

Multimodal Feedback in 3D Target Acquisition

Dalia El-Shimy*

Georgios Marentakis[†]

Jeremy R. Cooperstock[‡]

Centre for Interdisciplinary Research in Music, Media, and Technology
McGill University
Montreal, Canada

ABSTRACT

We investigated dynamic target acquisition within a 3D scene, rendered on a 2D display. Our focus was on the relative effects of specific perceptual cues provided as feedback. Participants were asked to use a specially designed input device to control the position of a volumetric cursor, and acquire targets as they appeared one by one on the screen. To compensate for the limited depth cues afforded by 2D rendering, additional feedback was offered through audio, visual and haptic modalities. Cues were delivered either as discrete multimodal feedback given only when the target was completely contained within the cursor, or continuously in proportion to the distance between the cursor and the target. Discrete feedback prevailed by improving accuracy without compromising selection times. Continuous feedback resulted in lower accuracy compared to discrete. In addition, reaction to the haptic stimulus was faster than for visual feedback. Finally, while the haptic modality helped decrease completion time, it led to a lower success rate.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Auditory (non-speech) feedback, Evaluation/Methodology, Haptic I/O, Input Devices and Strategies

1 INTRODUCTION

3D environments are gaining popularity beyond the fields of animation, graphic design or computer games. Facilitated by advances in computer graphics hardware and the development of visual languages and symbologies that work in 3D [9], everyday users now have access to and demand for simple applications that support 3D displays rendered in a 2D setting, and supplemented with depth cues. A common example is that of Google Earth, which allows users to navigate satellite images of the Earth's surface along longitude, latitude and height. In addition, many museums are now providing 3D virtual displays to enable visitors to examine high-resolution stereo models of valuable artifacts. However, elementary tasks that are common in 2D displays, such as selection, navigation, data input and manipulation cannot simply be imported into a 3D scene. For instance, a point-and-click method using a 2D arrow cursor would not be suitable for a volumetric target: an arrow and target may appear to overlap one another, even if there is a difference in depth between their positions. In addition, since a sphere rendered in 2D simply appears to be a disc, it is difficult to visually gauge how far it extends into the screen. Furthermore, most current 3D displays are only effective for precise navigation and positioning in specific circumstances, for instance, when good viewing angles or relevant positional cues such as shadows are available [7]. Even in the presence of such advantages, limits of screen dimension or resolution can prevent them from being effective. As a result, significant modifications of the content are often required

when rendering 3D scenes in 2D on devices with limited display capabilities, such as cell phones.

These observations motivate the development of improved paradigms for 3D object instantiation, positioning, and other parameter control, which effectively capitalize on our perceptual capabilities. This problem can be divided into the actions or gestures that the user needs to perform, the choice of sensor to acquire these input gestures, and appropriate feedback to indicate to the user what input has been recognized and/or what action has been performed. Our study investigates the comparative performance benefits of graphical, auditory, and haptic feedback, offered individually or in combination, when working in a 3D environment. As a starting point for such an effort, we consider the basic task of object or target "selection" as the necessary initial action required for almost all others.

2 RELATED WORK

2.1 3D Target Acquisition

Selection in 3D virtual reality can be accomplished through one of two techniques: using 3D point cursors that directly map the user's hand to the location of a 3D ray cursor, known as "hand extension", or by means of aperture-based selection techniques that project from the user's hand a ray that intersects targets for selection. A well-known example of the latter is the "laser gun" [5]. Ray cursors have been shown to result in faster reaction times [2], but some problems with this technique arise in dense environments where the ray might intersect multiple targets. Although a number of ways have been proposed to overcome this problem, there is no clear consensus on a solution that does not require a two-step selection process, the latter of which is to disambiguate depth cues. An alternative way to pointing in 3D environments is to use volumetric cursors, which use the 3D point technique with a volume rather than a point. This was found to be particularly useful for overcoming problems that arise when small targets are to be selected. In addition, this method is unaffected by the target density of the environment, since the 3D cursor is mapped in a unique way to the virtual reality coordinate system. However, a problem with this approach is that depth cues are confusing when the cursor and the target overlap, and it might be difficult for the user to decide when the cursor is completely encircling the target. Zhai et al. [10] refined this selection technique by introducing the silk cursor as a volume with semi-transparent sides that can provide partial occlusion when intersecting or positioned in front of a target. As a result, users may employ depth cues to gauge the relative position of the cursor and the target. Selection is performed by enveloping the target within the cursor volume. The authors conducted a series of user tests to compare target acquisition using a volumetric cursor with and without semi-transparent cues, in both stereoscopic and monoscopic display conditions. The silk cursor technique was found to be more significant than stereo in aiding task performance; however, the authors noted that the lack of continuous depth cues influenced the magnitude of errors.

Grossman and Balakrishnan [4] investigated pointing to 3D targets in 3D virtual environments. A classic selection approach, a small cursor and a large target displayed as a 3D wireframe, were

*e-mail: dalia@cim.mcgill.ca

[†]e-mail: gmarentakis@music.mcgill.ca

[‡]e-mail: jer@cim.mcgill.ca

used throughout the experiment. The task applied a reciprocal Fitts' law, with participants alternating selections between two targets, constrained to the xz -plane, and oriented in the direction of movement. Target depth and width, direction of movement, and distance to target all had a significant effect on trial completion time. Movements in the z -axis were found to be the most error-prone.

2.2 Feedback in Target Acquisition

Akamatsu et al. [6] performed a study to compare various combinations of tactile, auditory and visual feedback in a pointing task using a mouse-like device in a 2D selection task on a computer screen. They found no differences in overall response times or error rates, but significant variations in final positioning times, with tactile feedback offering the fastest performance, and visual the slowest. While the task was performed in a traditional 2D setting using a modified mouse, the authors provide a valuable overview and comparison of feedback modalities and their effects on user performance. Such concepts can subsequently be adapted to fit target selection within 3D scenes rendered in 2D (environments typically lacking in sufficient depth cues), and using input devices more suitable than the conventional mouse.

Cockburn and Brewster investigated multimodal feedback for the acquisition of small targets [1]. They asked users to select two thin vertical bars, in a reciprocal Fitts' law task, in eight conditions: *audio* (an audio cue was given when users were on target), *tactile* (the mouse would vibrate at 200 Hz when users were on target), *sticky* (the mouse control-display gain would reduce to 1/20th of its original value when the cursor was on target), in addition to four possible combinations of the above and a control condition with no feedback provided. Performance was fastest in the tactile plus sticky condition, and slowest in the normal condition. Mouse-over-target times were reduced as much as 15% with tactile and 9% with audio feedback. Stickiness also reduced overshooting. A follow-on experiment investigated target selection from a hierarchical menu using the same feedback cues. In contrast with the previous results, both selection times and mouse-over-target times increased with the provision of feedback cues. The authors suggest that this may be due to the presence of multiple targets, which therefore led to distracting noise.

Haptic feedback in the form of varying vibrotactile stimulation of the hand can provide cues for subtle changes in position. Oron-Gilad et al. [8] examined this within the context of guiding an operator toward a target. Tactile stimulation was provided in response to deviation from the point of origin, indicating the direction and magnitude of the positional error. Improved user performance was observed when both tactile and visual cues were present, indicating that both forms of feedback had minimal cross-modal interference.

2.3 Synthesis of the Literature Review

While Akamatsu et al. [6] observed the improvement in performance provided by multimodal feedback and Zhai et al. [10] studied the role of graphical feedback in conveying depth cues, we are interested in examining the relevant effects of the various modalities in a 3D target acquisition task, using an input device that is suitable for 3D interaction, and to test whether such feedback offers benefits beyond the Silk Cursor. In particular, we would like to determine whether feedback is more important when depth is considered, as this is a context in which users tend to have the greatest difficulty, and whether the importance of feedback is affected by the axis of movement, as these questions have not been addressed previously. Most importantly, as previous experiments comparing different feedback modalities have been conducted entirely with discrete feedback provided when the subject was on-target, we feel it is critical to verify whether the same results apply to the case of continuous feedback. This condition is relevant for instances where the computer does not "know" the intended target position and thus,

is unable to take advantage of other well-known techniques such as semantic pointing [3] for drawing the user more efficiently to the target. The set of experiments conducted are described below.

3 PRELIMINARY EXPERIMENT

The preliminary experiment largely repeated Zhai's Silk Cursor experiment but with a pointing device instead of a glove. This was done primarily to provide a wireless input device and one that would convey to the user the vibrational feedback as needed in the follow-up experiment, described in the next section. Users would confirm the selection by pressing a button on the Wiimote. Moreover, our choice of input device was justified by our interest in analyzing the more basic operation of 'pointing' rather than 'grasping'.

3.1 Apparatus and Materials

The virtual environment was rendered using OpenSceneGraph in a cave-like environment. A Nintendo Wiimote fitted with four Vicon infrared reflectors was used as an input device to control the cursor with 3-DOF positioning (no orientation control). The position of the Wiimote was tracked by the Vicon system and mapped to the position of the cursor on the screen. One 'unit' of cursor displacement in the virtual environment was designed to correspond to a 10 cm real-world displacement of the input device.

3.2 Method

The independent variables were the direction of movement x , y , z , and the type of cursor, as described below. A cursor with side length of 0.5 units was used throughout.

1. Wireframe cursor: the cursor is represented as a cube with completely transparent sides, with only the edges being visible (Figure 1(a)).
2. Silk cursor: A cube with semi-transparent sides is used. Users can make use of occlusion cues to gauge the position of the cursor in relation to the target (Figure 1(b)).

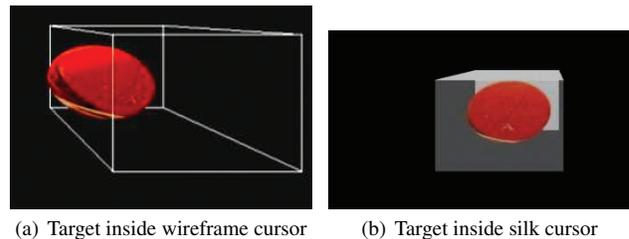


Figure 1: Preliminary experiment display examples.

For each cursor type, users completed four trial sets of twelve target acquisitions. Targets were spheres with a diameter of 0.25 units, located at the corners of a cube with sides three units long (for a total of eight possible locations). They appeared successively, with each new target offset relative to the previous one along a single edge of the cube. A script generated a different target sequence for each trial block and for each participant, such that an equal number of movements along each axis x , y , z and direction of movement (positive/negative) was performed. Overall, 16 repetitions for each plane of movement and cursor type were performed per participant.

The dependent variables were trial completion time and accuracy of the task, defined as the percentage of trials where users completely encapsulated the target.

3.3 Participants

For this study, 16 users, with a mean age of 23, all university students in either Engineering or Music Technology, participated.

3.4 Results

Analysis of user performance reveals that there is a tendency for the silk cursor to decrease completion time, while improving accuracy (see Figures 2(a) and 2(b)). A cursor (2) x -axis of movement (3) ANOVA on the log (base 10) transformed time to complete trials. The log transformed was used so that data would follow the normal distribution. There was a significant main effect of cursor $F(1,15) = 44.621, p < 0.001$ and axis of movement $F(2,30) = 10.015, p < 0.001$ on time to complete trials consistent with the results demonstrated by Zhai et al. [10]. Percent correct data did not follow the normal distribution and were therefore analyzed using a non-parametric Friedmann test, appropriate for within-subjects designs. There was a significant main effect of cursor type, $\chi^2(1,32) = 16.2, p < 0.001$, but no effect of movement plane, $\chi^2(1,32) = 16.2, p < 0.001$.

Post-hoc Bonferroni confidence interval adjusted comparisons showed that trial completion time was greatest in the z -direction, indicating that perception of target depth was limited, but there was no difference between x and y directions of movement.

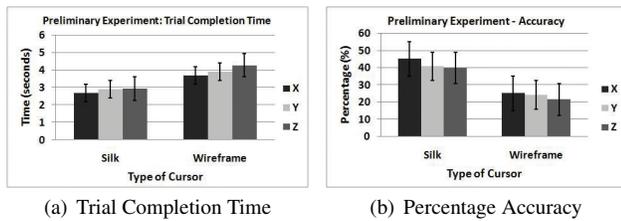


Figure 2: Preliminary experiment results (with error bars)

4 EXPERIMENT: DISCRETE VS. CONTINUOUS FEEDBACK

In the second round of experiments, we examine whether using multimodal feedback in addition to the silk cursor can improve user performance. We use two types of feedback: one where discrete feedback is given only when an acquisition is deemed successful by the system, and a second where continuous feedback with respect to the Euclidean distance to the target 'area' is given. The wireframe cursor was altogether omitted, since it was shown to be suboptimal to the silk cursor. Our intention is to understand user behaviour in these two fundamentally different feedback modes. To further increase acquisition accuracy along the z -axis, cubes rather than spheres were used as targets, since their depth is easier to perceive. This also helped provide better consistency with the cursor's shape. In addition, as ray cursors have been shown to improve performance [2], an extension was projected from the end of the cursor away from the screen, and the input device was modified to provide 5-DOF (rotation about the z -axis was ignored), thereby allowing users to feel as though they were directly manipulating a wand-like device (Figure 3).

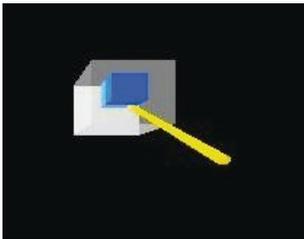


Figure 3: Modified Cursor and Target

4.1 Method

This experiment included one extra variable, as feedback type was either discrete or continuous. The feedback modes for each

case are described below. The experiment followed the same design as the preliminary experiment, with four blocks of 12 target acquisitions for each level of the variable feedback modality. Overall, 16 repetitions for each case of plane of movement and feedback modality were performed per participant. The dependent variables were the same as the preliminary experiment and the independent variables enumerated below.

Discrete Feedback Modalities

1. Visual: The silk cursor was used, providing the same partial occlusion cues as in the preliminary experiment, but the target colour also changed from blue to green when it was completely contained within the cursor.
2. Auditory: A 160 Hz sine wave was played when the target was completely contained within the cursor.
3. Haptic: The Wiimote vibrated once the target was completely contained within the cursor.

Continuous Feedback Modalities

1. Visual: The silk cursor was used, and users were able to make use of the partial occlusion cues it affords to determine whether a target was contained within it.
2. Auditory: A 160 Hz sine wave, with fixed duration of 200 ms was played on and off in accordance with the pulse train shown in Figure 4. At any given time, the duration of the 'on' intervals, x , was directly proportional to the distance, d , between the cursor and the target at that point. Thus, as the cursor became closer to the target, the sound was played more intermittently, and appeared to be continuous once the target was completely within the cursor.

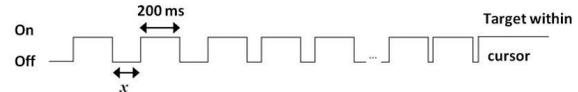


Figure 4: Pulse Train used in Continuous Feedback Experiment.

3. Haptic: The Wiimote vibrated on and off in accordance with the pulse train seen in Figure 4 as well, giving the impression that it vibrated continuously once the target was within the cursor.

4.2 Participants

In total, 24 participants were tested (12 for the continuous feedback case, and 12 for the discrete one): 20 males and 4 females, with a mean age of 26. All were university students. There was no overlap with participants in the first experiment.

4.3 Results

In comparison to the preliminary experiment, we note that the use of additional feedback modalities decreased completion time while increasing accuracy, both in the continuous and discrete cases. In particular, error along the z -axis was reduced, implying that using various perceptual cues can indeed compensate for limited depth perception. The issue then remains to examine whether such feedback should be discrete or continuous.

A feedback modality (3) \times axis of movement (3) \times feedback type (2) ANOVA was performed on the log transformed (base 10) trial completion times. The difference between time to complete selections between the discrete and continuous cases was not significant, $F(1,22) = 3.091$ & $p = 0.093$. There was a significant main effect of feedback modality and axis of movement on selection times, $F(2,44) = 11.515$ & $p < 0.001$, and $F(2,44) = 27.702$ & $p < 0.001$ respectively. Vibrational cues decreased both accuracy and completion time significantly with no difference between

the other two cues, as shown by Bonferroni confidence interval adjusted post-hoc tests. This indicates that users were more confident with and responded faster to the haptic stimulus, even though they were slightly off-target at selection time. The Wiimote’s vibration could have shifted the user’s hand slightly away from the correct position as he or she pressed the button for acquisition. Movements in the z axis were slowest, followed by movements along the y and x axes, all differences being significant based on Bonferroni confidence interval adjusted post-hoc tests. The faster performance using the vibro-tactile cue is consistent with the findings of [6].

Accuracy of selection did not follow the normal distribution, therefore a non-parametric (Kruskal-Wallis) test was performed. There was a significant main effect of variable feedback type on the percentage of trials completely encapsulating the target, $\chi^2(1, 216) = 90.712$ & $p < 0.001$ *. In the discrete case, there was a significant main effect of feedback modality ($\chi^2(2, 36) = 19.703$ & $p < 0.001$ (Friedmann) and movement axis $\chi^2(2, 36) = 7.197$ & $p = 0.05$ (Friedmann). Post-hoc Wilcoxon comparisons showed that tactile was less accurate than auditory and color ($z(36) = -3.009, p = 0.03, z(36) = -4.149, p < 0.001$) and also that color resulted in higher success rate than auditory ($z(36) = -2.835, p = 0.005$). In addition, higher success rates were found for movements in the x -axis compared to y and z , ($z(36) = -2.2, p = 0.028, z(36) = -2.18, p = 0.029$), but no difference was observed between z and y . No effect of the independent variables in the continuous feedback cases were observed.

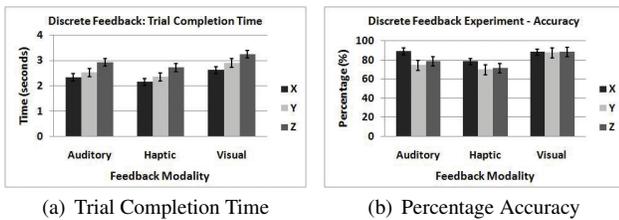


Figure 5: Results for Discrete Experiment (with error bars)

We find that trial completion times are quite comparable for both the discrete and continous cases. This is quite surprising, seeing as we would expect users to require more time when relying on their own judgement rather than the computer’s. In addition, we observed that as trials progressed, users began to use continuous feedback in a manner similar to discrete: they became accustomed to how a feedback modality should behave when the target was completely within the cursor, and thus simply moved the cursor using visual cues until they perceived the cue in an expected form. For instance, in the auditory case, instead of gauging the distance by comparing the differences in the sound signal emitted at various locations, users would perform a ballistic motion in the direction of the target and once a continual sound signal was heard, they would click for acquisition. There is a plausible explanation for this phenomenon: it is common knowledge that certain neural mechanisms are optimized for visual servoing, and thus, we are likely to exploit these first, before taking advantage of cues provided by other modalities. This is consistent with the observed behaviour of subjects using additional cues (in the continuous case) only for fine adjustments. We note, however, that there is a tendency for higher accuracy in the discrete case (Figure 5(b) vs. Figure 6(b)). This suggests that users were not able to perceive the shorter temporal gaps correctly as the cursor approached the target. Individual differences may also have played a role in these results.

5 CONCLUSIONS

A series of experiments were conducted to assess the benefit of various perceptual cues on the task of target selection. Results imply

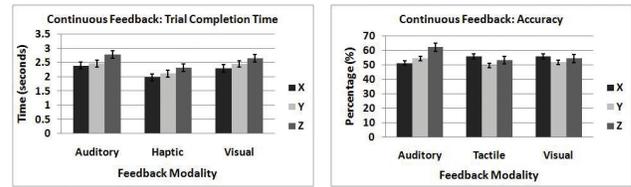


Figure 6: Results for Continuous Experiment (with error bars)

that subject performance improved when various perceptual cues were provided to indicate the possibility of successful acquisition. Interestingly, however, continuous feedback was not successful; users were less accurate in comparison to the discrete case and they tended to use continuous feedback in a similar manner as discrete.

This study has provided the basis for an ecologically valid, unobtrusive and easily understood paradigm for 3D target selection. Future work includes testing additional variations of the feedback modalities, moving to a stereoscopic display, and improving the design of the input device accordingly.

ACKNOWLEDGEMENTS

The authors would like to thank Mike Wozniowski for his invaluable help throughout, François Bérard for sharing his vast knowledge on related work, and Ronak Mallik for his participation in setting up this project. The research described here was supported by the Natural Sciences and Engineering Research Council of Canada and les Fonds de recherche sur la nature et les technologies.

REFERENCES

- [1] A. Cockburn and S. Brewster. Multimodal feedback for the acquisition of small targets. *Ergonomics*, 48(9):1129–1150, 2005.
- [2] J. L. D. Bowman, E. Kruijff and I. Poupyrev. An introduction to 3d user design interface. *Volume 10, 1*, pages 96–108, 2001.
- [3] N. Elmqvist and J.-D. Fekete. Semantic pointing for object picking in complex 3d environments. *Proceedings of graphics interface*, pages 243–250, 2008.
- [4] T. Grossman and R. Balakrishnan. Pointing at trivariate targets in 3d environments. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 447–454, New York, NY, USA, 2004. ACM.
- [5] J. Liang and M. Green. A highly interactive 3d modeling system. *Computer and Graphics*, pages 499–506, 1994.
- [6] T. H. M. Akamatsu, I.S. MacKenzie. A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics*, 38:816–827, 1995.
- [7] S. A. M. Tory, A. E. Kikpatrick and T. Moller. Visualization task performance with 2d, 3d, and combination displays. *IEEE Transactions on Visualization and Computer Graphics*, pages 2–13, 2006.
- [8] R. G. T. Oron-Gilad, J.L. Downs and P. Hancock. Vibrotactile guidance cues for target acquisition. *IEEE Transactions on Systems, Man and Cybernetics*, 37:993–1004, 2007.
- [9] C. Ware and G. Franck. Viewing a graph in a virtual reality display is three times as good as a 2d diagram. *IEEE Symposium on Visual Languages*, pages 182–183, 1994.
- [10] S. Zhai, W. Buxton, and P. Milgram. The “silk cursor”: investigating transparency for 3d target acquisition. In *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 459–464, New York, NY, USA, 1994. ACM.