Cone&Bubble: Evaluating Combinations of Gaze, Head and Hand Pointing for Target Selection in Dense 3D Environments

Ludwig Sidenmark, Zibo Sun, and Hans Gellersen

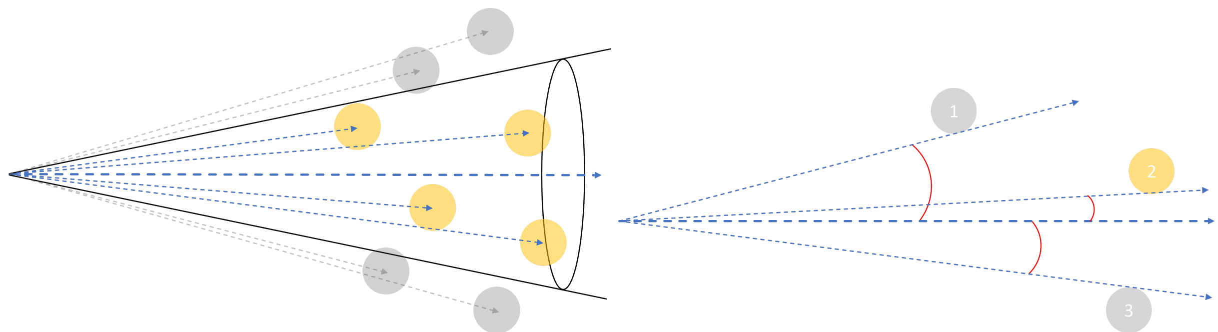

Figure 1: In the Cone&Bubble technique, users select a target by performing with two modalities in parallel via area cursors. The user points with a cone pointer (left) to indicate candidate targets used for selection. The N closest targets within the cone are indicated as candidates. The user points with the bubble ray technique (right) to indicate the final selected target. The target that is closest to the ray is indicated as the target wanted for selection.

Abstract

Target selection is an essential task in virtual reality (VR), but can be challenging when targets are small or in dense environments. In this work, we present Cone&Bubble techniques for easy selection of targets. Our techniques rely on two modalities for selection, one for rough candidate selection and one for precise final selection. The user first points with an area cursor to select candidate objects. The user then points with the second modality via the bubble cursor mechanism, which identifies the nearest target to the pointing ray. We investigate our technique with gaze, head, and controller pointing and compare the effect of each modality on each pointing role. Our results show that gaze pointing is most suitable for quick candidate selection and that controller pointing allows for easy direction changes and granular pointing for final selection. Head was the least performant modality due to its slow and strenuous nature.

Keywords: 3D Interaction, selection, area cursor, bubble cursor.

1 Introduction

Target acquisition represents an essential task for human-computer interaction and 3D user interfaces. In virtual reality (VR), the dominant selection technique is ray-casting from a hand-held controller. Ray-casting is widely used as it enables selection of targets at any distance. However, when targets are distant and small or close together than they can be difficult to acquire by ray-casting due to hand tremor [23].

In this work, we propose to combine two strategies for improving selection in virtual reality: selection of the target nearest to the pointer, originally introduced by the Bubble Cursor [13], and refinement of a coarse selection with a different modality [20] through cone-based volume pointing. The Bubble Cursor is based on a metaphor of dynamically resizing a cursor such that it always intersects only the nearest target. It relaxes the requirement for pointing precision by effectively increasing target width. However, the effect diminishes in dense environments, as the increase in target width is marginal when targets are close together. In the Cone&Bubble technique, we adopt a multimodal strategy where one modality is used for coarse pointing in 3D to pre-select a small number of targets, and a second modality to finalise the selection based on the Bubble mechanism.

Cone&Bubble is effective as cone pointing reduces the number of targets for Bubble raycasting (Figure 1). The cone width is adaptive for selection of a pre-set number of candidate targets. Targets within the cone may be positioned in close proximity but a Bubble ray can be cast to closer to the edge of the cone for effective disambiguation. Cone&Bubble is efficient as it uses two independent pointing modalities. A first modality needs to align the cone with the intended target before the second modality can complete the selection. However this does not require an explicit mode switch, and both modalities can move in parallel.

Cone&Bubble can be implemented with any combination of pointing modalities. In this work, we considered gaze, head and controller pointing, each for either cone casting or casting of the bubble ray. To gain insight into the effect of modality in either role, we compared all six permutations in a user study (N=12) on a selection task in a dense environment, with cone casting set to pre-select 4 candidate targets. Users were effective with all combinations. Gaze was found best suited for cone casting, as the eyes move faster than head or hand. Hand controller appeared best suited for refinement, as it is decoupled from the head-mounted display (HMD). Head, in contrast, was found least suitable for the final selection.

In sum, our contributions are:

- The Cone&Bubble technique for multimodal target selection in dense 3D environments
- An evaluation of gaze, head and hand combinations for coarse selection and bubble ray casting

2 Related Work

Our work is based on fundamental work on the comparison of modalities for target selection, multimodal interaction techniques, and target acquisition techniques.

2.1 Ray-casting

Ray-casting is a fundamental target acquisition technique where users shoot a ray from the user (e.g. the hand) out into the environment [3]. Whichever target intersects the ray is a candidate for selection. Ray-casting is popular as it is simple to understand by users and can select targets beyond their reach [28]. As such, ray-casting has been proposed for a wide variety of modalities, such as hands and hand-held controllers [14, 20], head [4, 20], gaze [16, 20], feet [22], and more. As a consequence, multiple studies have compared these modalities' ray-casting performance and user experience [11, 22, 27]. In general, hand- or controller-based ray-casting is preferred due to its high usability and control [11, 15]. However, gaze-based and head-based ray-casting have been shown to be a viable hands-free alternative: gaze for its speed and ease [11, 22], and head for its control [20, 27]. However, due to physiological factors such as hand tremor [19] or sensor noise in the case of eye tracking [10], pointing with ray-casting can become challenging for small or distant targets in dense environments. In this work, we tackle these limitations by combining area cursors for multi-modal target acquisition.

2.2 Area Cursors

Small target selection has been a long-lasting problem in Human-Computer Interaction and has received extensive attention for multiple devices [5, 29, 32]. In VR, this issue is extra relevant as targets can be presented at different depths from the user, making it difficult to control the size from the user's perspective (that is, the visual size) [6]. A common strategy to overcome this challenge is to increase the width of the pointing cursor or the targets. Early work proposed *area cursors* where users point to an area instead of a single point in the screen space [12, 17, 26]. Such techniques relax the amount of precision required for selection. The *Bubble cursor*, originally proposed for the desktop, is an area cursor where the target closest to the current cursor position is highlighted for selection [14]. The bubble cursor is advantageous as it requires minimal change from the typical desktop cursor while efficiently increasing the cursor's size. The bubble cursor principle has been shown to be effective for a wide variety of modalities such as gaze [8], and touch [1]. The principle has also been shown to be effective in 3D environments by adjusting the distance between the cursor and the target from Euclidian to angular distances [9, 21]. However, the effectiveness of the bubble cursor diminishes in dense environments, as targets are closely spaced together. As such, small target selection in dense environments remains a challenge. Our work leverages the bubble cursor mechanism for selection, and overcomes its weaknesses by combining it with a second area cursor preselection.

A second utility of area cursors is to minimise the potential candidates for selection to make final confirmation more robust while also reducing the amount of needed granularity. Previous techniques have leveraged this principle for progressive refinement of candidates [18], or to reduce the visual clutter of techniques by only adjusting a subset of targets [29, 30]. Examples include magnifying lenses where targets within the area cursor are increased in size to make selection and interaction easier [2, 8]. These techniques have proven to be effective in increasing selection performance and also in solving dense environment issues such as with the bubble cursor [8]. However, these techniques significantly alter the target appearance and usually add additional steps to the interaction, making it more cumbersome for users over time. In our work, we leverage such area cursors in a multi-modal fashion to allow preselection of candidates and final selection with the bubble cursor to be performed by users concurrently to increase performance and efficiency.

2.3 Multimodal Target Selection

Multimodal techniques have been extensively investigated to overcome the limitations of a single modality and to combine their strengths. An early example was the *MAGIC pointer*, where the cursor jumps to the gaze position at the start of a mouse movement, and the user performs final refined pointing with the hand [34]. This principle allows users to move effortlessly and quickly over large distances with gaze, while keeping the precision of the mouse. This principle has been widely adopted in different contexts such as pen [25], touch [33] and HMD-based interaction [16]. A common factor in these multimodal techniques is that a fast modality for rough pointing is combined with a second, more accurate modality. For example, Kytö et al. proposed to combine gaze-pointing with subsequent head- or controller-pointing [20]. Meanwhile, Sidenmark et al. combined head-based pointing with subsequent gaze-based gestures [30]. Furthermore, a common theme in these works is that each modality acts sequentially rather than in parallel. Previous work has proposed multi-modal techniques where for example gaze and head coordination [31], or gaze and hand coordination are leveraged and where movements are expected to happen concurrently [24]. Our techniques are designed to allow parallel pointing movements with each area cursor. As such, the suitability and efficiency of combinations of modalities remain unclear. Therefore, we compare different combinations of gaze, head, and hand pointing.

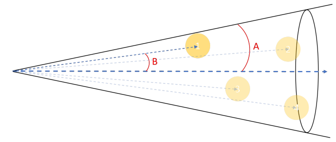

Figure 2: Cone pointer: the angular distance of target 1 to the ray(B) is compared to the visual angle radius(A)

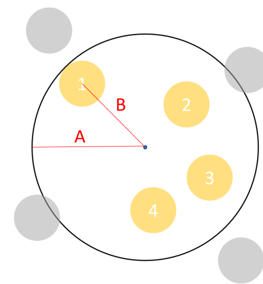

Figure 3: Cone pointer: functionality of the cone pointer from the user's perspective.

3 CONE&BUBBLE

Our *Cone&Bubble* consists of the following steps from a system perspective for target selection:

1. A pointing modality is used to preselect a subset of available targets as candidates for further interaction via cone pointing.
2. Candidate targets are highlighted.
3. A second pointing modality is used to indicate the intended target for final selection via a Bubble ray.
4. The user confirms the selection via a trigger (i.e., button click).

Note that the user can perform step one and three in parallel as they are associated with separate modalities. The user can change the current intended target for final selection by either moving the cone pointer to change the current candidates considered for the bubble ray or by moving the bubble ray to change the intended target from the current preselected candidates.

3.1 Cone Pointer

In our technique, users first select candidate objects by selecting a subset of targets via a cone pointer. The cone pointer is a 3D volume

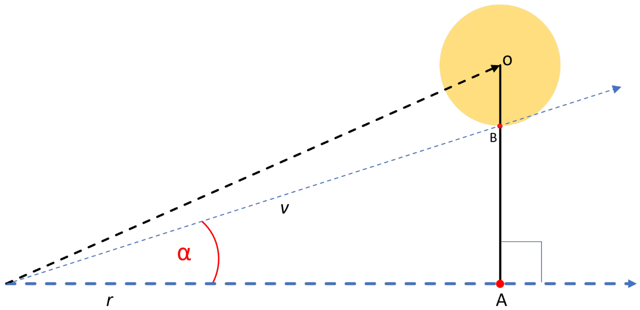

Figure 4: Finding the closest point of an object to a ray. We project the vector from origin to the centre(o) of the target onto the ray cast(r) and find point B, the closest point of the target to r. We then draw the vector from origin to B and consider the angle α.

pointer in the shape of a cone (Figure 2). While normal single-point cursors only select a single point in space, area cursors have a larger selection area, which can encompass multiple targets [17]. Area cursors have a bigger activation point and can, therefore, easily reach targets. From the point of view of the user, they will select targets in a circle in front of them (Figure 3). The cone pointer measures the angle between the central direction of the cone and the edge of each target hit. Given a visual angle radius (Figure 3A), the algorithm retrieves all targets with an angular distance to the ray (e.g, Figure 3B) lower than Figure 3A. The cone is defined by an angular radius R and a maximum number of candidates N to ensure that the area cursor does not preselect too many targets. If more than N targets are within the cone, the cone pointer will preselect the N closest targets to the centre direction.

3.2 Bubble Ray

After the user has indicated the intended target as a candidate, the user will point at and select the final object with a second modality using the bubble ray. The candidate target that is closest to the current pointing direction is highlighted as the final selection candidate. As recommended by previous work on bubble selection in VR [21], we use the angular difference between the pointing direction and the direction from the pointer origin and the target to decide the final target. The direction from the pointer to the target is defined by the directional vector from the pointing modality origin to the point of the object closest to the pointing direction. We find this closest point by using the projection from the centre of the target to the ray cast, as shown in Figure 4. If a target intersects with the pointing direction then that target is used as the final selection candidate. If the pointing direction intersects with multiple targets, we consider the first intersected target as the final target candidate.

3.3 Technique Permutations

Table 1 lists the combinations of modalities that we consider for our technique. We used basic modalities that are commonly available in modern head-mounted displays. The selection process follows the same pattern for every combination of modality: first, the user will direct the cone pointer to the approximate location of the final target, then they will use the bubble ray modality to disambiguate between the four candidate targets.

4 Methodology

We evaluated the performance and usability of our multimodal bubble selection techniques in a VR user study.

4.1 Task

Our task was inspired by previous work on Bubble Cursor techniques in VR [21], in which the participant had to select the correct target

Name	Cone Pointer		Bubble Ray
HeadHand	Head	+	Hand
HeadGaze	Head	+	Gaze
HandHead	Hand	+	Head
HandGaze	Hand	+	Gaze
GazeHead	Gaze	+	Head
GazeHand	Gaze	+	Hand

Table 1: Multimodal selection techniques.

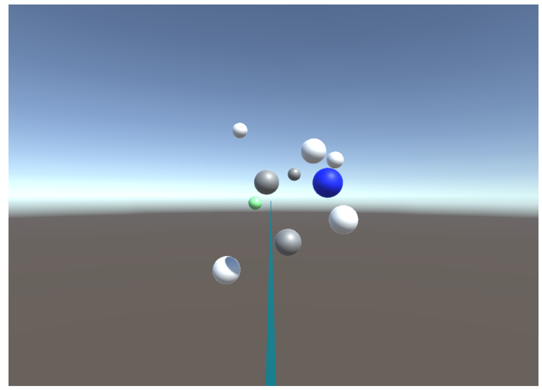

Figure 5: A target selection trial.

from many as quickly and as accurately as possible. Each trial involved ten spherical targets of different sizes (Figure 5). One of the targets had a distinct colour (blue) that the participant was tasked with selecting, while the other nine targets were distractors. All targets were spawned within an area of 30 degrees in front of the user direction. Distractor targets varied in 2-5 degrees in target size and we ensured that at least one target was within 1 degree of the correct target. We varied the target size of the correct target (*1, 3, 5°*) and the depth range of the correct target and distractor targets (*5-10, 15-20, 20-25 metres*). Participants could not finish a trial until they selected the correct target or until a 10 seconds had elapsed. The participants performed 3 repetitions of each condition. Trials were pre-generated to ensure that there was no overlap of occlusion between targets. The participants performed 6 techniques × 3 target sizes × 3 target depth ranges × 3 repetitions = 162 trials.

4.2 Participants and Apparatus

We recruited twelve participants (7M, 5F aged M:21.25, SD:1.05 years). All participants were students at *anonymised* and were recruited from student group chats or mailing lists. Two participants stated that they have regularly used virtual reality before, five expressed occasional use, and five had no experience with VR before. We developed the techniques and study environment in Unity (version 2021.3.9f1). We used an HTC Vive Pro Eye (display framerate 90Hz, eye tracker framerate 120Hz) to record eye and head movements, and an HTC Vive controller to record controller movements. The participants used the touchpad button on the HTC Vive controller for selection confirmation for all techniques. Based on pilot testing we selected $R = 30$ and $N = 4$. We used a One-euro filter on all modalities to minimise signal noise [7].

4.2.1 Visual Design of Technique

We use colour highlighting to help users keep track of target states. The target states and their respective visual feedback were defined as follows:

- **Neutral:** The default state of the target. No feedback is shown

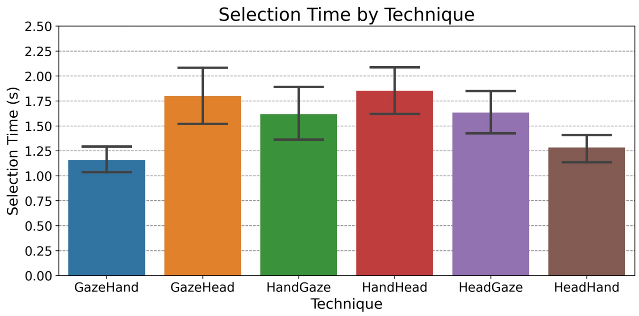

Figure 6: Average selection time for each technique. The error bars represent 95% confidence intervals of the means.

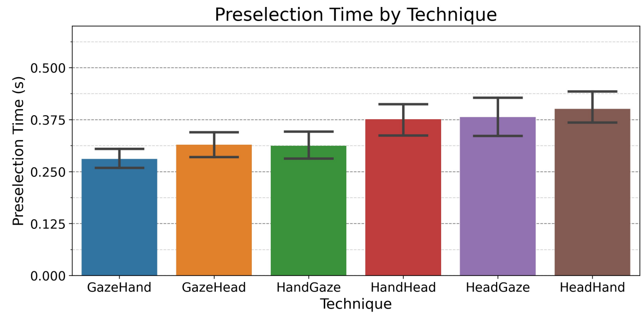

Figure 7: Average preselection time for each technique. The error bars represent 95% confidence intervals of the means.

on target

- **Candidate:** The targets are highlighted by the cone pointer and appear darker in colour.
- **Final candidate:** The target highlighted by the bubble ray. The target is highlighted in light green.
- **Selected target:** A confirmed selection of a final candidate via button click highlights the target in dark green.

The cone pointer was not visualised to reduce visual clutter. The bubble ray was visualised through a line or cursor, depending on the modality.

4.3 Procedure

On arrival, the participants received a short briefing on the procedure and purpose of the study. Participants then signed a consent form and answered a demographic questionnaire. The experimenter then explained how to use the techniques. The participants were then instructed to put on the HMD and perform the eye tracking calibration procedure. The participants then performed a short training session with the technique until they felt comfortable before starting data collection. After completing all selections with a technique, the participants removed the headset and answered a post-task questionnaire consisting of four 7-point Likert items about the technique. The participants were then instructed to rest until they were ready to proceed. The order of the technique was counterbalanced with a balanced Latin square. After finishing all the techniques, the participants responded to a final questionnaire about their preferences. The study took 45 minutes to complete. Participants were not compensated for their participation.

5 Results

The dependent variables of interest were selection time, error rate, preselection time, refinement time, and perceived usability. Unless otherwise stated, the analysis was performed with a three-way repeated measures ANOVA (α=.05) with Technique, Target depth and Target size as independent variables. When the assumption of Sphericity was violated, as tested with Mauchly's test, Greenhouse-Geisser corrected values were used in the analysis. QQ-plots were used to validate the assumption of normality. Bonferroni-corrected posthoc tests were used when applicable. Effect sizes are reported as partial eta squared (η_p^2). The Likert-scale usability data were analyzed using Friedman tests, and Bonferroni-corrected Wilcoxon signed-rank tests were used for the posthoc analysis.

5.1 Task Completion Time

We define the selection time as the time from the trial start until the correct target has been successfully selected. We only consider trials were the correct target was successfully selected without any previous erroneous selections. Outlier trials were defined as those trials with selection times more than 2 standard deviations from the grand mean. These trials were not considered for statistical analysis. In total, 252 (∼13%) trials were disregarded for this part of the analysis. Repeated measures ANOVA results showed no significant 3-way or 2-way interactions. However, we found a significant main effect for Technique ($F_{5,55}$=12.45, p<.001, η_p^2=.531), Figure 6). Posthoc results showed that GazeHand was significantly faster than GazeHead (p<.05), HandHead (p<.001) HeadGaze (p<.001). HeadHand was also significantly faster than GazeHead (p<.05), HandHead (p<.05), and HeadGaze (p<.05). Target size also had a main effect ($F_{2,22}$=38.46, p<.001, η_p^2=.778)). Posthoc results showed that the selection of targets of 1 degree was significantly slower than 3- (p<.001) and 5-degree targets (p<.001).

5.2 Preselection Time

We define preselection time as the time between start of the trial until the cone pointer first preselected the correct target. Only successful trials were used for this part of the analysis. Outlier trials were defined as those trials that had disambiguation times more than 2 standard deviations from the grand mean. These trials were not considered for statistical analysis. In total, 256 (∼13%) trials were disregarded for this part of the analysis. Repeated measures ANOVA showed no significant 3-way or 2-way interactions. However, we found a significant main effect for Technique ($F_{5,55}$=7.00, p<.01, η_p^2=.313), Figure 7). Posthoc results showed that GazeHand was significantly faster than HandHead (p<.001), HeadGaze (p<.05) and HeadHand (p<.001). We also found that GazeHead was significantly faster than HeadHand (p<.05). The results show that the fast gaze is attractive for area cursor pointing as users can quickly and effortly point in the general direction of target with ease. We also found a significant main effect for Depth ($F_{2,22}$=45.78, p<.001, η_p^2=.806)). Posthoc results showed that targets closer to the user took significantly more than disambiguation time than middle (p<.001) and far targets (p<.001). Finally, we also found a significant main effect for the target size ($F_{2,22}$=18.71, p<.001, η_p^2=.630)). It took significantly longer to preselect 1-degree targets than 3- (p<.001) and 5-degree targets (p<.001).

5.3 Refinement Time

We define refinement time from the time when the correct target was disambiguated with the cone pointer to a correct selection. Only successful trials were used for this part of the analysis. Outlier trials were filtered out as in preselection time. Results showed no significant 3-way or 2-way interactions. However, all main effects showed significant results. For Technique ($F_{5,55}$=11.33, p<.001, η_p^2=.507), Figure 8), GazeHand was significantly faster than GazeHead (p<.005) and HandHead (p<.005). HeadHand was significantly faster than GazeHead (p<.05) and HandHead (p<.05). These results indicated that the Hand is best suited for final refinement

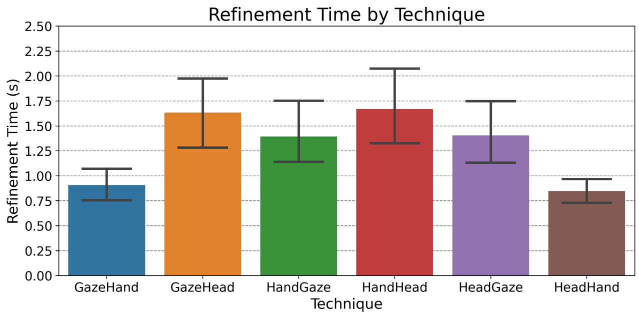

Figure 8: Average refinement time for each technique. The error bars represent 95% confidence intervals of the means.

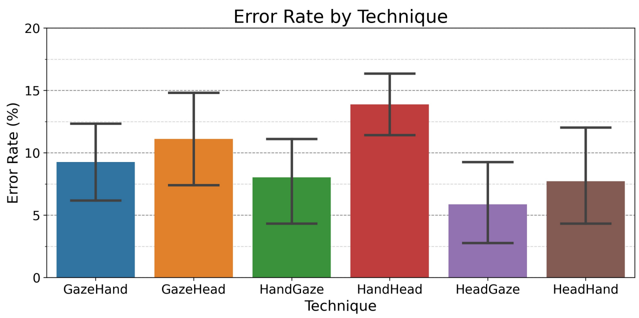

Figure 9: Average error rate for each technique. The error bars represent 95% confidence intervals of the means.

as users can make rapid movements with their hand while keeping visual attention on the target feedback. While users may be able to perform equally or more rapid movements with their eyes, users have to be more careful as to not move their eyes to far from targets which would make it difficult to focus on targets to visually process any feedback. We also found significant results for Depth ($F_{1.17,12.89}$=5.00, p<.05, η_p^2=.313)). Participants were significantly faster in refining to the correct targets in the nearest condition compared to the farthest (p<.001). Finally, the results showed significant differences in the refinement time for the target size ($F_{2,22}$=57.95, p<.001, η_p^2=.840)). A larger target size led to less accuracy required in refinement, leading to faster refinement times.

5.4 Error Rate

The error rate, i.e., the number of trials resulting in an error divided by the total number of trials, was computed to understand the accuracy of the techniques. An error was counted whenever a participant missed the target prior to a correct selection or if a participant failed to select the target before the 10-second trial timeout. As the error rate was positively skewed and violated the repeated measures ANOVA's assumption of normality, the number of errors was used as count data and an "underdispersed" Poisson regression model was fit to the data. We included all main effects in the regression. The analysis found that the overall model was significant ($\chi^2(8,N{=}648)$=21.80, p<.05). An investigation of the model effects revealed a significant main effect for Technique ($\chi^2(5)$=13.58, p<.05). Sequential Šidák pairwise comparisons (Figure 9) showed that HeadGaze had a significantly lower error rate than HandHead (p<.05). We also found a main effect for size ($\chi^2(1)$=.29, p<.05), where for each degree increase in target size, the amount of errors was reduced by 0.905.

5.5 Subjective Feedback

All participants indicated from 1 to 7 on each technique's perceived performance, ease, intuitiveness, and overall preference based on previous work [21]. Performance measured how fast and accurate users perceived the selection technique. Effortlessness rated how little effort was required to use the technique. This includes both the physical and mental fatigue caused by the technique. Intuitiveness refers to how easy it was to learn the technique. The results of the Friedman test showed a significant difference for all questions. For Performance ($\chi^2(5)$=26.55, p<.001, Figure 10a), we found that HeadHand was perceived to be more performant than GazeHead (p<.05) and HandHead (p<.05). Furthermore, we found that GazeHand was perceived as more performant than HandHead (p<.05). Friedman test showed a significant for Intuitiveness ($\chi^2(5)$=20.82, p<.001, Figure 10b). HandHead was considered significantly less intuitive than GazeHand (p<.05) and HeadHand(p<.05). The results also showed for Ease ($\chi^2(5)$=21.86, p<.001, Figure 10c), that GazeHand was significantly easier than GazeHead (p<.05) and HandHead (p<.05). Finally, for overall preference ($\chi^2(5)$=32.91, p<.001, Figure 10d), GazeHand was overall significantly more preferred than GazeHead (p<.05) and HandHead (p<.001). HeadHand was also significantly more preferred than HandHead (p<.05).

6 Discussion

We presented multimodal Cone&Bubble pointing techniques that allow users to, in parallel or in sequence, perform candidate preselection pointing with final fine-grained selection. Our user study results show that users can effectively perform selections with our techniques irrespective of the combination of modality. However, the results also showed clear advantages for specific combinations of modality and also specific roles for certain combinations. In general, gaze-pointing was suited for pre-selection while hand-pointing was best suited for final pointing and selection with the bubble ray.

6.1 Hand-based Pointing

The hand-directed bubble ray performed well with both head- and gaze-controlled cone pointers. The results showed that GazeHand and HeadHand had generally consistent well-performing selection times and error rates. Due to the candidate preselection, participants could perform large pointing movements in the general direction of the intended target, far outside the cone pointer significantly reducing the need for precise pointing. These movements were most effective with the hand as hand movements can be performed completely decoupled from gaze and head movements. This allows users to be more effective in their parallel pointing. However, the participants expressed some negative opinions about hand-directed cone pointer techniques. These opinions were further reflected in the performance data, which showed high error rates and selection times for HandHead and mixed performance by HandGaze. A potential issue could be the subtle use of highlighting for candidate selection rather than an explicit cursor. These techniques could be improved with better visual feedback to show the user the pointing direction of the cone, as some users expressed.

6.2 Head-based Pointing

Head-pointing was generally considered the worst modality for both cone pointing and pointing with the bubble ray. The performance of Head-based techniques was worse than with gaze and controller-based combinations. Furthermore, head-pointing was seen as more demanding and less intuitive compared to gaze- and hand-pointing. The poor suitability of the head for the bubble ray could be explained by the fact that large movements to fully leverage the increased target width enabled by preselection could not be performed, as these are cumbersome and may also significantly shift the field of view, which may confuse users. Especially when combined with gaze, significant shifts in the field of view may affect eye tracking quality and users'

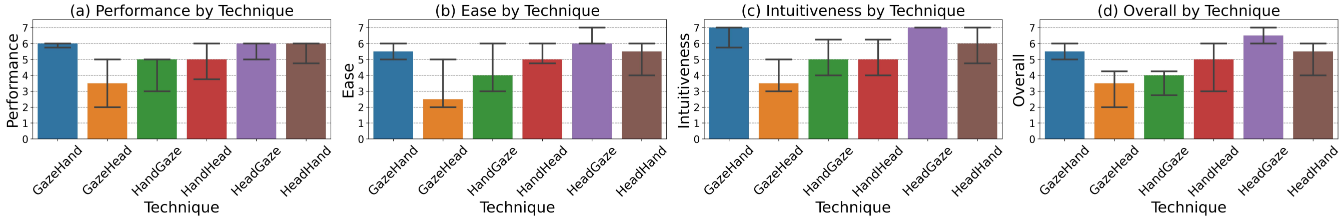

Figure 10: Median values of the subjective feedback from participants. Error bars represent the interquartile ranges of the medians.

ability to properly focus in the correct direction. However, the candidate selection performed by the head seems to play a more suitable role for this modality. HeadHand results showed good overall performance, whereas HeadGaze seems to be a slow but accurate method of selection. The larger range of the cone-pointer might favour the coarse and approximate pointing style of the head.

6.3 Gaze-based Pointing

Gaze-based bubble ray pointing showed unique results in terms of both performance and feedback. The selection time is never low enough to be competitive with the hand-pointing techniques. However, HeadGaze and HandGaze often outperform other techniques in terms of accuracy. Participants may have felt the need to take more time during selection when using their eyes, leading to slower but less error-prone selections. A possible explanation could be that users sometimes moved their gaze beyond the target border to leverage the bubble ray mechanism. However, they still needed to pay attention to the target, as feedback was indicated by colour change. As such, it is reasonable to conclude that techniques which require pointing outside the target border are less suitable for gaze in comparison to other modalities that can move more independently to our visual system. As such, similarly to head tracking, gaze movement can be applied successfully as a candidate selector. GazeHand was the most preferred technique by participants and one of the best techniques in terms of data. The low performance of GazeHead further reinforces the theory that head movement is better for pre-selection rather than final selection, as showcased by the low preselection time enabled by gaze. Additionally, natural eye-head coordination could have hampered GazeHead, as head movements could lead to uncomfortable eye positions, thus leading to less comfort and performance.

6.4 Limitations and Future Work

The study presented in this paper only addresses selection environments with stationary and nonoccluded targets. This does not simulate more realistic scenarios, where targets could overlap one another or even situations with moving targets, as has been investigated in prior work [21]. Future studies could study these scenarios and propose techniques that leverage parallel movements for even more challenging selection contexts.

Some participants faced issues related to the visual design of the techniques. Participants expressed frustrations with the hand-directed cone pointer, as it did not include a visualisation of the direction in which it was casting. On the other hand, the hand-directed bubble ray involved a visual ray emitting from the controller. The decision not to include a ray visualiser for this modality was made to maintain consistency in the visual design of the cursors, but in hindsight it is clear that this made the techniques less intuitive and more difficult to use. This likely impacted the performance data of the two techniques that involved hand-based cone casting. This was not an issue for gaze- and head-based cone pointing, as users could easily point in the general direction without any cursor-based visual feedback, further highlighting their suitability for the role of coarse preselection.

Finally, another consideration to be made is the colour design of the targets. There were two aspects of the design of the environment that occasionally caused confusion for participants: firstly, two participants expressed that they had gotten confused about which target was the final target they needed to select. Second, the colour of the targets inside the candidate selection was changed by darkening the default colour of the object. For techniques where users wanted to move their gaze away from targets (e.g. gaze-based bubble ray), and where the gaze ended up in the periphery, clearer target visualisation would be useful.

7 Conclusion

In this paper, we presented multimodal Cone&Bubble techniques in which two modalities are combined for coarse-grained preselection and final selection through area cursors. Our techniques allow users to perform these pointing tasks in parallel to increase the efficiency of movement and performance. We compared different permutations of our proposed techniques and investigated gaze, head, and controller-based hand-pointing for each role. The results showed that users could effectively select targets with all techniques. Furthermore, results showed that the hand is most suited for final selection due to its decoupling from gaze and head. Gaze was considered best for effortless pre-selection, while head pointing could be a useful alternative when eye tracking is not available. Our work shows how techniques can be developed for parallel movement of modalities rather than sequential as commonly seen in previous work, and opens up future design opportunities.

Acknowledgments

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 101021229, GEMINI: Gaze and Eye Movement in Interaction).

References

[1] D. Aliakseyeu, M. A. Nacenta, S. Subramanian, and C. Gutwin. Bubble radar: Efficient pen-based interaction. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '06, p. 19–26. Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1133265.1133271 2

[2] C. Appert, O. Chapuis, and E. Pietriga. High-precision magnification lenses. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, p. 273–282. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1753326.1753366 2

[3] F. Argelaguet and C. Andujar. A survey of 3d object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, 2013. doi: 10.1016/j.cag.2012.12.003 2

[4] R. Atienza, R. Blonna, M. I. Saludares, J. Casimiro, and V. Fuentes. Interaction techniques using head gaze for virtual reality. In *2016 IEEE Region 10 Symposium (TENSYMP)*, pp. 110–114, 2016. doi: 10.1109/TENCONSpring.2016.7519387 2

[5] H. Benko, A. D. Wilson, and P. Baudisch. Precise selection techniques for multi-touch screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, p. 1263–1272.

Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1124772.1124963 2

[6] D. A. Bowman, D. B. Johnson, and L. F. Hodges. Testbed evaluation of virtual environment interaction techniques. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '99, p. 26–33. Association for Computing Machinery, New York, NY, USA, 1999. doi: 10.1145/323663.323667 2

[7] G. Casiez, N. Roussel, and D. Vogel. 1 € filter: A simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, p. 2527–2530. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2207676.2208639 3

[8] M. Choi, D. Sakamoto, and T. Ono. Bubble gaze cursor + bubble gaze lens: Applying area cursor technique to eye-gaze interface. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA '20 Full Papers. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3379155.3391322 2

[9] W. Delamare, M. Daniel, and K. Hasan. Multifingerbubble: A 3d bubble cursor variation for dense environments. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491101.3519692 2

[10] 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 *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 1118–1130. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025599 2

[11] A. S. Fernandes, T. S. Murdison, and M. J. Proulx. Leveling the playing field: A comparative reevaluation of unmodified eye tracking as an input and interaction modality for vr. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2269–2279, 2023. doi: 10.1109/TVCG.2023.3247058 2

[12] A. Forsberg, K. Herndon, and R. Zeleznik. Aperture based selection for immersive virtual environments. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*, UIST '96, p. 95–96. Association for Computing Machinery, New York, NY, USA, 1996. doi: 10.1145/237091.237105 2

[13] 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. doi: 10.1145/1054972.1055012 1

[14] T. Grossman and R. Balakrishnan. The design and evaluation of selection techniques for 3d volumetric displays. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, p. 3–12. Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1166253.1166257 2

[15] J. P. Hansen, V. Rajanna, I. S. MacKenzie, and P. Bækgaard. A fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In *Proceedings of the Workshop on Communication by Gaze Interaction*, COGAIN '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3206343.3206344 2

[16] S. Jalaliniya, D. Mardanbeigi, T. Pederson, and D. W. Hansen. Head and eye movement as pointing modalities for eyewear computers. In *2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops*, pp. 50–53, 2014. doi: 10.1109/BSN.Workshops.2014.14 2

[17] P. Kabbash and W. A. S. Buxton. The "prince" technique: Fitts' law and selection using area cursors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, p. 273–279. ACM Press/Addison-Wesley Publishing Co., USA, 1995. doi: 10.1145/223904.223939 2, 3

[18] R. Kopper, F. Bacim, and D. A. Bowman. Rapid and accurate 3d selection by progressive refinement. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 67–74, 2011. doi: 10.1109/3DUI.2011.5759219 2

[19] R. Kopper, D. A. Bowman, M. G. Silva, and R. P. McMahan. A human motor behavior model for distal pointing tasks. *International Journal of Human-Computer Studies*, 68(10):603–615, 2010. doi: 10.1016/j.ijhcs.2010.05.001 2

[20] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise head- and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173655 1, 2

[21] Y. Lu, C. Yu, and Y. Shi. Investigating bubble mechanism for ray-casting to improve 3d target acquisition in virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 35–43, 2020. doi: 10.1109/VR46266.2020.00021 2, 3, 5, 6

[22] K. Minakata, J. P. Hansen, I. S. MacKenzie, P. Bækgaard, and V. Rajanna. Pointing by gaze, head, and foot in a head-mounted display. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3317956.3318150 2

[23] D. R. Olsen and T. Nielsen. Laser pointer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, p. 17–22. Association for Computing Machinery, New York, NY, USA, 2001. doi: 10.1145/365024.365030 1

[24] K. Pfeuffer, J. Alexander, M. K. Chong, and H. Gellersen. Gaze-touch: Combining gaze with multi-touch for interaction on the same surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, p. 509–518. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2642918.2647397 2

[25] K. Pfeuffer, J. Alexander, M. K. Chong, Y. Zhang, and H. Gellersen. Gaze-shifting: Direct-indirect input with pen and touch modulated by gaze. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, p. 373–383. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2807442.2807460 2

[26] J. S. Pierce, A. S. Forsberg, M. J. Conway, S. Hong, R. C. Zeleznik, and M. R. Mine. Image plane interaction techniques in 3d immersive environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, I3D '97, p. 39–ff. Association for Computing Machinery, New York, NY, USA, 1997. doi: 10.1145/253284.253303 2

[27] Y. Y. Qian and R. J. Teather. The eyes don't have it: An empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, p. 91–98. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3131277.3132182 2

[28] J. Schjerlund, K. Hornbæk, and J. Bergström. Ninja hands: Using many hands to improve target selection in vr. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445759 2

[29] 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 *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3580685 2

[30] L. Sidenmark, C. Clarke, X. Zhang, J. Phu, and H. Gellersen. Outline pursuits: Gaze-assisted selection of occluded objects in virtual 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. doi: 10.1145/3313831.3376438 2

[31] L. Sidenmark and H. Gellersen. Eye&head: Synergetic eye and head movement for gaze pointing and selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, p. 1161–1174. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3332165.3347921 2

[32] H. Skovsgaard, J. C. Mateo, J. M. Flach, and J. P. Hansen. Small-target selection with gaze alone. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, p. 145–148. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1743666.1743702 2

[33] S. Stellmach and R. Dachselt. Look & touch: Gaze-supported target acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, p. 2981–2990. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2207676.2208709 2

[34] S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (magic) pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, p. 246–253. Association for Computing Machinery, New York, NY, USA, 1999. doi: 10.1145/302979.303053 2