more relevant to VR, depth while precisely controlling an input device to achieve selection. This requires users to have a high level of cognitive ability and operational control [23]. Therefore, selecting moving targets in VR can be challenging. Improving the efficiency and accuracy of this process can enhance user performance and experience while offering more degrees of freedom, allowing users to tailor task difficulty to their individual preferences or skill levels.

Previous works have proposed various techniques to assist users in this task [29]. For example, the Comet technique [31] adds an additional selectable area at the tail of the moving target based on its speed, the Sticky Cursor technique [38] automatically follows the moving target when the cursor contacts the target, and the Ghost technique [31] provides a static proxy of the moving target. While supportive, some techniques often have a high learning curve, while others add additional visual elements, making the environment cluttered and, in turn, increasing the difficulty of distinguishing between targets. Additionally, these techniques are highly sensitive to specific parameters and scenarios, as they alter the interface layout [55].

Given the issues of these techniques, alternative approaches, such as feedback mechanisms, have been proposed to enhance user experience and task efficiency. One of the most intuitive methods to achieve this is via different sensory feedback, including visual, auditory, or tactile (and their combination). Their use could be in the different phases when selecting a target. For instance, when a mouse pointer enters the boundary of a target, users may receive signals such as visual highlights or vibrations, confirming that the pointer is correctly positioned and the object is selectable [12]. The advantages of feedback mechanisms lie in their low learning cost, minimal impact on interface layout, and independence from specific parameter constraints. However, compared to other assistive techniques, such as enlarging selectable areas, feedback mechanisms typically offer weaker assistive effects and, at times, have limited impact on improving user performance, posing challenges in practical applications [46]. To balance low learning costs, maintain a clean interface, and enhance user assistance, various improvements to traditional feedback mechanisms have been widely explored [4]. Particularly relevant to our work is *proximity-based feedback* [6, 25], which adjusts the intensity of feedback based on the distance between the intended target and the pointing device used. Proximity feedback includes *binary* feedback (providing feedback when the target is reached) and *continuous* feedback (gradually increasing feedback intensity as the user approaches the target).

Previous work using proximity-based feedback has primarily focused on selecting stationary targets or scenarios without visual assistance [6, 26]. In the context of moving targets, proximity-based feedback types may yield different outcomes, as individual users perceive the speed and direction of targets independently when attempting to point to moving targets [15]. To our knowledge, there is limited research on multimodal feedback mechanisms designed explicitly for selecting moving targets in VR. Given the potential and unique affordances of feedback mechanisms as intuitive and complementary input approaches, it is necessary to deepen our understanding of different feedback types based on proximity (such as binary, continuous, and integrated binary-continuous) and their effects on users' ability to select moving targets across various modalities, including visual, auditory, and tactile, providing deeper insights into the design of VR systems.

In this paper, we conducted two experiments to explore the effectiveness of different feedback types in assisting users' selection of moving targets across various modalities, including visual, audio, and haptic. We first developed binary and continuous feedback types based on the optimal parameters for users' perception of target direction, and then introduced a novel feedback type that integrates elements of both, which we refer to as *'partial'*. These three feedback types and three modalities led to nine feedback mechanisms (see Fig. 1). We then conducted a user study in which we compared the nine feedback mechanisms based on objective data (time and error rate) and subjective data (empirical scales and rankings), using target size and movement speed as task conditions. The results revealed the characteristics of each feedback type across different modalities (see Sec. 7). We found that (1) binary feedback effectively assists users in determining the selection time, (2) continuous feedback enables users to locate targets more quickly, and (3) partial feedback under the visual modality combines the advantages of both binary and continuous feedback. However, in auditory and haptic modalities, partial feedback results in redundant details because multiple types are used within a single channel. Previous studies on static target selection show that continuous feedback interferes with selection tasks because it requires more time for calibration despite its ability to locate targets quickly [6, 7]. In contrast, for selecting moving targets, continuous feedback significantly outperforms no feedback in both time and error rate because it supports locating targets quickly. Based on the study results, we then selected the most optimal feedback type for each modality and carried out a second experiment to investigate the effects of their combinations (see Sec. 10.1). The second study's findings indicate that while auditory feedback does not significantly enhance performance, it plays a crucial role in enhancing user experience, especially since it is more entertaining to use. Unlike selecting stationary targets, visual feedback remains the most important mechanism to enhance performance and reduce errors.

In short, the main contributions of this paper include: **(1)** Development and exploration of nine feedback mechanisms by combining visual, auditory, and haptic modalities with three proximity-based feedback types (binary, continuous, and partial) to optimize moving target selection in VR (see Study 1 in Sec. 5). **(2)** An exploration of the properties of each modality through various combinations, revealing the unique role of each in the process of selecting moving objects (see Study 2 in Sec. 8). **(3)** A set of design recommendations for when to use the various feedback mechanisms based on the task at hand (see Sec. 10.2) and some example scenarios.

2 Related Work

In this section, we first introduce common techniques for selecting moving targets. Then, we explore the application of multimodal feedback in selection tasks, focusing on proximity-based feedback types and their variations. We further discuss the optimization of proximity-based feedback types in VR environments.

2.1 Moving Target Selection

Previous studies have extensively investigated moving targets to enhance performance and user experience using various input methods and modalities. These input methods include, but are not limited to, styluses [34] and mice [35], while the modalities encompass the use of hands [36] and eye gaze [32]. In addition, previous work has attempted to understand the difficulty and uncertainty of selecting moving targets by establishing models to predict selection time and error rate. Hoffmann [33] introduced a steady-state position error model and developed a model for predicting movement time (MT) when selecting moving targets, revealing the relationship between target speed and selection time and demonstrating that increased target speed significantly prolongs selection time. From the perspective of predicting error rates, Huang et al. modeled the endpoint distribution of moving targets and established a trivariate Gaussian model for selecting 1D and 2D moving targets [35]. They further demonstrated that increasing the speed of a target significantly affects selection error rate.

Techniques for improving user efficiency in selecting moving targets are broadly divided into two main approaches: (1) reducing effective pointing distance and (2) using static proxies. Examples from the first group include techniques such as Area Cursors [65], which increase the size of the cursor, and Bubble Cursors [27], which dynamically change the cursor size based on the distance to the nearest target. VTE [28] also belongs to this group, as it transforms space into a Voronoi diagram, effectively increasing the selectable area of all targets until they fill the entire space. Likewise, DynaSpot adjusts the cursor size based on its movement speed. These techniques enhance selection performance by reducing the distance between the cursor and the target or expanding the target's selectable area. Hasan et al. proposed a method that increases the selection area by calculating the user's selection area based on the trajectory and speed of the moving target, similar to a comet's trail [31].

For enhancing static management, Hasan et al. proposed leaving a static "ghost" of the target based on its trajectory without stopping the target's motion [31]. Another method involves pausing all targets in the scene during the selection process [30]. However, most of these techniques do not consider occlusion issues specific to 3D environments or the importance of interface layout [5]. Hasan et al. acknowledged that their techniques could not be effectively used in dense environments because these methods increase the selection area, altering the scene layout [55]. This alteration leads to scene overlap and confusion, which is not ideal or even unacceptable in certain scenarios. Among the techniques that do not alter the layout, the Hook technique [11] predicts selection based on the distance of the user's cursor to possible targets. By visually emphasizing possible candidate targets, their approach supports improved selection time [44]. These techniques are sensitive to certain parameters and are often scenario-specific.

2.2 Multi-sensory Attentional Cues for target selection

Previous works have shown that providing multimodal feedback can significantly enhance task performance and reduce perceptual load without increasing the user's overall workload [20]. As a critical component of interactive interfaces, multimodal feedback for target selection tasks has been widely explored, particularly in scenarios involving static and small targets, but it is commonly applied to mouse-based selection and in mobile device contexts [13].

Anthony et al. investigated the role of visual feedback in touch screen interactions for children [4]. Their findings revealed that the absence of visual feedback led to an increase in average errors. However, it is important to note that an overreliance on visual feedback can potentially diminish auditory and tactile perceptual abilities, thereby increasing the load on the visual system [56].

For audio feedback, Akamatsu et al. discovered that it did not improve overall selection time but reduced the time users spent dwelling on the target [3]. Batmaz et al. found that auditory feedback could effectively increase task throughput [8]. The study of spatial audio in aiding selection reported in [52] and Canales et al.'s work [16] indicated that auditory feedback was more important than visual feedback in VR. Finally, Cockburn et al. found that audio feedback could effectively reduce the average selection time for small targets [19].

Haptic feedback has been applied to improve selection in mobile interactions, such as key selection in touch-screen keyboards [14]. Akamatsu et al. integrated haptics into the mouse for better interaction with targets [2]. Ahmaniemi et al. explored the effect of dynamic haptic feedback on finger control [1].

Li et al. examined the speed-accuracy trade-off between visual-audio-haptic feedback [45]. Multimodal sensory feedback has also been compared for tasks and positioning in constrained scenarios [9], and it has been shown that choosing the appropriate modality feedback can enhance user selection efficiency and reduce error rates [40]. To our knowledge, no research has explored the benefits of active selection of moving targets through multimodal feedback mechanisms. The only investigation into feedback mechanisms for selecting moving targets is by Li et al. [46]. However, their study merely required users to confirm the selection of moving targets within a time window, without involving any active aiming by the users, and thus does not reflect the characteristics of selecting moving objects. To address this gap, our current work investigates the impact of visual, auditory, and haptic modalities on selecting moving targets and explores the effects of combining different modalities.

2.3 Proximity-Based Feedback Types

VR systems, compared to real-world interactions, often provide insufficient feedback, resulting in imprecise target localization and reduced task efficiency [59]. To address this limitation, Ariza et al. [6] introduced proximity-based feedback, which extends traditional feedback mechanisms by correlating feedback intensity with the spatiotemporal relationship to the target. This approach aims to facilitate the rapid spatial localization of targets. In general, proximity-based feedback is categorized into two types: Binary and Continuous. Binary feedback is triggered when the user enters the target's selectable area, while continuous feedback modulates the intensity so that it is inversely proportional to the user's distance from the target.

Ariza et al. [6] have shown that proximity-based feedback can significantly enhance throughput in 3D target selection tasks, particularly during the correction phase, thus improving overall user performance. Its efficacy has been corroborated across various modalities and contexts. Gao et al. [25] reported that continuous auditory feedback based on proximity enhances the performance of trajectory-based finger gestures in 2D interfaces. In VR environments, proximity-based haptic feedback has been shown to be effective in navigation tasks [7, 47]. Furthermore, continuous proximity-based multimodal feedback has demonstrated improved performance in eyes-free target selection scenarios [26]. Lu et al. investigated the impact of various proximity feedback types on gesture-based pointing in VR, specifically to address depth perception challenges [50].

Building upon these findings, our research focuses on optimizing eyes-free target acquisition performance through continuous proximity-based multimodal feedback. We aim to examine the effects of different types of proximity-based feedback on the selection of moving targets. Additionally, we introduce the concept of partial feedback, which integrates elements of both *binary* and *continuous* feedback to enhance selection performance further.

3 Proximity-Based Feedback Types for Different Modalities

In this section, we introduce several proximity-based feedback types developed for different modalities based on prior research [6].

- ***Binary Proximity-Based Feedback:*** This type represents the most common type. Feedback is activated only when the cursor enters the target area; otherwise, no feedback is provided. The functional relationship between feedback intensity and distance is illustrated in Fig. 2a.
- ***Continuous Proximity-Based Feedback:*** As defined by Ariza et al. [6], this type continuously modulates intensity based on the distance between the user's cursor and the target center. The maximum feedback intensity is experienced when the cursor is at the target's center. The functional relationship between feedback intensity and distance is illustrated in Fig. 2b.
- ***Partial Proximity-Based Feedback:*** This novel approach combines both binary and continuous feedback types. Feedback is continuously given and varies based on the cursor's distance from the target center, reaching maximum intensity at the center. Likewise, for the binary aspect, feedback is activated only when the cursor is within the target area but in a different dimension (e.g., for partial-based haptic feedback, the frequency is continuously changing, but the intensity only changes once). The functional relationship between feedback intensity and distance for this combined approach is illustrated in Fig. 2c.

To enhance our understanding of the feedback mechanisms, we next outline the configurations of binary, continuous, and partial feedback types across different modalities. For ***Binary*** feedback, visual feedback is represented by the R value in RGB, which ranges from 0 to 255. The color red was specifically chosen for its high contrast in proximity cues and for its proven impact on attracting attention [43]. Regarding our audio and haptic modalities, we assigned optimal frequency ranges to each. For the auditory modality, we focused on frequencies between 0 and 250 Hz, with an em-

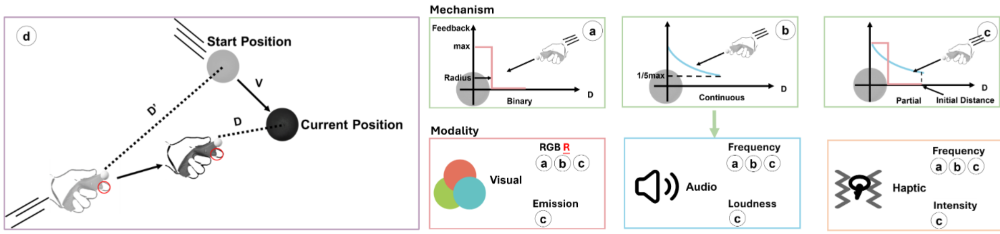

Figure 2: (d) The different components when selecting a moving target, (a-c) The function graphs of Binary, Continuous, and Partial feedback types showing the relationship between distance and feedback intensity.

phasis on around 250 Hz because it provides the highest localization accuracy [66] and has proven effective in tracking tasks [62]. Meanwhile, for haptic feedback, we employed vibration frequencies spanning 0 to 300 Hz, where 300 Hz has been demonstrated to be the highest sensitivity that the human finger can perceive [10].

For ***Continuous*** feedback, Visual feedback is represented by the R value in RGB, ranging from 51 to 255. Auditory frequency ranges from 50 to 250 Hz, while the haptic vibration frequency ranges from 60 to 300 Hz. The choice of highest audio and vibration frequency aligns with the optimal values identified in our binary feedback. In this type, 1/5 of the feedback is provided at the initial value, allowing users to perceive their distance to the target in real-time.

For ***Partial*** feedback, all sensory modalities are divided into two stages: Stage 1 (Continuous) and Stage 2 (Binary). Before reaching the target, feedback is in Stage 1. Once the cursor hits the target, it switches to Stage 2. In the visual modality, Stage 1 ranges from R(51) to R(255), while Stage 2 corresponds to emission intensity ranging from 0 to 2. In the auditory modality, Stage 1 frequency ranges from 50 to 250 Hz, aligning its upper limit with the binary and continuous modes, while Stage 2 loudness spans 0 to 6 dB. In the haptic modality, Stage 1 frequency ranges from 60 to 300 Hz—-again matching the same upper limit—-while Stage 2 intensity ranges from 0 to 120.

4 Research Question

Our review has pointed to a noticeable gap in our understanding of the correlation between user behavioral patterns and different feedback mechanisms in selecting moving targets and, in particular, of the impact of binary, continuous, and partial feedback types across visual, auditory, and haptic modalities. To bridge this gap and help frame our work, we formulated four research questions:

RQ1: *What are the properties of the partial proximity-based feedback in different modalities?* As described in Sec. 2.3, previous work has investigated the impact of binary and continuous feedback on user performance. Nevertheless, the integration effects of binary and continuous feedback remain unexplored. Therefore, it is necessary to investigate whether the partial feedback type can combine the advantages of continuous and binary feedback and identify its potential drawbacks.

RQ2: *Can adding different feedback types across various modalities enhance users performance in selecting moving targets?* Numerous studies have shown that feedback mechanisms can greatly enhance user selection performance. This improvement is observed for both static and moving targets, particularly in situations where precise aiming is not required, and users do not need to actively move their bodies to input [6, 46]. However, the exploration of feedback types has only been conducted in static target selection tasks. As such, further exploration of the properties of these feedback types for moving targets is still needed.

RQ3: *What are the properties of binary, continuous, and partial feedback types in different modalities?* In previous work on static target selection tasks, researchers have already discussed the throughput of binary and continuous feedback in different modalities. However, objective data (such as completion time and error rate) alone cannot fully evaluate the effectiveness and usability of feedback types across different modalities. In particular, in various task scenarios, designers need to provide appropriate feedback types based on specific needs. Therefore, it is necessary to discuss the properties of different feedback types within each modality.

RQ4: *Can the combination of multiple modalities further enhance user performance and experience?* In previous studies, multimodal integration has been shown to effectively enhance user performance in certain scenarios, while in other cases, it can lead to feedback overload Sec. 2.2. This is related to whether the chosen feedback type for each modality is appropriate. Therefore, we aim to explore the combination of the optimal feedback properties selected for each modality from **RQ3** and investigate the differences between unimodal, bimodal, and trimodal combinations to explore further their role in selecting moving targets.

5 User Study 1: Impact of Different Feedback Types on Various Modalities

The goal of this study is to compare and evaluate user performance and experience across three modalities, three feedback types, and a baseline condition for selecting moving targets.

5.1 Participants, Apparatus, and Materials

We recruited 20 participants (12 females and 8 males), aged between 18 and 30 (M = 22.7, SD = 3.4), with diverse educational backgrounds (computer science, arts, robotics, etc.) from a local university. All participants had normal or corrected-to-normal vision, and none reported issues with color vision. All participants were right-handed. Fifteen participants reported being familiar or very familiar with VR HMDs. The experiment was conducted on an Intel Core i9 processor PC with an NVIDIA GTX 4060 GPU. The program was developed using C#.NET and ran on the Unity3D platform. We used a Meta Quest 3 to provide the virtual environment and its controllers for interaction.

5.2 Test Environment and Task

In the experimental setup, a small ball was positioned at the tip of the right controller. This was treated as the collision range for the handle, representing the most basic form of the virtual hand [64]. A starting sphere was placed 20 cm in front of the user, requiring the user to collide with it to initialize the task. This ensured that users could re-acquire depth perception at the beginning of each task. One second after initialization, a moving sphere was generated 60 cm in front of the user as the selection target, moving in a randomly chosen direction along the X-Y plane.

The task required users to move the small ball on the controller as quickly and accurately as possible to the center of the target sphere. Participants had to press the trigger button on the controller to capture the target. A selection was deemed successful if the button was pressed while the small ball intersected the target sphere. Upon a correct selection, the system proceeded to the next trial. If the target was missed, the target sphere remained visible. Once a correct selection was made, the current target sphere disappeared, and participants needed to return their arms to the initial position to touch the starting sphere again for the next task.

Given that moving targets could be difficult to select if located outside the user's field of view, we employed a bounded space anchored in the virtual environment that would position all moving targets in front of the user, following the approach by Chen et al. [18]. This approach would allow us to analyze the effects of target speed, width, and technique without confounding factors, such as searching for out-of-view targets. The bounded space was designed based on a reachable workspace within arm length [58], with a radius of 60 cm. To prevent users from anticipating the rebound path of the moving object, the outline of our bounded space was invisible. Given that previous work [63] showed that the direction of motion relative to the target could lead to different targeting strategies, we added a bouncing feature to the boundary of the bounded space so that whenever the target hits the boundary, it would reflect and bounce away. Since the boundary was designed to be transparent, participants could not see its specific area. Consequently, they were unable to determine whether the target would change direction due to bouncing and could not predict the target's next movement in real time.

5.3 Experimental Design and Procedure

The experiment utilized a $(3 \times 3 + 1) \times 2 \times 2$ within-subjects design with four independent variables: Modality, Feedback Type (with the "+1" representing the Baseline, where no feedback was provided, and thus neither Modality nor Feedback Type applied, creating a distinct standalone condition.), Target Width, and Target Speed, leading to 40 experimental conditions. To minimize any carry-over effects, the 2 Width $\times$ 2 Speed conditions was counterbalanced via the Latin Square approach. The order of Modality and Feedback Type was randomly arranged. The conditions for each independent variable were as follows: Modality (Visual, Audio, and Haptic), Feedback Type (Binary, Continuous, and Partial), Target Width (6 cm, 8 cm), and Target Speed (1 m/s, 1.2 m/s).

Each participant spent approximately 40 minutes completing the experiment, which was divided into four distinct sessions. First, participants filled out a demographic questionnaire to provide their personal information and received a brief introduction to the experiment. They then underwent a minimum of five minutes of training to familiarize themselves with the techniques and learn about the tasks and procedures of the formal experiment. After the training, they proceeded to the formal trials for each condition as specified in the experimental design. At the end of each session, participants completed a short version of the User Experience Questionnaire (UEQ-S) [60], followed by a brief rest. Finally, upon completing all conditions, participants joined a semi-structured interview to discuss their preferences and provide feedback.

5.4 Evaluation Metrics

We employed a set of dependent variables encompassing both objective and subjective measurements.

5.4.1 Objective Measurements

For the objective measurements, we used the average results per condition and participant for statistical analysis. The objective metrics included: ***Selection Time***, which refers to the time (in seconds) taken to select the target in each trial, and ***Error Rate***, which is the number of incorrect selections in each trial, specifically instances where the user pressed the trigger button without intersecting the target sphere.

5.4.2 Subjective Measures

We also evaluated the techniques based on subjective measurements, including user experience and preference rankings. The subjective metrics included: ***User Experience Questionnaire (UEQ-S)***, a short version comprising eight items to measure user experience in terms of Pragmatic Quality, Hedonic Quality, and Overall User Experience, and ***Preference Ranking and Interview***, where participants ranked the three techniques based on their preferences for different modalities and provided reasons at the end of the experiment.

6 Results

6.1 Objective Results

From the experiment, we collected 9600 trials (20 participants $\times$ (3 Modalities $\times$ 3 Feedback Type) + Baseline) $\times$ 2 Widths $\times$ 2 Speeds $\times$ 12 repetitions). To analyze selection time and error rate, we removed outliers where selection times exceeded three standard deviations from the mean for each condition (261 trials, 2.71%). Such outliers are typically removed as they likely do not represent typical selection performance (e.g., minor distractions during the experiment) and can skew the results for a particular condition [61]. We assessed the normality of the data using both Shapiro-Wilk tests and Q-Q plots, confirming that completion time and error rate were normally distributed. We employed a two-way repeated measures ANOVA (RM-ANOVA), with modality and feedback type as the within-subjects factors, and used Bonferroni-adjusted pairwise comparisons to analyze selection time and error rate for each condition. Since we are interested in the performance of feedback types under different modalities, we present a detailed analysis of the effects and interactions between feedback types within each modality.

6.1.1 Selection Time

Fig. 3 a, b, and c illustrate the selection times across modalities, feedback types, and all mechanisms. Results from the RM-ANOVA tests revealed significant main effects of both MODALITY ($F_{1.731,32.882} = 57.102, p < 0.001, \eta_p^2 = 0.750$) and FEEDBACK TYPE ($F_{2.120,40.281} = 81.919, p < 0.001, \eta_p^2 = 0.812$) on selection time. Additionally, a significant interaction effect between MODALITY $\times$ FEEDBACK TYPE ($F_{3.626,68.891} = 22.418, p < 0.001, \eta_p^2 = 0.541$) was found for selection time.

Post-hoc tests showed that, in terms of modality, the *BaseLine* ($M = 2.49, S = 0.94$) had a significantly longer selection time compared to other modalities ($p < 0.001$). The *Haptic* modality was significantly faster than both *Visual* ($M = 2.14, S = 0.82$), ($p = 0.027$) and *Audio* ($M = 2.17, S = 0.70$), ($p = 0.007$). Regarding feedback type, post-hoc tests indicated that the *BaseLine* had a significantly longer selection time than all other feedback types (all $p < 0.001$). The *Continuous* feedback type ($M = 2.02, S = 0.73$) was significantly slower than the *Partial* (($M = 1.89, S = 0.64$), ($p < 0.001$). Further comparisons showed that *Visual Continuous* feedback mechanism (($M = 2.11, S = 0.83$) was slower than *Visual Partial* ($M = 1.85, S = 0.61$).

6.1.2 Error Rate

Fig. 3 d, e, and f, respectively, display the error rate between modalities, between feedback types, and across all conditions. Results from the RM-ANOVA tests revealed significant main effects of both MODALITY ($F_{1.709,30.764} = 33.923, p < 0.001, \eta_p^2 = 0.653$) and FEEDBACK TYPE ($F_{1.480,26.643} = 36.599, p < 0.001, \eta_p^2 = 0.670$) on error rate. Additionally, a significant interaction effect between MODALITY $\times$ FEEDBACK TYPE ($F_{4.023,72.423} = 15.137, p < 0.001, \eta_p^2 = 0.457$) was found for error rate.

Post-hoc tests showed that, in terms of modality, the *BaseLine* ($M = 0.28, S = 0.09$) had a significantly higher error rate compared to other modalities ($p < 0.001$). The *Haptic* ($M = 0.12, S = 0.05$)

modality had a significantly lower error rate than *Audio* ($M = 0.17, S = 0.07$), ($p = 0.006$). Regarding feedback type, post-hoc tests indicated that the *BaseLine* had a significantly higher error rate than all other feedback types (all $p < 0.001$). The *Binary* ($M = 0.13, S = 0.04$) feedback type had a lower error rate than the *Continuous* ($M = 0.17, S = 0.08$), ($p = 0.05$). The *Partial* feedback type ($M = 0.13, S = 0.05$) also had a lower error rate than the *Continuous* ($p = 0.006$). Further comparisons showed that *Visual Continuous* feedback mechanism ($M = 0.18, S = 0.06$) had a higher error rate than *Visual Partial* ($M = 0.11, S = 0.04$), ($p = 0.043$).

6.2 Subjective Results

We performed RM-ANOVAs and pairwise comparisons with Bonferroni adjustments to the ART-transformed questionnaire results, including UEQ scores, and a semi-structured interview with ranking.

6.2.1 User Experience

RM-ANOVA revealed a significant main effect on overall user experience for both MODALITY ($F_{3,57} = 16.84, p < 0.001, \eta_p^2 = 0.312$) and FEEDBACK TYPE ($F_{3,57} = 20.82, p < 0.001, \eta_p^2 = 0.274$). A significant interaction effect between MODALITY × FEEDBACK TYPE ($F_{9,171} = 26.19, p < 0.001, \eta_p^2 = 0.413$) was found for user experience.

Technique	Pragmatic	Hedonic	Overall
BaseLine	-1.75	-2.82	-2.28
Visual Binary	1.30 (>avg.)	-0.32	0.49
Visual Continuous	-0.25	0.38	0.07
Visual Partial	2.00 (>avg.)	-1.02	0.49
Audio Binary	2.05 (>avg.)	2.35 (exc.)	2.20 (exc.)
Audio Continuous	1.35 (>avg.)	1.55 (>avg.)	1.45 (>avg.)
Audio Partial	1.15 (>avg.)	2.10 (>avg.)	1.63 (>avg.)
Haptic Binary	2.55 (exc.)	1.08 (>avg.)	1.82 (>avg.)
Haptic Continuous	0.72	0.77 (>avg.)	0.74 (>avg.)
Haptic Partial	1.23 (>avg.)	-0.10	0.57

Table 1: Results from the short version of User Experience Questionnaires (UEQ-S), showing the pragmatic quality, hedonic quality, and overall quality of each technique. In the table, ">avg." means "above average" (orange cells), "exc." means "excellent" (yellow cells).

Post-hoc tests showed that, in terms of MODALITY, *BaseLine* ($M = -2.28, S = 1.09$) had a significantly lower score compared to *Audio* ($M = 1.76, S = 0.59$) and *Haptic* ($M = 1.04, S = 0.92$), (all $p < 0.001$). *Audio* had a significantly higher score compared to *Visual* ($M = 0.35, S = 1.13$), ($p = 0.017$). Regarding feedback types, post-hoc tests indicated that *BaseLine* had a significantly lower score than all others (all $p < 0.001$). *Binary* feedback type ($M = 1.50, S = 0.78$) was significantly higher than *Continuous* ($M = 0.75, S = 0.91$), ($p = 0.004$) and *Partial* (($M = 0.89, S = 0.96$), ($p = 0.012$). Further comparisons showed that *Haptic Binary* feedback mechanism ($M = 1.82, S = 0.48$) was significantly higher than *Haptic Partial* ($M = 0.57, S = 0.91$), ($p = 0.002$). *Audio Binary* feedback mechanism ($M = 2.20, S = 0.42$) was higher than *Audio Continuous* ($M = 1.45, S = 0.61$), ($p = 0.027$). (See Tab. 1)

6.2.2 Interview and User Ranking

Fig. 3 (g, h, and i) shows participants' rankings of different FEEDBACK TYPE based on their preferences for each MODALITY. In the ***Visual Modality***, 11 participants (55%) ranked the *Partial* feedback type as their top choice, followed by *Binary* (8 participants, 40%) and *Continuous* (1 participant, 5%). *BaseLine* was unanimously ranked last (20 participants, 100%). Sixteen participants (80%) preferred *Partial* and *Binary*, stating these feedback types "provide a sense of confidence" and "help clarify if the target object is fully reached" (P2-P4, P11). P6 noted, "The Partial feedback type makes tracking the target simpler and allows for quicker distance estimation to the target."

For the ***Auditory Modality***, 11 participants (55%) favored the *Binary* feedback type, followed by *Partial* (5 participants, 25%) and *Continuous* (4 participants, 20%). Seven participants (5%) found the continuously varying audio in *Continuous* and *Partial* types uncomfortable. P1, P2, and P4 remarked, "The wide frequency changes in a short time cause discomfort when selecting." P6 mentioned: "Auditory feedback is only noticeable during long-distance movement; subtle frequency changes are undetectable, especially when close to the target." Similar to the visual modality, 15 participants (75%) indicated that *Binary* and *Partial* feedback enhanced their confidence. It is worth mentioning that users reported a perceived delay between hearing the sound and pulling the trigger during the interview.

In the ***Haptic Modality***, 12 participants (60%) preferred the *Binary* feedback type, followed by *No Feedback* (5 participants, 25%) and *Partial* (3 participants, 15%). Seventeen participants (85%) reported that continuous vibration caused hand numbness, with P13-P16 stating that constant vibration makes selection during movement inaccurate. Notably, five participants preferred *No Feedback* over *Continuous* and *Partial* feedback type in this modality.

7 Discussion

7.1 Effects of Partial Proximity-Based Feedback Type in Different Modalities (RQ1)

Our objective data indicate that the *Partial* feedback outperforms both Binary and Continuous feedback in terms of selection time and error rate, with this advantage being particularly pronounced in the visual modality. In the auditory and haptic modalities, *Partial* feedback performs comparably to the Binary feedback but consistently surpasses the Continuous feedback. This pattern is further corroborated by subjective data, where the *Partial* feedback scored lower in pragmatic and hedonic qualities for auditory and haptic modalities. User feedback highlighted difficulties in processing multiple layers of feedback, particularly when dealing with continuous audio and vibration signals simultaneously. This finding aligns with prior research suggesting that overlaying multiple feedback types within a single channel can lead to user confusion and discomfort [21, 37], especially in complex tasks involving spatial aiming at moving objects [41]. Despite these challenges, our objective data demonstrate that the *Partial* feedback can significantly enhance task efficiency, indicating its potential to optimize performance in complex scenarios without substantially compromising user experience.

7.2 Effect of Adding Feedback In Different Modalities (RQ2)

RM-ANOVAs revealed a significant influence of adding feedback on movement time and error rates, which is consistent with previous studies [26]. On the UEQ scale, the *Baseline* scored the lowest in both Pragmatic and Hedonic qualities. User rankings and interviews revealed that participants found it challenging to grasp depth and motion direction in selection tasks without feedback, describing the feedback-free selection as very monotonous. Our results demonstrate that, for the selection of moving objects that require active spatial aiming, adding feedback as an auxiliary cue is essential to enhance user performance and experience.

Notably, we found that *Continuous* feedback also outperforms the *Baseline* for selecting moving objects. This finding contrasts with previous conclusions about stationary object selection. For

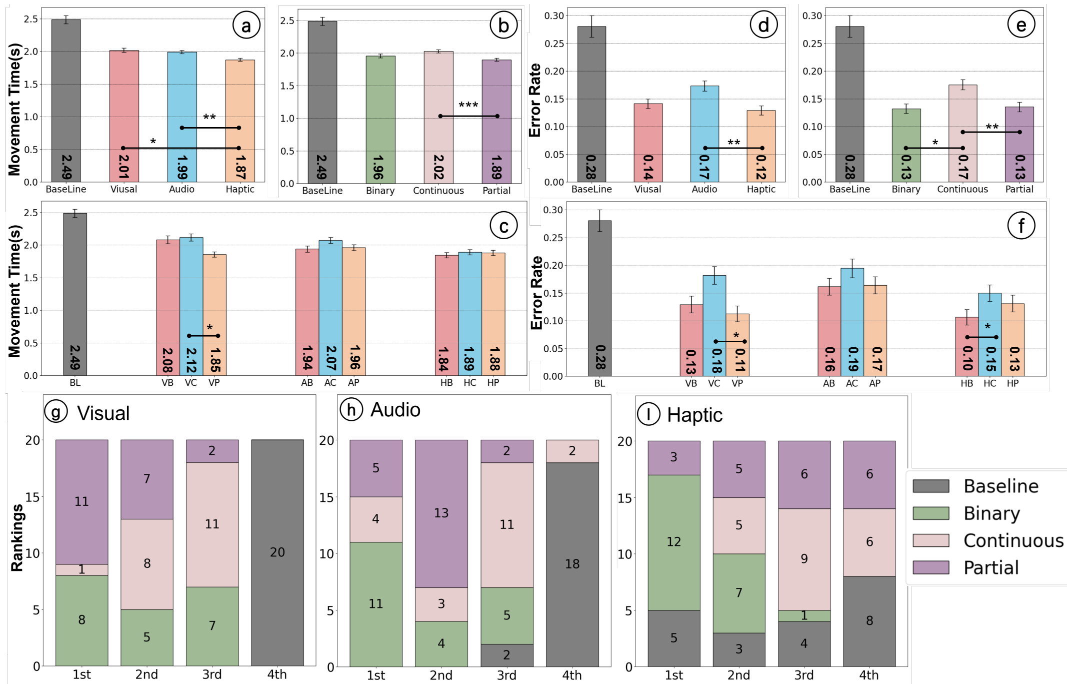

Figure 3: Objective measurement plots: (a-c) Movement time for the selection task, (d-f) Selection task error rate, categorized by (a, d) Modality, (b, e) Feedback type, and (c, f) All conditions combined. (g-i) Ranking of each Feedback type across different Modalities. Error bars represent 95% confidence intervals (*p $<$ 0.05, **p $<$ 0.01, ***p $<$ 0.001). Note: All Baseline comparisons with other conditions have p-values $<$ 0.001, but these are not explicitly marked.

example, in a previous study, Ariza et al. [6] suggested that with stationary objects, users sometimes get used or accustomed to pointing to the object (since it is static and its position remains the same across several trials), leading to lower accuracy demands for target localization when selecting stationary objects. In such cases, users may not fully take advantage of the benefits offered by *Continuous* feedback. However, in our task, since the target's initial movement direction is not known a priori, and its direction changes upon bouncing off invisible transparent boundaries (see Sec. 5.2), users cannot predict based on prior knowledge or experience the target's location. As a result, the *Continuous* feedback changes help users better understand the spatial relationship between the controller and the target, helping them adjust the direction of their hand movement. Additionally, in target selection tasks, we can divide the selection action into two main parts: (1) rapid target localization and (2) calibration [48]. Ariza et al. noted that one drawback of *Continuous* feedback in static target selection is that users, by quickly approaching the target through the feedback, experience a reduced calibration area, which affects accuracy. However, in moving target selection, the issue of a small calibration area does not arise, making *Continuous* feedback more advantageous and helpful in this context.

7.3 Evaluation of the Attributes of Different Feedback Types Across Various Modalities (RQ3)

In the **visual modality**, the Partial feedback type significantly outperformed the Continuous feedback in terms of selection time. It exhibited the same error rate as Binary feedback, which was better than Continuous feedback. This suggests that Binary feedback in the visual modality allows users to better gauge the timing of their actions, thus reducing error rates [22]; this may be because users are more sensitive to binary changes in feedback (from absence to presence). Although with continuous feedback, users reported being better able to grasp spatial relationships while moving (such as whether they were approaching an object), the timing of the selection may be less helpful, leading to a higher error rate. Partial feedback, combining the advantages of both Continuous and Binary feedback, showed the best preference in user experience, with 55% of users ranking it as the best visual feedback type. Therefore, we selected Partial feedback as the optimal type for the visual modality.

For the **auditory modality**, while the difference was not statistically significant, Binary feedback proved superior in terms of both selection time and error rate. Subjectively, Binary feedback also outperformed Continuous and Partial feedback in both pragmatic and hedonic evaluations. User interviews revealed that, for continuous audio feedback, users can detect frequency changes when initially approaching the object quickly. However, in the proximity zone, these changes become less noticeable, and users are not sensitive enough to discern the reduced distance [42]. Consequently, we chose the Binary feedback as the best for the auditory modality.

With the **haptic modality**, Binary feedback also showed the best performance in selection time and error rate compared to Continuous. Users found continuous haptic vibrations uncomfortable, leading to hand numbness and decreased sensitivity over time [53]. Subjective data supported this, with Continuous and Partial feedback rated lower in user experience and workload compared to Binary feedback. 60% of users selected Binary feedback as the best haptic feedback type. Based on these findings, we selected Partial feedback for the visual modality and Binary feedback for both auditory and haptic modalities as the optimal types.

8 User Study 2 - Impact of Multimodal Combinations

The second study aims to compare and evaluate user performance and experience by combining the best feedback identified in Study 1 across different modalities. Selecting inappropriate feedback types can lead to poor design, often resulting in negative experimental outcomes [5]. Based on the discussion of the Study 1 results, we have chosen Partial feedback for the visual modality, Binary feedback for the auditory modality, and Binary feedback for the haptic modality as the best feedback type for unimodal interactions. Specifically, we aim to explore the differences between multimodal and unimodal interactions under optimal feedback type conditions.

8.1 Participants, Apparatus, and Materials

This study took place about 10 days after Study 1 and involved 16 participants (6F/10M), between 19 and 30 years old (M = 23.1, SD = 3.8), from the same university campus for this study. All participants had normal or corrected-to-normal vision, and all were right-handed. Nine participants reported being familiar or very familiar with VR HMDs. We used the same apparatus and materials as in Study 1 (see Sec. 5.1). Eight of the participants had also participated in the first study.

8.2 Design, Procedure, and Metrics

We employed the same experimental setup and tasks as in Study 1 (see Sec. 5.2). The experiment utilized a 7 × 2 × 2 within-subjects design with three independent variables: Feedback Type, Feedback Width, and Target Speed, resulting in 28 experimental conditions. The seven Feedback Types are *Visual*, *Audio*, *Haptic*, *Visual-Audio*, *Visual-Haptic*, *Audio-Haptic*, and *Visual-Audio-Haptic*. We used the same metrics as in Study 1 (see Sec. 5.4).

9 Results

9.1 Objective Results

We collected 5,376 trials (16 participants × 7 Feedback Types × 2 Widths × 2 Speeds × 12 repetitions). We discarded the outliers (78 trials, 2.6%) to analyze the selection time and error rate. We employed RM-ANOVA with Greenhouse-Geisser correction to analyze the effect of each factor. Pairwise comparisons with Bonferroni adjustment were used for technique comparison. Since we were interested in how the techniques were affected by different environmental factors, we focused only on the effects and interactions related to the factor Feedback Type.

9.1.1 Selection Time

Fig. 4a displays the selection times across feedback types. Results from the RM-ANOVA tests revealed significant main effects of FEEDBACK TYPE ($F_{3.25,48.74} = 5.204, p < 0.001, \eta_p^2 = 0.258$) on selection time. Post-hoc tests showed that *Audio* ($M = 1.93, S = 0.18$) was significantly longer than *Visual-Audio* ($M = 1.74, S = 0.13$), ($p = 0.013$) and *Visual-Audio-Haptic* ($M = 1.68, S = 0.09$), ($p = 0.013$), ($p < 0.001$). *Haptic* ($M = 1.88, S = 0.25$) was significantly longer than *Visual-Audio-Haptic* ($p = 0.05$).

9.1.2 Error Rate

Fig. 4b shows the error rates between feedback types. Results from the RM-ANOVA tests revealed significant main effects of FEEDBACK ($F_{2.931,3.720} = 6.667, p = 0.001, \eta_p^2 = 0.308$) on error rate. Post-hoc tests showed that *Visual-Audio-Haptic* ($M = 0.081, S = 0.032$) had a significantly lower error rate compared to *Visual* ($M = 0.124, S = 0.046$), ($p = 0.01$), *Audio* ($M = 0.147, S = 0.054$), ($p = 0.003$), *Haptic* ($M = 0.117, S = 0.039$), ($p = 0.003$), ($p = 0.028$), *Visual-Audio* ($M = 0.112, S = 0.029$), ($p = 0.048$), and *Audio-Haptic* ($M = 0.147, S = 0.049$), ($p = 0.002$). *Visual-Haptic* ($M = 0.094, S = 0.036$) had a significantly lower error rate compared to *Audio-Haptic* ($p = 0.002$).

9.2 Subjective Results

As in the first study, we performed RM-ANOVAs and pairwise comparisons with Bonferroni adjustments to ART-transformed questionnaire results, including UEQ scores, and a semi-structured interview.

9.2.1 User Experience

RM-ANOVA revealed a significant main effect on *Pragmatic Quality* ($F_{6,90} = 7.592, p < 0.001, \eta_p^2 = 0.292$), *Hedonic Quality* ($F_{6,90} = 9.437, p < 0.001, \eta_p^2 = 0.347$), and *Overall User Experience* ($F_{6,90} = 8.856, p < 0.001, \eta_p^2 = 0.371$) among FEEDBACK TYPE. Post-hoc tests showed that, regarding *Overall User Experience*, *Visual-Audio-Haptic* ($M = 2.60, S = 0.27$) had a significantly higher score compared to *Visual* ($M = 1.10, S = 0.43$), ($p < 0.001$), *Audio* ($M = 1.33, S = 0.83$), ($p = 0.03$), *Haptic* ($M = 1.27, S = 0.65$), ($p < 0.001$), *Visual-Audio* ($M = 1.19, S = 0.58$), ($p < 0.001$), *Visual-Haptic* ($M = 1.72, S = 0.56$), ($p = 0.004$), and *Audio-Haptic* ($M = 0.99, S = 1.37$), ($p = 0.0014$). *Visual-Haptic* ($M = 1.72, S = 0.58$), was significantly higher than *Visual* ($p = 0.03$).

Technique	Pragmatic	Hedonic	Overall
Visual	1.64 (>avg.)	0.56	1.10
Audio	0.97	1.68 (>avg.)	1.33
Haptic	1.37	1.16	1.27
Visual-Audio	1.49	0.89	1.19
Visual-Haptic	2.02 (>avg.)	1.43 (>avg.)	1.73 (>avg.)
Audio-Haptic	1.0	0.98	0.99
Visual-Audio-Haptic	2.57 (exc.)	2.62 (exc.)	2.60 (exc.)

Table 2: Results from the short version of User Experience Questionnaires (UEQ-S), showing the pragmatic quality, hedonic quality, and overall quality of each technique. In the table, ">avg." means "above average" (orange cells), "exc." means "excellent" (yellow cells).

9.2.2 Interview and User Ranking

Fig. 4 shows the participants' rankings of different FEEDBACK TYPE based on their preferences. Since the unimodal mechanisms were extensively discussed in Study 1, this section focuses on the user interviews regarding bimodal and trimodal feedback combinations.

For the ***Feedback Type Combinations***, 10 participants (62.5%) preferred the *Visual-Audio-Haptic* combination, followed by *Visual-Haptic* (3 participants, 18.7%), *Visual-Audio* (2 participants, 12.5%), and *Haptic* alone (1 participant, 6.25%). Thirteen participants (21.6%) stated that multimodal combinations "*increased their confidence in gauging the distance to the target and helped clarify whether the target was fully reached*." For instance, P2-P4 and P11 noted, "*Although adding audio feedback sometimes caused delays in reaction, it significantly enhanced the entertainment value, preventing the selection task from becoming too monotonous*." These reported reaction delays are consistent with findings from previous research [54], which demonstrated that people's response times to haptic stimuli are 28% shorter than those to auditory stimuli. Besides, no additional delay was observed between each modality.

Regarding the ***Visual-Haptic Combination***, P6, P8, and P10 reported that "*Haptic feedback often comes the fastest and is the most intense stimulus, but prolonged vibration can lead to hand numbness*." Notably, the *Visual-Audio-Haptic* combination was highlighted by most participants as providing the best balance, as it integrates the advantages of visual cues and haptic feedback with the added value of audio, enhancing their experience without overwhelming them.

10 Discussion

10.1 Evaluating Multimodal Combinations (RQ4)

In this study, objective analysis of time and error rate showed that feedback mechanisms involving visual input consistently performed well. Mechanisms without visual feedback, such as Au-

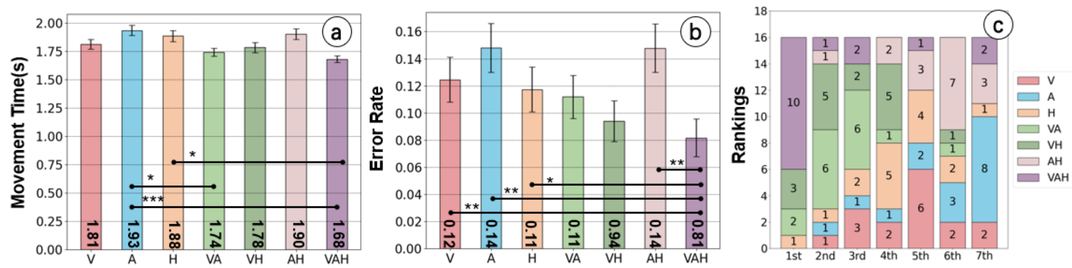

Figure 4: Objective measurement plots: (a) Movement time for the selection task, (b) Selection task error rate. (c) Ranking of each Mechanism. Error bars represent 95% confidence intervals (*p $<$ 0.05, **p $<$ 0.01, ***p $<$ 0.001).

dio and Audio-Haptic, exhibited poor performance in terms of error rates. When comparing bimodal and unimodal interactions, both Visual-Audio and Visual-Haptic outperformed unimodal interactions. Subjectively, users felt that auditory feedback significantly enhanced the hedonic aspect of tasks, while visual feedback was deemed most critical for pragmatic performance. The trimodal combination, which incorporates the best feedback mechanisms, effectively integrated these advantages. Additionally, 62.5% of users rated this as the best mechanism. Hence, our data indicate that the trimodal modality is the most effective for tasks involving active targeting of moving objects, both objectively and subjectively. Previous studies, such as the one conducted by Li et al. [46], reported poor performance for trimodal feedback. We attribute these results to the inclusion of suboptimal feedback mechanisms in their design, such as discrete visual feedback with color changes (e.g., turning red), continuous auditory feedback with volume changes, and continuous haptic feedback with varying vibration frequencies, rather than to the complexity of information.

10.2 Design Recommendations and Application

Based on the results of the two studies, we have distilled the following four design recommendations with their practical applications.

DR1. For rapid and accurate target selection, use *Visual Partial* feedback as the foundation. This provides continuous visual tracking during aiming, reducing selection time. For high-precision tasks, supplement the interaction with *Auditory* and *Haptic Binary* feedback to improve timing and accuracy. *Application:* This approach can be used in a VR football match (see Fig. 5a), where players need to be selected accurately without disrupting the game flow.

DR2. When the objective is to locate a target quickly, a combination of *Continuous* feedback mechanisms across *Visual*, *Auditory*, and *Haptic* modalities can work well. Real-time feedback based on proximity assists users in rapidly approaching the target. Our findings suggest that in scenarios where quickly locating the target is required, adding the *Partial* or *Binary* mechanism as an additional feedback channel may be unnecessary and could overwhelm the user with excessive information. *Application:* This approach can be applied to Bubble cursor-type interactions in VR (see Fig. 5b), where speed of target acquisition is prioritized over precision.

DR3. For entertainment-focused applications, consider adding *Binary Auditory* feedback. Our UEQ scale results show that this enhances the hedonic aspect of user experience. However, note that *Auditory* feedback alone minimally impacts task performance. For challenging selection tasks, additional measures may be needed to boost entertainment value. *Application:* This strategy can be applied to VR shooting games (see Fig. 5c), where entertainment value is the primary consideration.

DR4. In prolonged tasks involving moving object selection, minimize or avoid *Continuous* and *Partial* mechanisms for *Auditory* and *Haptic* feedback. Extended exposure to changing sounds and vibrations can cause user discomfort and desensitization, reducing their effectiveness over time. *Application:* This recommendation can be helpful in extended VR typing and chat sessions (see Fig. 5d) that require sustained, prolonged interaction.

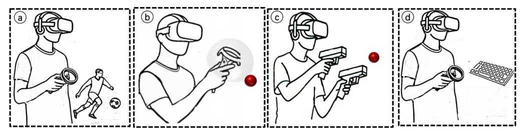

Figure 5: (a) Watching a soccer match and selecting a moving player. (b) Using Bubble Cursor technique to select a moving target. (c) Playing a shooting game. (d) Typing while walking.

11 Limitations and Future Work

We have identified several limitations in our work that suggest directions for future research. First, we constrained moving targets to a circular boundary, whereas, in real-world scenarios, objects can move over much larger areas. We also used only two sets of speed and size parameters and did not consider extreme situations—like targets moving at very high speeds—where continuous mechanisms might excel due to faster target detection. Additional studies are needed to confirm this possibility. Second, we focused on single-target selection, excluding multiple-target selection, to create a controlled environment for examining feedback mechanisms and modalities. As a result, our findings primarily apply to single-target tasks and may not extend to scenarios where users must manage multiple distractions or targets. Future work could explore how these mechanisms and modalities perform under more complex conditions. Lastly, we are interested in investigating a wider range of frequencies within a single sensory channel to deepen our understanding of how different feedback modalities function. By broadening the scope of our experiments, we hope to gain more comprehensive insights into designing effective feedback mechanisms for diverse real-world applications.

12 Conclusion

In this work, we introduced several feedback approaches and examined their performance to optimize the selection of moving targets in virtual reality (VR). Our work found that in unimodal channels with (1) binary feedback reduces user error rates, (2) continuous feedback helps users to approach targets quickly, and (3) partial feedback effectively provides additional information without diminishing user experience. We also combined and evaluated these feedback types to understand the role of each modality in multimodal settings for moving object selection tasks. Our results show that (1) visual is the most important modality for determining moving target direction, (2) audio feedback can enhance user engagement and sense of fun, and (3) haptic feedback can significantly reduce error rates. Based on our findings, we provide several recommendations and insights regarding multimodal feedback to assist in the design and framing of techniques for selecting moving objects in VR.

References

[1] T. Ahmaniemi. Effect of dynamic vibrotactile feedback on the control of isometric finger force. *IEEE transactions on haptics*, 6(3):376–380, 2012. 3

[2] M. Akamatsu and I. S. MacKenzie. Changes in applied force to a touchpad during pointing tasks. *International Journal of Industrial Ergonomics*, 29(3):171–182, 2002. 3

[3] M. Akamatsu, I. S. MacKenzie, and T. Hasbroucq. A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics*, 38(4):816–827, 1995. 3

[4] L. Anthony, Q. Brown, J. Nias, and B. Tate. Examining the need for visual feedback during gesture interaction on mobile touchscreen devices for kids. In *Proceedings of the 12th international conference on Interaction Design and Children*, pp. 157–164, 2013. 2, 3

[5] F. Argelaguet and C. Andujar. A survey of 3d object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, 2013. 3, 7

[6] O. Ariza, G. Bruder, N. Katzakis, and F. Steinicke. Analysis of proximity-based multimodal feedback for 3d selection in immersive virtual environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 327–334, 2018. 2, 3, 4, 7

[7] O. J. Ariza N., M. Lange, F. Steinicke, and G. Bruder. Vibrotactile assistance for user guidance towards selection targets in vr and the cognitive resources involved. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 95–98, 2017. 2, 3

[8] A. U. Batmaz, K. Yu, H.-N. Liang, and W. Stuerzlinger. Improving effective throughput performance using auditory feedback in virtual reality. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction*, pp. 1–11, 2022. 3

[9] J. D. Bell and K. L. Macuga. Goal-directed aiming under restricted viewing conditions with confirmatory sensory feedback. *Human Movement Science*, 67:102515, 2019. 3

[10] K. R. Boff and J. E. Lincoln. Engineering data compendium. human perception and performance. user's guide. 1988. 4

[11] P. Boffi, A. Kouyoumdjian, M. Waldner, P. L. Lanzi, and I. Viola. Bagginghook: Selecting moving targets by pruning distractors away for intention-prediction heuristics in dense 3d environments. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 913–923, 2024. 3

[12] S. Brewster. The design of sonically-enhanced widgets. *Interacting with Computers*, 11(2):211–235, 1998. 2

[13] S. Brewster, F. Chohan, and L. Brown. Tactile feedback for mobile interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, p. 159–162. Association for Computing Machinery, New York, NY, USA, 2007. 3

[14] S. Brewster, F. Chohan, and L. Brown. Tactile feedback for mobile interactions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 159–162, 2007. 3

[15] A.-M. Brouwer, T. Middelburg, J. B. J. Smeets, and E. Brenner. Hitting moving targets. *Experimental Brain Research*, 152(3), 2003. 2

[16] R. Canales and S. Jörg. Performance is not everything: Audio feedback preferred over visual feedback for grasping task in virtual reality. In *Proceedings of the 13th ACM SIGGRAPH Conference on Motion, Interaction and Games*, MIG '20. Association for Computing Machinery, New York, NY, USA, 2020. 3

[17] B. Chen, X. Yan, X. Hu, D. Kao, and H.-N. Liang. Impact of tutorial modes with different time flow rates in virtual reality games. *Proc. ACM Comput. Graph. Interact. Tech.*, 7(1), may 2024. 1

[18] Y. Chen, K. Katsuragawa, and E. Lank. Understanding viewport- and world-based pointing with everyday smart devices in immersive augmented reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2020. 5

[19] A. Cockburn and S. A. Brewster. Multimodal feedback for the acquisition of small targets. *Ergonomics*, 48:1129 – 1150, 2005. 3

[20] N. Cooper, F. Milella, C. Pinto, I. Cant, M. White, and G. Meyer. The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *PLOS ONE*, 13(2):1–25, Feb. 2018. Publisher: Public Library of Science. 3

[21] K. Darvish, L. Penco, J. Ramos, R. Cisneros, J. Pratt, E. Yoshida, S. Ivaldi, and D. Pucci. Teleoperation of humanoid robots: A survey. *IEEE Trans. Robot.*, 39(3):1706–1727, 2023. 6

[22] T. Djajadiningrat, K. Overbeeke, and S. Wensveen. But how, donald, tell us how? on the creation of meaning in interaction design through feedforward and inherent feedback. In *Proc. 4th Conf. Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, pp. 285–291, 2002. 7

[23] T. Feuchtner and J. Müller. Ownershift: Facilitating overhead interaction in virtual reality with an ownership-preserving hand space shift. In *Proc. 31st ACM Symp. User Interface Software and Technology*, UIST '18, p. 31–43. ACM, New York, NY, USA, 2018. 1

[24] I. S. Fitton, E. Dark, M. M. Oliveira da Silva, J. Dalton, M. J. Proulx, C. Clarke, and C. Lutteroth. Watch this! observational learning in vr promotes better far transfer than active learning for a fine psychomotor task. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24. Association for Computing Machinery, New York, NY, USA, 2024. 1

[25] B. Gao, H. Kim, H. Lee, J. Lee, and J.-I. Kim. Effects of continuous auditory feedback on drawing trajectory-based finger gestures. *IEEE Transactions on Human-Machine Systems*, 48(6):658–669, 2018. 2, 3

[26] B. Gao, Y. Lu, H. Kim, B. Kim, and J. Long. Spherical layout with proximity-based multimodal feedback for eyes-free target acquisition in virtual reality. In J. Y. Chen and G. Fragomeni, eds., *Virtual, Augmented and Mixed Reality. Multimodal Interaction*, pp. 44–58. Springer International Publishing, Cham, 2019. 2, 3, 6

[27] T. Grossman and R. Balakrishnan. The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, p. 281–290. Association for Computing Machinery, New York, NY, USA, 2005. 2

[28] M. Guillon, F. Leitner, and L. Nigay. Static voronoi-based target expansion technique for distant pointing. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, AVI '14, p. 41–48. Association for Computing Machinery, New York, NY, USA, 2014. 2

[29] T. J. Gunn, P. Irani, and J. Anderson. An evaluation of techniques for selecting moving targets. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, p. 3329–3334. Association for Computing Machinery, New York, NY, USA, 2009. 1

[30] A. A. Hajri, S. Fels, G. Miller, and M. Ilich. Moving target selection in 2d graphical user interfaces. In *Human-Computer Interaction–INTERACT 2011: 13th IFIP TC 13 International Conference, Lisbon, Portugal, September 5-9, 2011, Proceedings, Part II 13*, pp. 141–161. Springer, 2011. 3

[31] K. Hasan, T. Grossman, and P. Irani. Comet and target ghost: techniques for selecting moving targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 839–848. Association for Computing Machinery, New York, NY, USA, 2011. 1, 2, 3

[32] J. Hild, C. Kühnle, and J. Beyerer. Gaze-based moving target acquisition in real-time full motion video. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, ETRA '16, p. 241–244. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2857491.2857525 2

[33] E. R. HOFFMANN. Capture of moving targets: a modification of fitts' law. *Ergonomics*, 34(2):211–220, 1991. 2

[34] J. Huang, F. Tian, X. Fan, H. Tu, H. Zhang, X. Peng, and H. Wang. Modeling the endpoint uncertainty in crossing-based moving target selection. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376336 2

[35] J. Huang, F. Tian, X. Fan, X. L. Zhang, and S. Zhai. Understanding

the uncertainty in 1d unidirectional moving target selection. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2018. 2

[36] J. Huang, F. Tian, N. Li, and X. Fan. Modeling the uncertainty in 2d moving target selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, p. 1031–1043. Association for Computing Machinery, New York, NY, USA, 2019. 2

[37] X. Huang, M. Halwani, R. Muthusamy, A. Ayyad, D. Swart, L. Seneviratne, D. Gan, and Y. Zweiri. Real-time grasping strategies using event camera. *J. Intell. Manuf.*, 33(2):593–615, 2022. 6

[38] S. Ikeno, C.-M. Chang, and T. Igarashi. Sticky cursor: A technique for facilitating moving target selection. In *International Conference on Human-Computer Interaction*, pp. 467–483. Springer, 2023. 1

[39] M. V. Ilich. *Moving target selection in interactive video*. PhD thesis, Univ. British Columbia, 2009. 1

[40] A. Jaimes and N. Sebe. Multimodal human-computer interaction: A survey. *Comput. Vis. Image Underst.*, 108(1–2):116–134, oct 2007. 3

[41] S. Jeong, J. Kim, and J. Lee. The differential effects of multisensory attentional cues on task performance in vr depending on the level of cognitive load and cognitive capacity. *IEEE Trans. Vis. Comput. Graph.*, 30(5):2703–2712, 2024. 6

[42] V. O. Knudsen. The sensibility of the ear to small differences of intensity and frequency. *Phys. Rev.*, 21:84–102, Jan 1923. 7

[43] M. Kuniecki, J. Pilarczyk, and S. Wichary. The color red attracts attention in an emotional context. an erp study. *Front. Hum. Neurosci.*, 9, 2015. 3

[44] B. Lee, S. Kim, A. Oulasvirta, J.-I. Lee, and E. Park. Moving target selection: A cue integration model. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2018. 3

[45] T. Li, D. Wang, C. Peng, C. Yu, and Y. Zhang. Speed-accuracy tradeoff of fingertip force control with visual/audio/haptic feedback. *International Journal of Human-Computer Studies*, 110:33–44, 2018. 3

[46] Y. LI, D. WU, J. HUANG, F. TIAN, H. WANG, and G. DAI. Influence of multi-modality on moving target selection in virtual reality. *Virtual Reality & Intelligent Hardware*, 1(3):303–315, 2019. Human-computer interactions for virtual reality. 2, 3, 4, 9

[47] R. W. Lindeman, J. L. Sibert, E. Mendez-Mendez, S. Patil, and D. Phifer. Effectiveness of directional vibrotactile cuing on a building-clearing task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05. Association for Computing Machinery, New York, NY, USA, 2005. 3

[48] G. Liu, R. Chua, and J. T. Enns. Attention for perception and action: task interference for action planning, but not for online control. *Experimental Brain Research*, 185(4):709–717, Mar 2008. doi: 10.1007/s00221-007-1196-5 7

[49] J. Liu, B. Ens, A. Prouzeau, J. Smiley, I. K. Nixon, S. Goodwin, and T. Dwyer. Datadancing: An exploration of the design space for visualisation view management for 3d surfaces and spaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. 1

[50] Y. Lu, B. Gao, H. Tu, W. Luo, and H. Kim. Effects of different proximity-based feedback on virtual hand pointing in virtual reality. In N. Magnenat-Thalmann, V. Interrante, D. Thalmann, G. Papagiannakis, B. Sheng, J. Kim, and M. Gavrilova, eds., *Advances in Computer Graphics*, pp. 608–620. Springer International Publishing, Cham, 2021. 3

[51] Y. Luo, J. Wang, R. Shi, H.-N. Liang, and S. Luo. In-device feedback in immersive head-mounted displays for distance perception during teleoperation of unmanned ground vehicles. *IEEE Transactions on Haptics*, 15(1):79–84, 2022. doi: 10.1109/TOH.2021.3138590 1

[52] G. N. Marentakis and S. A. Brewster. Effects of feedback, mobility and index of difficulty on deictic spatial audio target acquisition in the horizontal plane. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, p. 359–368. Association for Computing Machinery, New York, NY, USA, 2006. 3

[53] S. McIntyre, A. O. Holcombe, I. Birznieks, and T. Seizova-Cajic. Tactile motion adaptation reduces perceived speed but shows no evidence of direction sensitivity. *PLOS ONE*, 7(9):1–12, 09 2012. 7

[54] A. W. Ng and A. H. Chan. Finger response times to visual, auditory and tactile modality stimuli. In *Proceedings of the international multiconference of engineers and computer scientists*, vol. 2, pp. 1449–1454. IMECS Hong Kong, 2012. 8

[55] M. Ortega. Hook: Heuristics for selecting 3d moving objects in dense target environments. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 119–122, 2013. 2, 3

[56] S. Oviatt, A. DeAngeli, and K. Kuhn. Integration and synchronization of input modes during multimodal human-computer interaction. In *Referring Phenomena in a Multimedia Context and Their Computational Treatment*, ReferringPhenomena '97, p. 1–13. Association for Computational Linguistics, USA, 1997. 3

[57] S. Pei, D. Kim, A. Olwal, Y. Zhang, and R. Du. Ui mobility control in xr: Switching ui positionings between static, dynamic, and self entities. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24. Association for Computing Machinery, New York, NY, USA, 2024. 1

[58] S. Pheasant and C. M. Haslegrave. *Bodyspace: Anthropometry, Ergonomics and the Design of Work, Third Edition*. CRC Press, 3 ed., 2006. 5

[59] N. Ryu, H.-Y. Jo, M. Pahud, M. Sinclair, and A. Bianchi. Gamesbond: Bimanual haptic illusion of physically connected objects for immersive vr using grip deformation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. 3

[60] M. Schrepp, J. Thomaschewski, and A. Hinderks. Design and evaluation of a short version of the user experience questionnaire (ueq-s). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6):103–108, 12/2017 2017. 5

[61] R. Shi, Y. Wei, X. Hu, Y. Liu, Y. Yue, L. Yu, and H.-N. Liang. Experimental analysis of freehand multi-object selection techniques in virtual reality head-mounted displays, 2024. 5

[62] C. Trepkowski, A. Marquardt, T. D. Eibich, Y. Shikanai, J. Maiero, K. Kiyokawa, E. Kruijff, J. Schöning, and P. König. Multisensory proximity and transition cues for improving target awareness in narrow field of view augmented reality displays. *IEEE TVCG*, 28(2):1342–1362, 2022. 4

[63] J. R. Tresilian. Hitting a moving target: Perception and action in the timing of rapid interceptions. *Perception & Psychophysics*, 67(1):129–149, Jan. 2005. 5

[64] J. Wentzel, G. d'Eon, and D. Vogel. Improving virtual reality ergonomics through reach-bounded non-linear input amplification. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2020. 4

[65] A. Worden, N. Walker, K. Bharat, and S. Hudson. Making computers easier for older adults to use: area cursors and sticky icons. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, p. 266–271. Association for Computing Machinery, New York, NY, USA, 1997. 2

[66] W. A. Yost and X. Zhong. Sound source localization identification accuracy: Bandwidth dependencies. *The Journal of the Acoustical Society of America*, 136(5):2737–2746, 11 2014. 4

A Supplemental Material Instructions

Speed	Width	Time Mean (s)	95% CI
1.0	0.06	2.054	± 0.141
1.0	0.08	1.894	± 0.127
1.2	0.06	2.212	± 0.178
1.2	0.08	1.898	± 0.134
1.4	0.06	2.315	± 0.158

Table 3: Mean Time and 95% CI by Speed and Width