Interaction Design of Dwell Selection Toward Gaze-based AR/VR Interaction

Toshiya Isomoto
isomoto@iplab.cs.tsukuba.ac.jp
University of Tsukuba
Tsukuba, Ibaraki, JAPAN

Shota Yamanaka
syamanak@yahoo-corp.jp
Yahoo Japan Corporation
Chiyoda, Tokyo, JAPAN

Buntarou Shizuki
shizuki@cs.tsukuba.ac.jp
University of Tsukuba
Tsukuba, Ibaraki, JAPAN

ABSTRACT

In this paper, we first position the current dwell selection among gaze-based interactions and its advantages against head-gaze selection, which is the mainstream interface for HMDs. Next, we show how dwell selection and head-gaze selection are used in an actual interaction situation. By comparing these two selection methods, we describe the potential of dwell selection as an essential AR/VR interaction.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; *Empirical studies in interaction design.*

KEYWORDS

Gaze-based interface, target selection, intent prediction, Midas-touch problem

1 INTRODUCTION

Dwell selection has been the mainstream method since the gaze-based interface was proposed [Jacob 1990]. Although numerous researchers have put their efforts into dwell selection to solve the critical issue of unintended selection, which is referred to as the Midas-touch problem [Jacob 1990], dwell selection still faces this issue. Dwell selection for a target requiring much time to understand its content (e.g., a target consisting of a text and image) is not always achievable [Zhang et al. 2011], particularly in contrast to the successful achievement of dwell selection for a simple target. When selecting such non-simple targets, it is difficult to predict whether the user intends to select a target or not with only the time threshold as the required time varies depending on the content. Although many studies have pointed out the advantages of dwell selection highly appreciated for AR/VR interaction, such as "easy to use" and "straightforward," they also reported that users generally dislike this selection method as a result of the frequent occurrence of Midas-touch problems (e.g., [Lu et al. 2021]).

Many selection methods that uses a combination of gaze and other modalities have been proposed to solve the Midas-touch problem. Specifically, with the recent development of an HMD with a built in eye-tracker, a selection method using a combination of gaze and head-rotation (i.e., the direction of the head) has been researched. This method is referred to as head-gaze selection, in which users can select a target by dwelling on the button while their head is directed toward the target (e.g., [Lu et al. 2020; Lutteroth et al. 2015]). Given that no interaction is performed by dwelling on a target, a few Midas-touch problems occur. Even though head-gaze selection has better results than current dwell selection against the occurrences of the Midas-touch problem, the potential of dwell selection, in terms of easy and straightforward selection, remains attractive.

In this paper, we show an interaction design of dwell selection toward gaze-based AR/VR interaction and describe the potential of dwell selection.

2 INTERACTION DESIGN OF DWELL SELECTION

We show the use of dwell selection and head-gaze selection and compare the two methods to show the potential of dwell selection to be an essential AR/VR interaction.

With dwell selection (Figure 1, top), a user can display food information by dwelling on it and close the display by dwelling on a button. In the example, there are many possibilities where the Midas-touch problem can occur when selecting food and beverages because their content, such as price and nutrition, depends on each item. Accordingly, the user must take much time to understand the content, and there are possibilities for causing the Midas-touch problem. Occurrences of the Midas-touch problem could irritate users due to the frequent unwanted display of information. Recently, a few studies have begun developing the robustness of dwell selection against the Midas-touch problem with machine-learning based intent prediction (e.g., [Bednarik et al. 2012; David-John et al. 2021]). Although such dwell selection has a great possibility of preventing the Midas-touch problem, improving the prediction performance is still necessary for robust selection. We should use a dwell selection with intent prediction for non-simple targets due to the necessity of robustness against the Midas-touch problem and use a dwell selection with only a time threshold for a simple target because robustness has already been achieved.

With head-gaze selection (Figure 1, bottom), the contrast to dwell selection are the separation of target and buttons and the buttons' position. Most buttons are consistently positioned at the outermost of the user's field of view (FOV) to avoid the Midas-touch problem. That is, the selection is performed by first moving the gaze onto a
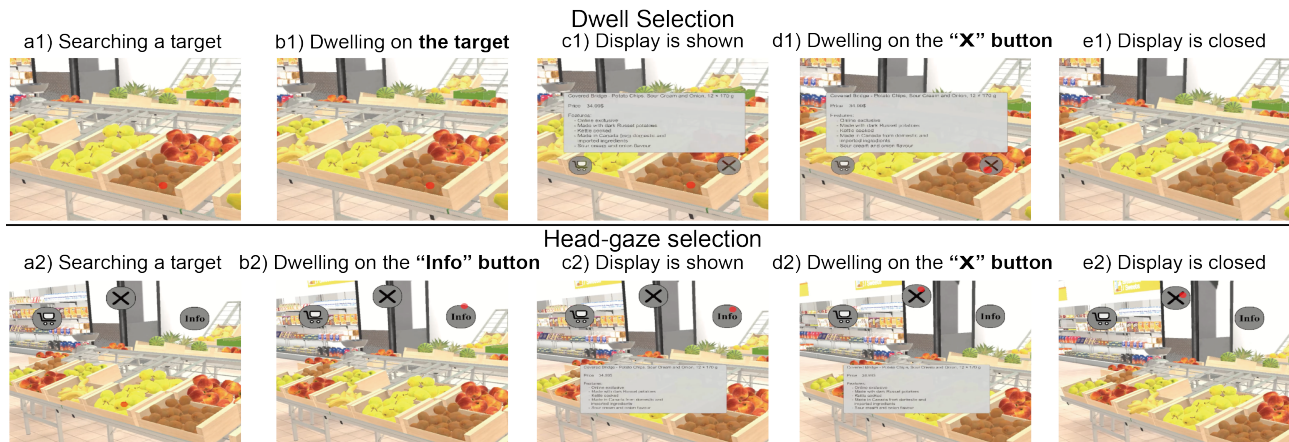
Figure 1: Schema of two selection methods in a shopping situation. These images are made by cropping the original image almost into the user's field of view (FOV). (Top) The user of dwell selection can show information and corresponding buttons by dwelling on a target (a1–c1) and close them by dwelling on the "X" button (d1 and e1). (Bottom) The user of head-gaze selection can show information by dwelling on the "info" button positioned at the outermost of the FOV (a2–c2) and close it by dwelling on the "X" button positioned at the outermost of the FOV (d2 and e2). The red point shows the gaze position of the user.

button positioned at the outermost of the FOV and then dwelling on it. In Figure 1, selecting a button shows and closes the information mapped into the button. Even though head-gaze selection requires additional eye movement and buttons, it achieves high usability [Lu et al. 2020] and thus has been the promised interface for an HMD.

Having no need for additional eye movement increases dwell selection's potential. For example, one important potential is smaller physical demand. Although head-gaze selection requiring additional eye movement and buttons achieves high usability [Lu et al. 2020], moving the eyes toward the outermost of the FOV could be physically demanding for users [Sidenmark and Gellersen 2019]. With dwell selection, a user can select a target just by dwelling on the target (Figure 1b1), followed by searching for a target (Figure 1a1) without additional eye movement. Another potential is the ability of free interface design. The outermost of the FOV that may affect the user's attention least is useful for stacking notifications and displaying status bars in the same manner as in desktop interfaces. Also, dwell selection is helpful for combined interactions with the mouse, touch, and voice. These combined interactions use dwell selection to determine the user's intent in the same manner as the principle of dwell selection of "what the user looks at is what the user wants." Thus, dwell selection with precise intent prediction will be essential for a gaze-assisted interface.

3 CONCLUSION

We showed an interaction design of dwell selection toward gaze-based AR/VR interaction, describe the potentials of dwell selection, and pointed out two factors requiring dwell selection to be such an interface. One is improving the user's intent prediction for dwell selection without any additional eye movement and buttons, and the other is investigating the performance of dwell selection with a target that requires much time to understand its content. We believe that the proposed dwell selection will become an essential gaze-based AR/VR interaction.

REFERENCES

Roman Bednarik, Hana Vrzakova, and Michal Hradis. 2012. What Do You Want to Do Next: A Novel Approach for Intent Prediction in Gaze-based Interaction. In *Proceedings of the 2012 ACM Symposium on Eye Tracking Research & Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 83–90. https://doi.org/10.1145/2168556.2168569

Brendan David-John, Candace Peacock, Ting Zhang, T. Scott Murdison, Hrvoje Benko, and Tanya R. Jonker. 2021. Towards Gaze-Based Prediction of the Intent to Interact in Virtual Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual, Germany) *(ETRA '21 Short Papers)*. Association for Computing Machinery, New York, NY, USA, Article 2, 7 pages. https://doi.org/10.1145/3448018.3458008

Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-based Interaction Techniques. In *Proceedings of the 1990 CHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/97243.97246

Feiyu Lu, Shakiba Davari, and Doug Bowman. 2021. Exploration of Techniques for Rapid Activation of Glanceable Information in Head-Worn Augmented Reality. In *Symposium on Spatial User Interaction* (Virtual Event, USA) *(SUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 14, 11 pages. https://doi.org/10.1145/3485279.3485286

Feiyu Lu, Shakiba Davari, Lee Lisle, Yuan Li, and Doug A. Bowman. 2020. Glanceable AR: Evaluating Information Access Methods for Head-Worn Augmented Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (IEEEVR '20)*. Association for Computing Machinery, New York, NY, USA, 930–939. https://doi.org/10.1109/VR46266.2020.00113

Christof Lutteroth, Moiz Penkar, and Gerald Weber. 2015. Gaze vs. Mouse: A Fast and Accurate Gaze-Only Click Alternative. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 385–394. https://doi.org/10.1145/2807442.2807461

Ludwig Sidenmark and Hans Gellersen. 2019. 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* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1161–1174. https://doi.org/10.1145/3332165.3347921

Xinyong Zhang, Pianpian Xu, Qing Zhang, and Hongbin Zha. 2011. Speed-Accuracy Trade-off in Dwell-Based Eye Pointing Tasks at Different Cognitive Levels. In *Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-Based Interaction* (Beijing, China) *(PETMEI '11)*. Association for Computing Machinery, New York, NY, USA, 37–42. https://doi.org/10.1145/2029956.2029967