Designing gaze gestures for gaming: An investigation of performance

Conference Paper · January 2010

DOI: 10.1145/1743666.1743740 · Source: DBLP

5 authors, including:

Howell O. Istance
De Montfort University

Aulikki Hyrskykari
University of Tampere

Lauri Immonen
University of Tampere

Stephen Vickers
De Montfort University

Designing Gaze Gestures for Gaming: an Investigation of Performance

Howell Istance
Department of Informatics
De Montfort University
Leicester, UK
hoi@dmu.ac.uk

Aulikki Hyrskykari
TAUCHI
Department of Computer Sciences
University of Tampere, Finland
ah@cs.uta.fi

Lauri Immonen
TAUCHI
Department of Computer Sciences
University of Tampere, Finland
lauri.immonen@cs.uta.fi

Santtu Mansikkamaa
TAUCHI
Department of Computer Sciences
University of Tampere, Finland
santtu.mansikkamaa@cs.uta.fi

Stephen Vickers
Department of Informatics
De Montfort University
Leicester, UK
svickers@dmu.ac.uk

Abstract

To enable people with motor impairments to use gaze control to play online games and take part in virtual communities, new interaction techniques are needed that overcome the limitations of dwell clicking on icons in the games interface. We have investigated gaze gestures as a means of achieving this. We report the results of an experiment with 24 participants that examined performance differences between different gestures. We were able to predict the effect on performance of the numbers of legs in the gesture and the primary direction of eye movement in a gesture. We also report the outcomes of user trials in which 12 experienced gamers used the gaze gesture interface to play World of Warcraft. All participants were able to move around and engage other characters in fighting episodes successfully. Gestures were good for issuing specific commands such as spell casting, and less good for continuous control of movement compared with other gaze interaction techniques we have developed.

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

Keywords: eye tracking, gaze gestures, gaze control, feedback, gaze and gaming

1 Introduction

The context for this work is designing interaction with Massively Multiplayer Online Games by eye gaze only. In particular, we are interested in role playing games and virtual communities. The target user group is people with motor impairments who wish to play games, such as World of Warcraft, or participate in virtual communities, such as Second Life. A high bandwidth input modality is needed for this, and simple mouse emulation by gaze is not sufficient to facilitate an adequate range or pace of interaction.

New gaze-based interaction techniques have to be found that fit (i) the particular game, (ii) the user in terms of their particular impairments and preferences, and (iii) the eye tracking equipment the user has in terms of its accuracy. User interfaces to MMORPGs (Massively Multiplayer Online Role Playing Games) enable a player to control his or her character's locomotion through a 3D graphical world, fight other characters, communicate with other players, and manipulate objects at the interface, such as an equipment pouch. In addition to the well-established issues of gaze-based interaction, this situation requires time-constrained, if not real-time, interaction, which is not the case with 2D desk top applications.

If dwell-click techniques are used for selection by gaze where icons are located at the edges of the screen, then a number of issues arise. Visual attention is diverted away from the centre area of the screen where most of the action takes place. The player has to look at a "cast spell" icon until it times out and the spell is launched, and then the player has to look at it again to cast another. Furthermore the size of the icon may be small leading to the familiar issues of inaccuracy when dwell clicking on this kind of icon. An interaction technique is needed that allows the player to look at the centre of the screen, is fast, and is not constrained by the need to maintain the gaze point within small targets. Previously, we have studied various ways to address these issues ([Istance, Bates, Hyrskykari & Vickers, 2008 and Istance, Hyrskykari, Vickers & Chaves, 2009]. In this paper we investigate the use of gaze gestures as a means of overcoming these problems.

We define a gaze gesture as..

> *"A definable pattern of eye movements performed within a limited time period, which may or may not be constrained to a particular range or area, which can be identified in real-time, and used to signify a particular command or intent."*

Most actions or commands in a game like World of Warcraft have user definable key bindings. We have built a layer of software as 'middleware' that is capable of recognizing patterns of eye movements in relation to areas displayed on a screen, and which generates keyboard events in response to these. Gaze gestures have so far been mainly used as a means of entering

ETRA 2010, Austin, TX, March 22 – 24, 2010.
© 2010 ACM 978-1-60558-994-7/10/0003 $10.00

text by eye. We have extended their use into interaction with 3D graphical environments and games.

In this paper we report an experiment with 24 participants to investigate factors in the design space of gaze gestures and the impact of these on performance and skill building.

Prior to presenting this experiment and its results in detail, we review the previous work with gaze gestures. We also wanted to study how well the designed gaze gestures work when actually used in games context, so we asked the experiences of 12 gamers who used the gaze gestures to play World of Warcraft. In section 5 we report their subjective opinions of their experience and also a study of how easily the designed gestures are made accidentally. We finish the paper by giving conclusions on what we learned of gaze gestures in the experiment and from the test play session.

2 Previous work

Gestures are a familiar concept in the context of other input devices; gestures made for example by stylus, mouse, hand or even body have been used in giving commands or feeding information to a computer. Even if the notion of 'gaze gestures' is relatively new, there has been studies that track gaze paths which can in a broad sense be considered as using gaze gestures.

2.1 Entering text using gaze gestures

There have been several different approaches to using gaze gestures for text entry. In the traditional dwell based eye writing system the dwell time sets a determinate limit for the typing speed. Thus, gesture based systems have appeared to be one possible solution to get rid of this constraint.

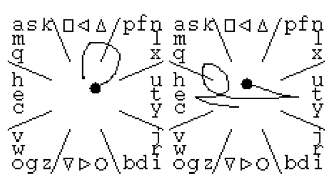

Figure 1. Quikwriting the letter 'f' and the word 'the' [Perlin, 1998].

Quikwriting [Perlin, 1998], a Graffiti-like stylus writing system, has been used as the basis for gaze sensitive writing systems. In Quikwriting the user enters characters with a stylus by drawing a continuous line on the surface. The characters are arranged around the starting position into 8 zones (Figure 1). A character is chosen by dragging the stylus from the centre to the zone where the needed character lies. If the character is the middle character in its zone, like 'n' in the top-right zone, the stylus is dragged back to the centre and 'n' is typed. To type other than the middle characters from a zone the stylus is homed via adjacent zones. Generating the letter 'f' is shown left in the Figure 1. To get 'p' the stylus should be 'home'd via the second adjacent zone, i.e. the top-left zone.

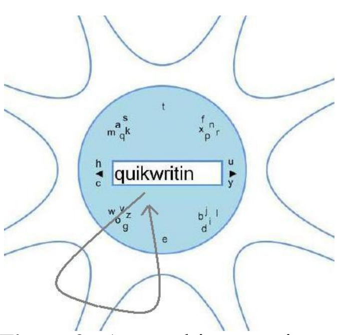

Figure 2. A gaze driven version of Quikwriting [Bee and Andrè, 2008]

In 2008 Bee and Andre built and tested a gaze driven version of Quikwriting (Figure 2). As gaze is used both as an input device and to view feedback, the hints for characters could not be displayed in the gaze sensitive zones, since the need of checking the hints would have disturbed making the gestures.

Another approach to using gestures for text entry is to make the shape of the gesture resemble the printed or handwritten shape of the character (Graffiti type of writing). This could make learning the gesture alphabets easier. Wobbrock (2008) built and evaluated such a system, EyeWrite (Figure 3). In their experiments with the system they found that it was somewhat slower than traditional on-screen keyboard dwell time typing, but it resulted in less mistakes. Thus, there seemed to be a speed-accuracy trade-off between these two approaches. However, the learning curve suggested that with practice the speed of using gestures approaches the speed of dwell time typing. In addition, EyeWrite was considered to be less fatiguing than on-screen keyboard typing.

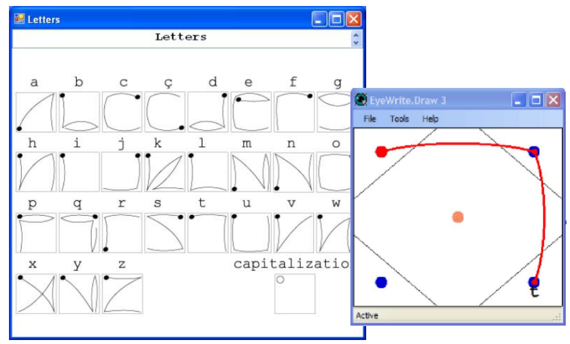

Figure 3. The gesture alphabet of the EyeWrite implementation of EdgeWrite and EyeWrite in action: writing a character 't '[Wobbrock et al., 2008].

Other text entry systems that share a gaze gesture approach include e.g. VisionKey [Kahn, Heynen, and Snuggs, 1999], pEYEwrite [Huckauf and Urbina, 2008]. For a review of these, see [Heikkilä and Räihä, 2009].

2.2 Gaze gestures in the interface

Beyond text entry systems, 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.

To find out how users are able to do these kind of gestures they made an experiment in which the users made a square clockwise and counter clockwise 4-legged gesture by looking at corners of a window (they could be interpreted e.g. as 'ok' and 'cancel'). Another set of gestures they experimented with, were the ones in Figure 4: two 6-legged gestures and one 4-legged gesture. Nine participants performed these gestures on different backgrounds, one having the square with helping lines giving support for the eye movements, another a spreadsheet document with a grid of lines and the third one was a blank (gray) background.

The participants were able to perform all the gestures so that the gesture recognizer registered them with the exception that four

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

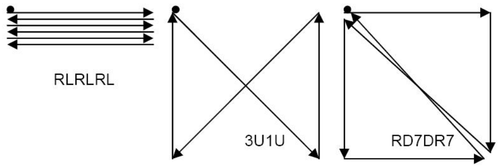

Figure 4. The three gestures chosen for an experiment [Drewes & Schmidt,2007]

of the nine participants failed to do the last, most complicated 6-legged gesture on blank background. It is well known that fixating "on nothing" is hard. The average time required to do one leg of a gesture was 557 ms.

An interesting feature in Drewes' and Schmidt's gesture recognizer is that the gestures are not location bound, which means that the algorithm is constantly "on watch" and a gesture can be launched from any position on the screen. Also, the algorithm is interpreting the direction between each recorded gaze position and if the direction is the same as the previous then the stroke is considered to be continuing. This makes the gesture sizes scalable and can be made in whatever size or aspect ratio.

In addition to the work above, there are at least two other studies on gaze gestures. Heikkilä and Räihä (2009) have been interested in using gaze gestures in the context of an eye driven drawing application. In their experiment the participants performed 2-legged, 3-legged and 4-legged gestures on both empty background and on a background with a visual guidance to do the gestures. The times per leg varied from 824 ms (in a 2-legged gesture forming a L shape on an empty background) to 1190 ms (in a 4-legged gesture with visual guidance). Mollenbach, Hansen, Lillholm and Gale (2009) discuss using simple single stroke gaze gestures combined with dwell buttons. In that context they studied single stroke gestures and the mean time they got for one leg gesture (a stroke from one side of a screen to the other side) was 334 ms.

3 Design of a gesture scheme

We used a scheme, which is a modified version of Perlins (1998) Quickwriting, and similar to that used by Bee and Andre (2008) in their work on gesture based text entry. Our version used a reduced number of regions or zones so that 12 different gestures could be recognized. We wanted the player to be able to make control gestures while looking generally in the central part of the screen. Thus, for us the use of active regions located in the centre of the screen is an acceptable restriction on where gestures would be recognized. The zones themselves were made semi-transparent so the player could see the avatar and surrounding part of the game world through the zones.

The gestures were made using 5 active zones. They were either 2-legged or 3-legged as shown in Figure 5 giving a total of 4 possible 2- legged and 8 possible 3 -legged gestures. The first of these target zones was called the major zone and the second, the minor zone. 2-legged gestures have major zones only. We wished to understand how different attributes of the design of the gesture scheme affected user performance, particularly in terms of the speed and reliability of performing the gestures.

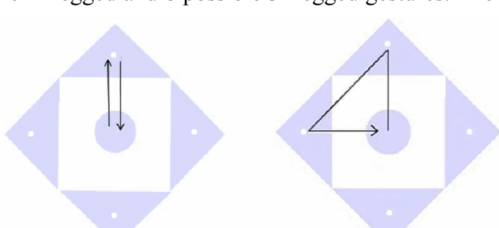

Figure 5 2-legged and 3-legged gestures, starting and ending in the centre.

3.1 Parameters investigated during pilot testing

a) Size of the gesture zones and distance from the centre to the inner edge of the active zone.

We opted not to test locating gesture zones at the edges of the window area as the amplitude of the gesture legs would be unnecessarily large. We did test the size of regions shown in Figure 6 against a set of regions which were 200 pixels greater than those shown. There was no significant difference in the pilot trial. We chose to continue with the smaller of the two sets.

b) Impact of adding fixation targets within the zones.

The initial trials showed that gaze points were clustered around the corners of the triangular zones in the absence of any other fixation lock. Adding small circular cutouts in the centre of the triangles had the effect of attracting gaze points to this feature.

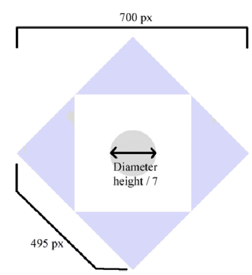

Figure 6 Size of the used gesture area.

c) Maximum allowable durations for gestures

The time for both 2 and 3-legged gestures was studied and the time-out period was set to 2 seconds for the main experiment. This was further revised after the experiment when gestures were incorporated into the gaze interface for the game.

d) Feedback

We investigated the impact of providing visual feedback by changing the colour of the zones, but we found this to be too distracting as pilot participants reported waiting for the feedback for each leg. For the experiment, a simple click sound was given as feedback that the complete gesture had been recognized within the timeout period. No feedback was given for an incorrect gesture.

e) Gesture timing

The initial implementation of the gesture recognizer was based on an "eye mouse", where the cursor was attached to the gaze point. We then used the operating system timestamp of mouse events as the cursor moved in and out of the gesture regions. However, when studied more closely this was found to be far too unreliable in view of the very short time durations of the gestures. There was a substantial lag after the point of gaze entered a region, before a mouse over event was generated. Thus, this approach was abandoned in favour of the one described in section 3.2.

3.2 Implementing the gesture recognizer

A valid *gesture* was accepted as a sequence of fixations, which begins with the last fixation in the centre zone before leaving it, one or more fixations in the major zone, followed by one or more fixations in the minor zone (for 3-legged gestures), and terminated with the first fixation back in the centre zone. Any no-zone fixations were allowed in the sequence. Let us use references T, B, L, R, C and N references for top, bottom, left, right, centre, and no-zones, respectively. Multiple sequential fixations in the same zone are replaced with one fixation and all "N" zones are removed from the sequence. For example, "C-T-C" and "C-T-L-C" are the valid 2 and 3-legged gestures as seen in Figure 5. An invalid gesture sequence is one that does not start and end with "C", exceeds the sequence timeout period, or does not produce any of the defined valid sequences.

The sequence had to occur within a 2-second time period. The time for the sequence began with the time of the last gaze point in the first fixation in the sequence, to the time of the first gaze point in the last fixation. Without this constraint, a variable amount of time could be spent looking in the centre zone at the beginning and/or the end of the sequence.

A *fixation* was defined as being 5 or more gaze points falling within a tolerance region centered around the average x value and the average y value of the previous gaze points in the sequence of gaze points. The tolerance region was defined to cover a visual angle of one degree. The location of the fixation was defined as the rolling average of the x and y coordinates of its component gaze points. The location was hit tested at the end of fixation to whether or not it fell inside a zone. The gaze points were delivered to the application every 15 or 16 ms by the eye tracker.

The implementation produced 3 logs, one of individual gaze points, one of fixations, and one of gesture sequences.

4 Experiment – performance in making gestures

Within the gesture scheme described in Section 3, we chose to investigate the effects on performance of 3 factors. By understanding how these factors impact performance, we hope to be able to devise a reliable and efficient gesture system by reducing the impact of negative factors. The factors were the following three.

a) The number of legs in a gesture: 2 or 3
As stated earlier, we count from the end of the starting fixation to the start of the terminating fixation. So a minimal 2-legged gesture would consist of saccade-fixation-saccade, and a minimal 3-legged gesture would consist of saccade-fixation-saccade-fixation-saccade (see Figure 5). If we simply assume that fixation durations are much longer than saccade durations then we would expect the durations for 3-legged gestures to be slightly less than 2 times the durations of the 2-legged gestures (2 fixations and 3 saccades versus 1 fixation and 2 saccades).

b) Principal direction of the eye movements in the gesture: vertical/horizontal or oblique
This is an important difference in the context of gaming. Vertical and horizontal gestures map well to directions of character movement, compared with diagonal (or oblique) movements. If the gestures are not used for movement then this natural mapping is less important (except perhaps in the case of camera control). However we would expect more accidental gestures where the principle components are vertical and horizontal compared with oblique eye movements. Here it is possible that a person's natural eye movements result in an unintentional gesture (a 'Midas' gesture). This factor can be manipulated by rotating the gesture detection regions by 45° resulting in a diamond and a square shape respectively (Figure 7).

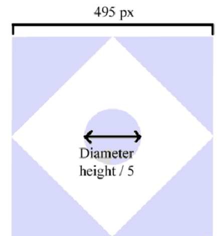

Figure 7 By setting the height and width of 495 x 495 px of the used square, the length of required saccades for gestures stays the same than in the 700 x 495 size diamond.

c) Direction of the first movement in the gesture: leftward or rightward
We suspected that there could be an effect on performance due to the direction of the first movement due to reading behaviour. It has been found that the perceptual span field is asymmetric [Rayner, 1995]. The span extends 14–15 character spaces to the right of fixation, but on the left only to the beginning of the fixated word, or 3–4 character spaces. This depends on cultural background, and we thought that since out participants are western readers, that might result the right first movements being faster than left first movements.

These three factors are represented in the eight gestures shown in Figure 8.

4.1 Participants

24 participants were recruited from staff and students attending a summer school in the university. There were 13 male and 11 females, with an average age of 38. No participant reported any ocular muscular defects that would have adversely affected their performance in the experiment. 12 participants had uncorrected vision, 11 had vision corrected with spectacles, and 1 had vision corrected with contact lenses.

4.2 Task and procedure

Participants were not required to learn the gesture sequences. Instead, they were presented with the 5 regions (top, right, bottom, left, centre) against a blank white background in the centre of screen. The 8 gestures were displayed as images at the edges of the screen, 4 on each side, and each were identified by a number. After an initial short training period, the participant was asked to make one of the 8 gestures 5 times. They were asked to make each gesture as quickly as possible, but they were told it was not necessary to make the set of 5 as quickly as possible.

A click sound was given as feedback after each one of the set was recognized, and a ping sound was given after the 5th successful gesture. After all 8 gestures had been repeated 5 times in this way, the complete set of eight gestures was repeated in the same order on 2 further occasions with a short pause for recovery in between. The next gesture that a participant was required to make, was announced by the trial leader verbally as a number, and participants were able to see the required pattern at the edge of the screen.

Participants were advised that they could withdraw at any time. There was no reward given for participation. The complete data collection period, including introduction calibration, training and 3 blocks of 8 gestures took between 20 - 30 minutes. The order in which the 8 gestures were presented to a participant was counterbalanced using a Latin Square.

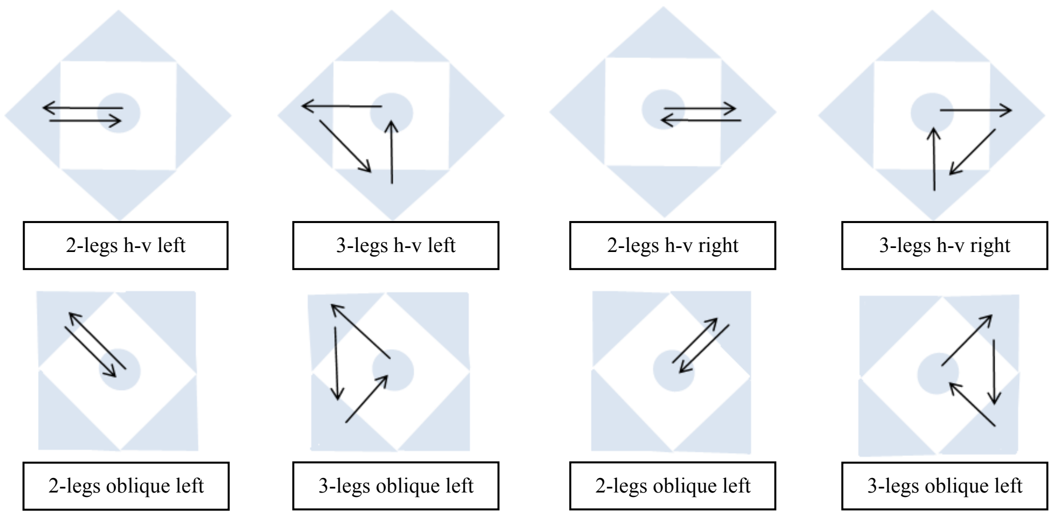

Figure 8 Four examples of horizontal-vertical (square layout) and four oblique (diamond layout) gestures.

4.3 Apparatus

All trials were conducted in a research office. A Tobii X120 eye tracker was mounted beneath a Samsung SyncMaster 959nf 19" CRT display. The participant was seated approximately 60 cms from the display. As described earlier, the centre of each zone was marked with a small circle as a fixation feature. The visual angle subtended between the marker in the centre zone and the markers in each of the triangular zones was about 7°.

4.4 Design

To summarise, the experiment was a 2 x 2 x 2 within-participants design, with 24 participants, 3 blocks of 8 conditions, 5 trials per block, giving a total of 120 trials per participant and 2,880 trials in total. The dependent variables were time to complete each valid gesture and errors. Errors were counted as either valid gestures that took longer than the 2 second time out, or gestures that were not the one that a participant was being asked to make at that time.

4.5 Results

We observed during the trials that frequently, when asked to make a new gesture, a participant would check the required pattern before and also during the first gesture. This led to timeout errors before the first gesture was successfully made and to very long times for the first gesture in the set of 5 compared with the other 4. We decided to remove the data from the first of the 5 gestures in all sets in all conditions. The average time to complete the remaining 4 gestures was used as the single score from a gesture in a given block. The data from each participant consisted of 24 scores, 1 for each of the 8 gestures in each of the 3 blocks.

4.5.1 Effect of practice

To gauge the learning effect, the times in the 3rd block were compared with those of the 1st block across all gestures and all participants. There were 8 data points in each block for each of the 24 participants giving 192 values for each block. Although the order in which gestures were completed was counterbalanced between participants, each participant performed the gestures in the same order. Therefore we can examine the differences between the blocks using a paired t-test.

There was a significant improvement in time to complete a gesture between block 1 and block 3 (Table 1). As a consequence, it was decided to discard the data from blocks 1 and 2 from subsequent analysis, and use the data from block 3.

Table 1 Time to complete a single gesture.

Time (ms) per Gesture	Block 1 (8 Gestures)	Block 3 (8 Gestures)
n	192	192
Mean	719	687
Stddev	332	290

Paired t-test, $p=0.04$

4.5.2 Main effect of number of legs in a gesture

As stated earlier, we would expect a difference in performance such that the time to complete a 3-legged gesture is slightly less than 2 times that required for a 2-legged gesture.

Table 2 Comparison of time to complete 2 and 3-legged gestures (block 3 only)

Time (ms) per Gesture	2 Legs (4 Gestures)	3 Legs (4 Gestures)
n	96	96
Mean	493	880
Stddev	332	290

Paired t-test p (1 tail) < 0.0001

The difference between the times for a 2-legged and a 3-legged gesture was highly significant, as expected (Table 2). The average time for a 2-legged gesture was about 0.5 second which compares favourably with dwell times commonly used in gaze communication for experienced users. The 3-legged gestures take longer and the ratio between them is 1.78 to 1, which was similar to what we expected on the basis of a simple comparison of the minimal number of saccades and fixations (less than 2 to 1).

4.5.3 Main and simple effects of the primary direction of eye movement in a gesture.

The main effect of the primary direction (direction of the first gesture to the major zone) was not significant (Table 3).

Table 3 Main effect of primary direction of movement (block 3 only)

Time (ms) per Gesture	Oblique (square) (4 Gestures)	H/V (diamond) (4 Gestures)
n	96	96
Mean	689	684
Stddev	282	298

Paired t-test p (1 tail) ≈ 0.69

However, the main effect includes both 2 and 3-legged gestures. The effect of the primary direction of movement is likely to be more pronounced for 2-legged gestures than for 3-legged. It is more interesting to look at the simple effect of direction of primary movement for 2-legged and 3-legged gestures separately.

Table 4a Simple effect of primary direction of movement for 2-legged gestures (block 3 only)

Time (ms) per gesture	Oblique: 2 legs (2 gestures)	H/V: 2 legs (2 gestures)
n	48	48
Mean	507	480
Stddev	125	150

Paired t-test p (1 tail) ≈ 0.05

4b 3-legged gestures (block 3 only)

Time (ms) per Gesture	Oblique: 3 legs (2 Gestures)	H/V: 3 legs (2 Gestures)
n	48	48
Mean	872	888
Stddev	278	269

Paired t-test p (1 tail) ≈ 0.25

There is small, but significant, difference between the times to complete 2-legged gestures where the primary direction of eye movement was horizontal/vertical compared with those with the primary direction being oblique. (Table 4a). There was no significant difference between the primary directions of movement for 3-legged gestures (Table 4b). These had either 2 oblique movements and 1 horizontal or vertical movement, or vice-versa. Adding a movement in the non-primary direction may have masked any small differences between the 2 primary directions.

Here there is a significant main effect, and gestures that begin with a leftward move are completed more quickly than those that begin with a rightward move (Table 5).

Table 5 Main effect of direction of first eye movement (block 3 only)

Time (ms) per Gesture	Left First (4 Gestures)	Right First (4 Gestures)
n	96	96
Mean	669	704
Stddev	272	307

Paired t-test p (2 tail) < 0.02

Rather surprisingly, the source of the effect lies within the 3-legged gesture, and there is no difference in gesture completion for 2-legged gestures between those that begin with a leftward movement compared with a rightward movement, as shown in Tables 6a and 6b.

Table 6a Simple effect of direction of first eye movement for 2-legged gestures (block 3 only)

Time (ms) per Gesture	Left First (4 Gestures)	Right First (4 Gestures)
n	48	48
Mean	493	494
Stddev	144	133

Paired t-test p (2 tail) ≈0.96

6b 3-legged gestures (block 3 only)

Time (ms) per gesture	Left First (4 Gestures)	Right First (4 Gestures)
n	48	48
Mean	846	914
Stddev	255	287

Paired t-test p (2 tail) < 0.01

4.6 Analysis of errors

Errors were categorized as either being a valid gesture in terms of an allowable sequence of zones, but taking longer than the 2 second timeout period, or being a recognizable gesture but not the one that the participant was being asked to make at the time. In most cases, the latter category applied to 3-legged gestures where one of the regions was missed, so that it was recognized as a 2-legged gesture instead. In the introduction to section 4, it was noted that participants frequently referred to the gesture diagram during the first gesture. Consequently, all errors that were made before the first of the 5 repeated gestures in each set were ignored.

The frequency of errors in each category for block 3 only are show in Table 7. The current error analysis does not adequately detect attempts to make 2 legged gestures where the major region was missed out. The error data is therefore more reliable for 3 legged gestures. The total number of errors for these was 55, summing across both categories. These arose from 480 correct gestures (24 participants x 4 3 legged gestures x 5 gestures in the block). This represents an total error rate of 11%.

The result in section 4.5.4 was that gestures that began with a rightward movement first were significantly slower than gestures that began with a leftward movement. One reason could be that the former were perceived as being more difficult to make, which could be reflected in a greater number of errors made in rightward first gestures. Table 8 shows the errors made in block 3 for left first and right first gestures respectively.

The probability of this occurring by chance is p = 0.62 (chi square = 0.23, 1 df) and thus we cannot conclude that the number of made errors explains why rightward first gestures are slower.

Table 7 Frequency of time out (A) and wrong gesture (B) errors for block 3 for each of the gesture parameter combinations

	Primary Oblique				Primary H-V				Total	
	First Left		First Right		First left		First Right			
Error Type	A	B	A	B	A	B	A	B	A	B
2 Legs	1	1	3	0	1	0	0	0	5	1
3 Legs	1	6	4	11	4	10	6	13	15	40
Total	2	7	7	11	5	10	6	13	20	41

Table 8 Frequency of time out (A) and wrong gesture (B) errors in block 3 separated by direction of the first eye movement in the gesture.

Error Frequency		Left first	Right First	Total
Error Type	A	7	13	20
	B	17	24	41
Total		24	37	61

5 Evaluating gestures during free game play

5.1 User experiences

We built a gaze gesture driven interface to support locomotion and fighting, and tested this in World of Warcraft with 12 able-bodied participants, all of whom were experienced gamers. We used the diamond shaped gesture regions and mapped the locomotion controls to the 4 2-legged gestures. Just like in most 3-D environments, the 'W','A', 'S' and 'D' keys can be used to control the movements of the players avatar in the game. A top region gesture switched a stream of 'W' key events on, and another top gesture switched the stream off. This caused the character to move forward. A bottom region gesture did the same for 'S' key events, causing backward movement. A left region gesture sent one 'A' event causing a turn to the the left and a right region gesture sent one 'D' key event causing a turn to the right.

The eight 3-legged gestures were assigned to commands for targeting other characters (in order to attack them), for launching attack spells and for launching healing spells. Similar commands were grouped into the same gesture region for ease of learning. The configuration of the gesture interface and the circular icons interface is shown in Figure 9.

The players were asked to freely locate and attack monster characters for about 5 minutes. This came at the end of an experiment where participants used gestures and other interaction techniques for locomotion and spell casting in a series of structured tasks. These took in total about 20 minutes to complete. In this study we actually compared different interfaces in real playing situations, but due to lack of space the study will be reported in detail elsewhere.

The outcome was very positive. All players were able to use the gesture driven interface to successfully move around in the game and to target and attack monster characters, after very little practice. Control over locomotion using gestures was experienced to be difficult, particularly during fighting. Turning was achieved by nudging their own character around in a series of discrete steps and this was effortful and time consuming. Although this was not a problem during locomotion over long distances that

Figure 9 Using Gaze Gestures to control the game. The four triangle areas (one highlighted just to make it visible in this figure) are displayed as transparent layers, each one having the small round "hole" in it to help the player's focus in the triangle.

require occasional changes in direction. Gestures were considered to be very effective however for issuing discrete commands such as spell casting.

5.2 Frequency of accidental gestures during game play

There is a danger of course that unintentional gestures will result from normal eye movements. For the game playing described in section 5.1, the maximum time for the gestures was reduced from 2 seconds to 800 ms for 2-legged gestures, and 1600 ms for 3-legged gestures. This was done to try to reduce the likelihood of unintentional gestures. In a separate small study we asked 2 of the 12 gamers to play World of Warcraft for 15 minutes using keyboard and mouse while their eye movements were recorded.

We examined the frequency of gestures detected with both the diamond and the square shaped gesture zone configurations. We expected that the diamond configuration of regions would lead to a higher frequency of unintended gestures as we expected a greater frequency of vertical and horizontal eye movements, then oblique movements. The results are shown in Table 9.

Table 9 Accidental gestures during 18 minutes of free play.

	Gesture	player 1	player 2
Diamond	2 leg: up	3	12
	2 leg: right	5	2
	2 leg: bottom		2
	2 leg: left	1	
	Total	9	16
Square	2 leg: upper left	1	
	2 leg: upper right	1	2
	2 leg: lower left	1	1
	Total	3	3

The observed data matched the expected data with far fewer unintended gestures where the primary direction was oblique compared with being horizontal and vertical (an average of 3 in the former case compared with 12.5 in the latter in 18 minutes of continuous play). It is noteworthy, but not surprising that no unintended 3-legged gestures were detected. Reliability of the chosen gesture scheme evidenced by few unintentional gestures is an important factor in the design of a gaze gesture based interface.

6 Discussion and Conclusions

We have investigated some of the design space of gaze gesture schemes intended for interacting with Massively Multiplayer Online Role Playing Games (MMORPGs). People can learn to make 2 and 3-legged gaze gestures fairly reliably after only a short amount of practice. The average times for completion of 2-legged and 3-legged gestures are 490 and 880 ms respectively, which compares favourably with dwell periods that are used in desktop applications for command selection by gaze. This means of interaction comes with the advantage of not having to fixate accurately on small targets in order to select them.

We were able to predict fairly well the ratio of completion times between 2 and 3-legged gestures on the basis of a simple comparison of the number of fixations and saccades in each. We were also able to predict the difference between 2-legged gestures where the primary direction of eye movement was, in one case, horizontal and vertical, and in the other, oblique. We did find a difference between gestures where the first movement was leftward, and where it was rightward, although this was confined to 3-legged gestures. We expected a possible effect in the other direction, so this was an unexpected effect. We were not able to find support for that in the literature. Becker (1991) states that there is tentative evidence that upward directed movements reach higher velocities than downward ones, but does not state anything about left and right bound movements. In fact, Abrams, Meyer and Kornblum (1989) found that there is no speed difference between left and right saccades. As we are not able to offer an explanation for this, we do not offer it as a significant finding. We are how-ever encouraged to look further into the area of modelling user performance when making different kinds of gaze gesture in order to be able to predict user performance with different schemes and variations on these.

In previous studies the time measured for a leg in a gaze gesture has varied a lot, e.g. Drewer and Schmidt: 557 ms, Heikkilä and Räihä 824-1190 ms and Mollenbach et al. 334 ms (see Section 2). Our figures vary from 247 ms (= 493/2) to 293 ms (= 880/3) for a 3 and 2-legged gestures. Why do the times vary so much and why are our times less than others? When we are comparing these small times and difference between them, we are working with single saccades and often single fixations and accuracy and consistency in time measurement is important. Using readymade fixation detecting algorithms provided by the manufacturer without reporting the exact parameters used for fixation detection is problematic. The 'eye-mouse' approach is also unreliable as the operating system needs to recognise gesture regions; update the cursor position; callback any mouse over events; and so on. How fast this sequence happens will partly depend on what other processes are doing at the same time on the same machine. We think that to report times for gestures reliably it is necessary to work at the sub-fixation level with times of gaze points that end and begin the starting and terminating fixations respectively.

We believe that gaze gestures are an effective means of interacting with MMORPGs, particularly for tasks which involve selection of commands, rather than continuous control of locomotion. 12 experienced gamers were able to use a gaze gesture only interface for free game play with World of Warcraft after very little training. We have shown that the rate of unintentional gestures during game play is much lower with the square configuration of regions (mostly oblique eye movements) compared with the diamond shaped configuration mostly (horizontal and vertical movements). If the directions of the gestures are not important (as they are in the case of locomotion tasks) then the square configuration is better to use in the games interface. It is likely however that the strength of gestures as an interaction technique for gaze–based gaming lies in its combination with other gaze-based interaction techniques, rather than trying to use it exclusively for all tasks..

References

ABRAMS, R. A., MEYER, D. E., AND KORNBLUM, S. 1989. Speed and accuracy of saccadic eye movements: Characteristics of impulse variability in the oculomotor system. *Journal of Experimental Psychology: Human Perception and Performance 15*, 3, 529-543.

BECKER. W. 1991. Saccades. In Carpenter, R. H. S. (ed) *Eye movements. Vision and visual dysfunction; vol.8,* Basingstoke: Macmillan, 95-133.

BEE, N., AND ANDRE, E. 2008. Writing with Your Eye: A Dwell Time Free Writing System Adapted to the Nature of Human Eye Gaze, *Lecture Notes in Computer Science 5078*, Springer Berlin / Heidelberg, 111-122

DREWES, H. & SCHMIDT, A. 2007. Interacting with the Computer using Gaze Gestures. In *Proceedings 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. In *Journal of Eye Movement Research 3*, 2:1, 1-14.

HUCKAUF, A. & URBINA, M. H. 2008. Gazing with pEYEs: towards a universal input for various applications. In Proc. of the 2008 Symposium on Eye Tracking Research & Applications-ETRA '08. ACM Press, New York, NY, pp. 51–54.

ISTANCE, H., BATES, R., HYRSKYKARI, A., AND VICKERS, S. 2008. Snap clutch, a moded approach to solving the Midas touch problem. *Proc. of 2008 Symposium on Eye Tracking Research & Applications (ETRA '08).* ACM, 221-228.

ISTANCE, H., VICKERS, S., & HYRSKYKARI, A. 2009. Gaze-Based Interaction with Massively Multiplayer On-line Games. *In Proceedings of the 27th international Conference Extended Abstracts on Human Factors in Computing Systems,CHI '09. Conference.* 4381-4386

KAHN, D. A., HEYNEN, J., & SNUGGS, G. L. 1999. Eyecontrolled computing: The VisionKey experience. In *Proceedings of the Fourteenth International Conference on Technology and Persons with Disabilities. CSUN'99*, Los Angeles, CA.

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

PERLIN, K. 1998. Quikwriting: continuous stylus-based text entry. *Proc. of 11th Annual ACM Symposium on User interface Software and Technology (UIST '98),* 215-216, a demo available at http://mrl.nyu.edu/~perlin/demos/Quikwrite2.html

RAYNER, K. 1995. Eye movements and cognitive processes in reading, visual search, and scene perception. In Findlay, J. M., Walker, R., & Kentridge, R. W. (Eds.) *Eye Movement Research: Mechanisms, Processes and Application.* New York: Elsevier, 3-21.

WOBBROCK, J.O., RUBINSTEIN, J., SAWYER, M., AND DUCHOWSKI, A.T. 2008. Longitudinal evaluation of discrete consecutive gaze gestures for text entry. In *Proc. 2008 Symposium on Eye Tracking Research & Application.ETRA '08.* ACM, 11-18.