Gaze Gestures or Dwell-Based Interaction?

Aulikki Hyrskykari,* Howell Istance** and Stephen Vickers**
* University of Tampere, **De Montfort University
ah@uta.fi, hoi@dmu.ac.uk, svickers@dmu.ac.uk

Abstract

The two cardinal problems recognized with gaze-based interaction techniques are: how to avoid unintentional commands, and how to overcome the limited accuracy of eye tracking. Gaze gestures are a relatively new technique for giving commands, which has the potential to overcome these problems. We present a study that compares gaze gestures with dwell selection as an interaction technique. The study involved 12 participants and was performed in the context of using an actual application. The participants gave commands to a 3D immersive game using gaze gestures and dwell icons. We found that gaze gestures are not only a feasible means of issuing commands in the course of game play, but they also exhibited performance that was at least as good as or better than dwell selections. The gesture condition produced less than half of the errors when compared with the dwell condition. The study shows that gestures provide a robust alternative to dwell-based interaction with the reliance on positional accuracy being substantially reduced.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces – Evaluation/methodology; Input devices and strategies.

Keywords: eye tracking, gaze and gaming, gaze gestures, assistive input devices, physically disabled user groups.

1 Introduction

Gaze gestures as a means of issuing commands is a relatively recent technique for gaze-based interaction. Dwell-based interaction risks unintentional commands if the dwell interval is too short, and slow interaction if the interval is too long. Furthermore the size of the objects that can be selected by dwell is constrained by the accuracy of the eye tracker being used.

Gestures have the potential to avoid these two problems. The hazard of unintentional gesture commands is easier to avoid by designing gestures that are not likely to happen by mistake. Gestures are determined by saccades rather than by fixations. There is the potential for faster interaction as times to make saccades are shorter than dwell times. Importantly, gestures need not rely on users fixating on small on-screen targets. However, instead of selecting a command represented by a visible on-screen object, a gesture has to be mapped to a command. This requires an effective means of reminding the user of what this command is without disturbing the gesture, which is not a trivial design problem.

In this paper we report a direct comparison of performance and subjective appraisal between gestures and dwell-based interaction for tasks associated with game playing.

2 Related work

There have not been many studies of using gaze gestures to actively control the computer. However, Drewes and Schmidt (2007) built a general gesture recognizer and evaluated its performance. The gesture scheme they designed was inspired by FireGestures[1], a Firefox web browser plug-in, which recognizes mouse gestures that are composed of strokes into four directions left, up, right and down. The gestures in their system were composed from eight strokes consisting also the diagonal directions. Also Heikkilä and Räihä (2009) are studying gaze gestures in the context of developing a gaze-based drawing application and Mollenbach, Hansen, Lillholm and Gale (2009) discuss using simple single stroke gaze gestures combined with dwell buttons. In addition, stylus handwriting systems have inspired to implement eye operated text input systems, which can be seen as gaze gesture systems. Reviews of those can be found in Heikkilä's and Räihä's paper (2009) and also in [Skovsgaard, Räihä & Tall, 2011].

In our previous work (Istance et al. 2010) we studied the performance of gaze gestures when used to perform tasks in an MMORPG (Massive Multiuser Online Role Playing Games). In such games, the player's gaze has to be maintained in the centre of the screen, where the action takes place, for most of the time. Consequently, dwell clicks on command icons at the edges of the screen draw the user's attention away from gameplay. Gaze gestures allow for commands to be issued without having to look away from the centre of the screen. The gestures were recognized from a sequence of fixations in five gaze-aware regions on screen. A gesture always started from the round region in the center and visited one (2-legged) or two (3-legged) of the surrounding triangle regions (see Figure 1). All gestures were concluded by a fixation back in the centre zone. This allowed a total

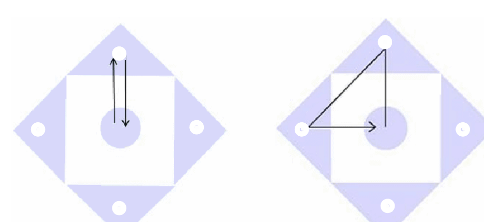

Figure 1 Examples of 2-legged and 3-legged gaze gestures experimented by Istance et al. (2010).

1 https://addons.mozilla.org/en-US/firefox/addon/6366

ETRA 2012, Santa Barbara, CA, March 28 – 30, 2012.
© 2012 ACM 978-1-4503-1225-7/12/0003 $10.00

of 4 x 2-legged and 8 x 3-legged different gestures to be made.

These experiments demonstrated that these type of gestures were feasible and effective. Users could learn how to make the gestures fairly reliably after only a short amount of practice.

3 Comparison of gaze gestures and dwell icons for discrete command selection

The advantage of maintaining visual attention in the centre of the screen while giving commands could also be obtained by arranging dwell objects representing commands around the players character in the centre of the screen. In this study, we wanted to compare whether commands could be executed more quickly and accurately using gaze gestures when compared with an equivalent number of dwell-activated buttons arranged in a circle around the central character.

We used an MMORPG (World of Warcraft) as the test application in our study. A total of 12 gestures were available, 4 x 2-legged gestures and 8 x 3-legged gestures (see Fig. 2a, described in more detail in Fig. 3a). This corresponds to 12 dwell buttons arranged around the central character (Fig. 2b and Fig. 3b). We expected the gesture condition to be more efficient in terms of speed and accuracy. This was because it would not be necessary to maintain the gaze point within the area bounded by an icon. There is clearly a trade-off between the size of the icon and the extent to which the radial arrangement of icons obscures the view of the game area. The icons were made semi-transparent.

3.1 Participants and apparatus

12 participants took part in the experiment. All were able bodied university students. We chose to recruit participants who were experienced gamers so as to reduce the amount of effort and learning required in playing the game. Eleven were male, there was one female, and the average age was 27.8 years. The eye tracker used was Tobii T60 integrated into a 17" TFT monitor for which the accuracy reported by the manufacturer is 0,5 degrees.

3.2 Experiment set-up

The duration of a gesture was taken to be the period of time from the last gaze point, of the last fixation in the centre zone - to the first gaze point, of the first fixation back in the centre zone. Using data from individual gaze points within fixations, rather than relying on operating system event time stamps to derive the timing of the duration of the gesture, removes the lag and variability due to other concurrent operating system tasks.

The player's character remained static throughout the tasks. The playing area in front of their character was populated with monsters of 'Level 1' or 'Level 2' ability. We used the level of the monsters as a means of instructing the player whether to use 2-legged or 3-legged gestures: 2-legged gestures were used for the attack, healing, and emote commands for Level 1 monsters, while 3-legged gestures were used for Level 2 monsters. Whether a monster was Level 1 or Level 2 could be determined by referring to the character description at the top left of the screen.

Two similar composite tasks were carried out. A Level 1 monster was selected with a target command, after which the player launched a specific attack spell, followed by a specific healing spell, followed by a command for specific emote. All of these actions required 2-legged gestures. If the target command resulted in a Level 2 monster being selected, the command was repeated until a Level 1 monster was found. This was repeated for 3 x Level 1 monsters. The second task was to repeat this task by attacking 3 x Level 2 monsters, using 3-legged gestures. The attack and healing spells, and emotes used were different in the second task.

Figure 3a shows the commands associated with gaze gestures in the experiment. The commands are grouped for ease of learning: attack spells at the bottom, healing spells on the right, emote commands at the top and the target command on the left. The same arrangement of commands was mapped onto the 12 dwell-activated icons, as shown in Figure 3b. The dwell icons subtended a visual angle of 1.3 – 1.4 degrees.

To provide base-line data, a mouse condition was included in addition to the gesture and dwell conditions. In the mouse condition, the participant used the same icons as in dwell condition, but selected with the mouse. The order of the mouse, gestures and dwell-button conditions was counterbalanced between participants.

Our earlier study [Istance et al., 2010], showed that the average time to complete 2-legged gestures was 493 ms, and 880ms for 3-legged gestures. The dwell period had to be set to some fixed

Figure 2a: *Giving 12 commands by four 2-legged and eight 3-legged gestures (one of the four regions highlighted just to make it visible in this figure). A gesture crib sheet in the lower right of the screen)*

Figure 2b: *Giving 12 commands by dwell buttons (three of the buttons highlighted just to make them visible in this figure)*

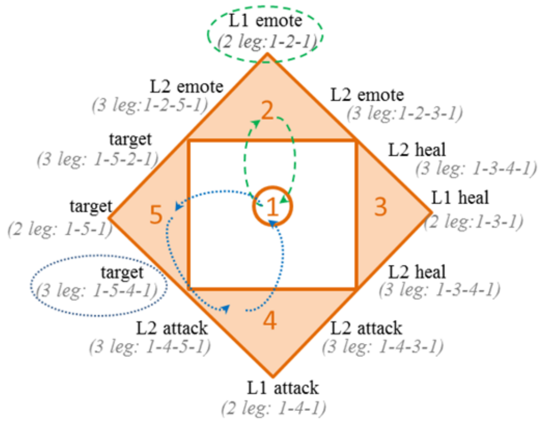

Figure 3a: *12 gaze gestures associated with commands. For example, the Level 1 emote command is performed by looking at the centre, glancing at the region 2 and back to the centre. (1-2-1). associated with 12 dwell buttons.*

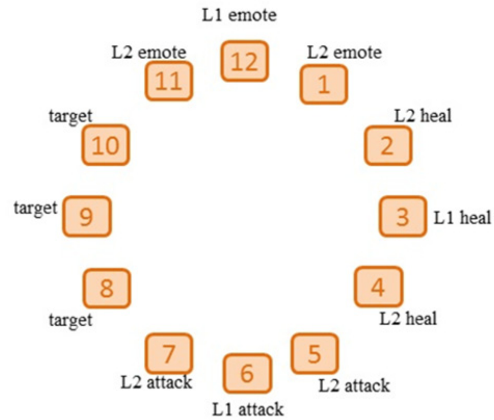

Figure 3b: *Target and emote commands, attack and heal spells associated with 12 dwell buttons.*

value. It was set to be the average of these (700ms). By equalizing the times required to issue a single command, either by gesture or dwell, the impact of accuracy on overall task duration could be compared in terms of time as well as error frequencies. We expected that the tasks using 3-legged gestures would take at least 4.6 seconds longer (880-493 x 4 commands x 3 repetitions) than tasks using 2-legged gestures. We expected no similar difference between targeting Level 1 and Level 2 monsters in the dwell selection or the mouse condition.

Results

The trials were analysed by marking up screen-capture videos with observed instances and durations of errors. The following categories of errors were defined:

- Selection error: any incorrect but valid gesture or incorrect icon selection after successfully targeting a monster at the required level.
- Missed (gesture or click): any attempt to make a gesture that is not recognized, or an icon selection where the click event misses then target.

The average times to complete the 2 tasks in each of the 3 conditions are shown in Table 1. The times shown include any time spent in error conditions. The gesture condition showed less than half of the errors and shorter observed overall task completion times compared with the dwell condition. As expected, task completion times using 2-legged gestures were observed to be shorter than those using 3-legged gestures.

Table 1 shows the total number of errors summed across all participants. The error counts between 2-legged and 3-legged gestures are similar except for the higher number of selection errors in the 2-legged gesture condition. This is largely due to occasions where one of the gesture zones was missed when trying to make a 3-legged gesture and it was recognised as a 2-legged gesture instead.

The difference between task completion times with 2-legged gestures and dwell button selection is not significant, but is close to being so (p=0.058). This is based on a 2-tailed pair-wise t-test between the times in both conditions. The difference in task completion times between 3-legged gestures and dwell buttons is not significant (p=0.56).

Task with Level 1 monsters (3 repeats)		**Gestures (2 legged)**	**Dwell Keys**	**Mouse**
time (s)	**mean**	**46.3**	**66.0**	**27.0**
	n	12	12	12
	sd	11.8	34.7	8.0
errors (nr)	total selection	3	7	-
	total missed	38	96	-

Task with Level 2 monsters (3 repeats)		**Gestures (3 legged)**	**Dwell Keys**	**Mouse**
time (s)	**mean**	**56.2**	**61.9**	**32.3**
	n	12	12	12
	sd	15.3	36.0	9.6
errors (nr)	**total selection**	16	3	-
	total missed	41	95	-

Table 1: The task completion times and errors from the three conditions, when attacking Level 1 and Level 2 monsters.

4 Subjective comparison of gestures and dwell selection

To compare the use of gestures with dwell selection in a free game-play situation, two conditions were devised where participants used a gesture-based interface and a dwell selection based interface respectively to move, search for and attack monsters.

The commands associated with the 2-legged gestures in Figure 3a were replaced with locomotion controls. The gestures to the left and right had the effect of turning the player's character by a fixed amount in each direction. The up gesture started movement in the forward direction and another gesture in the same direction stopped it. Similarly, a down gesture caused the character to move backwards. In the dwell selection condition, the radial arrangement of icons was retained but locomotion was controlled by looking into active zones away from the centre of screen (Istance et. al. 2009). In the mouse condition, the same icon arrangement was used but locomotion was controlled by mouse and keyboard.

Participants were asked to find and attack monsters in the game for 3 minutes in each condition. The participant's own level was high enough to ensure they were never killed. The same 12 participants took part and the order of conditions was counterbalanced between them. After the three conditions had been completed, participants completed a questionnaire. All questions used a 7-point fully labelled rating scale. A recording of the three conditions was played back and participants were encouraged to comment about the usability of each technique, in the style of 'retrospective think aloud' data collection.

All participants were able to successfully attack and kill more than one monster in each of the 3 conditions. The participants rated both gaze conditions as being poorer that the mouse condition (Fig. 4). The active region/dwell icon condition was rated slightly poorer than the gestures.

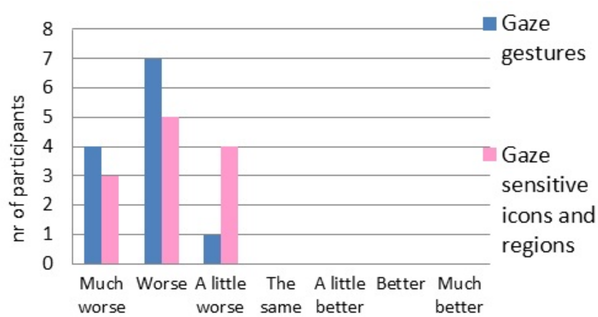

Figure 4: *Sense of control compared to using keyboard and mouse*

The data about the level of enjoyment is shown in Figure 5. There is no difference between gestures and dwell selection

Participants were asked which combination of gaze techniques would be preferable if they were to control the game by gaze only. 7 of the 12 participants said they would prefer to use active screen regions for locomotion and gestures for giving commands during fighting. 3 of the participants said they would prefer active screen regions for locomotion and dwell activated icons for giving all or most commands (2 did not give a clear opinion).

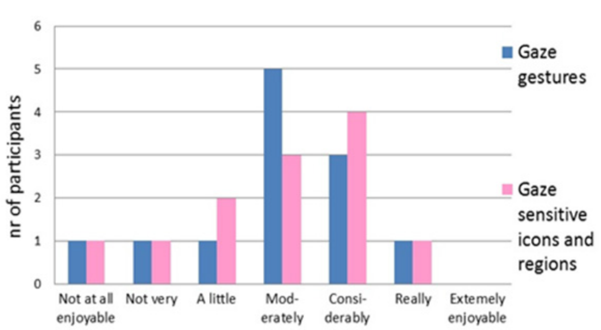

Figure 5: *How enjoyable it was to control the game using gaze.*

5 Discussion and Conclusions

To our knowledge the effectiveness of dwell selection and gaze gestures has not been compared before. The data presented in this paper applies only to able-bodied participants. If game players with physical disabilities are the target user group for gaze-based game playing, it cannot be assumed that the differences reported here apply. A separate study reported in this conference showed that two groups of participants with cerebral palsy and muscular dystrophy respectively had considerable difficulty in making gaze gestures.

The results from this study with able bodied gamers show that gaze gestures are an effective means of issuing commands during the course of game play when compared with dwell selection. We have been able to demonstrate the expected advantage of gaze gestures over dwell selection of command icons in terms of the numbers of errors made. Gaze gestures can be made during game play without having to move visual attention away from the centre of screen. The activation zones for the gestures do not obscure the view of the area around the character.

Gaze gestures represent a good potential interaction technique where access to a limited number of commands is required quickly and without reliance on a high level of accuracy. Gestures are better suited to giving discrete commands and less suited to continuous control adjustments, such as the direction of locomotion. Not surprisingly, neither of the gaze-only control interfaces, gestures or dwell selection, give able-bodied participants the same sense of control compared with a mouse and keyboard. The enjoyment ratings are neither negative, nor are they overwhelmingly positive. This is encouraging given that this was the first occasion that participants had used gaze control for integrated locomotion and command selection. Gaze interaction may well contribute to a more immersive and enjoyable gaming experience, when used as an additional modality as also observed by Smith and Graham [2006]. Further work is needed to investigate how far gaze gestures could be used in conjunction with mouse and keyboard as a means of increasing the input bandwidth of the game control interface.

6 Acknowledgements

We wish to express our thanks to Lauri Immonen and Santtu Mansikkamaa for running the experiments and helping to analyse the data.

References

DREWES, H. & SCHMIDT, A. (2007), Interacting with the Computer using Gaze Gestures. *Proc. of INTERACT 2007. Lecture Notes in Computer Science 4663,* Springer, 475–488.

HEIKKILÄ, H., AND RÄIHÄ, K.-J. (2009), Speed and Accuracy of Gaze Gestures. *Journal of Eye Movement Research 3*, 2:1, 1-14.

ISTANCE, H., HYRSKYKARI, A., IMMONEN, L., MANSIKKAMAA, S., AND VICKERS, S. (2010) Designing Gaze Gestures for Gaming: an Investigation of Performance. *Proc. of 2010 Symposium on Eye Tracking Research & Applications (ETRA '10).* ACM, 323-330.

MOLLENBACH, E., HANSEN, J. P., LILLHOLM, M., & GALE, A. G. (2009). Single stroke gaze gestures. *Proc. of the 27th international Conference Extended Abstracts on Human Factors in Computing Systems CHI'09.* ACM, New York, NY, pp. 4555–4560.

SKOVSGAARD, H., RÄIHÄ, K-J., AND TALL, M. (2011), Computer control by gaze. *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies* (Majaranta et al. eds.), IGI Global, Hershey, PA, USA, 78-102.

SMITH, D. AND GRAHAM, N. (2006) Use of eye movements for video game control. *Proc. of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (ACE '06). ACM, New York, NY, USA, Article 20.