Supporting Making Fixations and the Effect on Gaze Gesture Performance

Howell Istance
University of Tampere
Tampere, Finland
howell.istance@staff.uta.fi

Aulikki Hyrskykari
University of Tampere
Tampere, Finland
ah@staff.uta.fi

ABSTRACT

Gaze gestures are deliberate patterns of eye movements that can be used to invoke commands. These are less reliant on accurate measurement and calibration than other gaze-based interaction techniques. These may be used with wearable displays fitted with eye tracking capability, or as part of an assistive technology. The visual stimuli in the information on the display that can act as fixation targets may or may not be sparse and will vary over time. The paper describes an experiment to investigate how the amount of information provided on a display to assist making fixations affects gaze gesture performance. The impact of providing visualization guides and small fixation targets on the time to complete gestures and error rates is presented. The number and durations of fixations made during gesture completion is used to explain differences in performance as a result of practice and direction of eye movement.

Author Keywords

Gaze gestures; gaze gesture performance; fixation targets; fixation duration.

ACM Classification Keywords

H.5.2 User interfaces: Input devices and strategies.

INTRODUCTION

The use of eye tracking technology is becoming more widespread and can be integrated with mobile devices, public displays and wearable displays, as well as it's more established use with desktop computers. This provides both information about where a person is looking on a display or in the world, and opportunities to use gaze-based interaction techniques. These techniques divide into three groups: those based on extended fixations (or dwells); those based on patterns of saccadic eye movements, or gaze gestures; and those based on recognizing smooth pursuit patterns of motion. [18].

Wearable near eye displays are becoming increasingly popular as consumer electronic devices. The VR headset (of type HTC Vive) can be fitted with relatively high quality eye tracking capability [27]. It will also be possible to fit gaze tracking capability to see-through 'smartglasses' (of type Epson Moverio), which can project information onto the user's view of the world. It is likely that the position of these glasses with the respect to the head will be less stable than the VR headset. It is also likely that less processing power will be available to smartglasses leading to lower sample rates of gaze position.

Saccadic gaze gestures are deliberate patterns of eye movements that can be identified and used to give a particular command [15], for example, to zoom into the display, to show the next email, or to switch off the augmented information. They offer a hands-free interaction technique that is less reliant on accurate measurement of gaze position, and less reliant on maintaining good calibration. Dwell-based interaction requires both of these features.

There are several use cases for saccadic gaze gestures that suggest themselves. One is using smartglasses, say at a workplace such as engineering maintenance, without voice commands and the use of the hands [30]. Another is hands free interaction with mobile displays [2, 8]. A significant use case is as an assistive technique for motor impaired users. This could be either with mobile devices that have limited eye tracking capability, or in gaze control of applications where dwell is problematic, e.g. games [15]. Gestures may be less suitable however when interacting with, or being seen by, other people if the reasons for the eye movements are misinterpreted [1].

Making gaze gestures is quite different from other visual tasks, such as reading or visual search. Gestures involve making deliberate eye movements, and generally extracting little or no information from the component fixations. By contrast in reading, eye movements are directed by characteristics of the task (top-down processes) and of the visual stimulus (bottom-up processes) [24].

Normally the visual system requires a target for the eyes to fixate upon. There may be suitable fixation targets within the displayed data or there may not be. The question then arises of whether specific fixation targets should be provided in order for someone to be able to make these patterns of eye movements efficiently and reliably. If fixation targets are to

CHI 2017, May 06-11, 2017, Denver, CO, USA
© 2017 ACM. ISBN 978-1-4503-4655-9/17/05…$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025920

be superimposed on the display surface, then they should have minimal impact on the information displayed upon it.

The research question addressed in this paper is what the effect on gaze gesture performance will be of different amounts of visual information specifically aimed at supporting making fixations. We aim to help interaction designers to understand how to support making gaze gestures while viewing an information display.

We present a detailed study of how specific support in making fixations impacts gesture speed and error rates. The work is novel in that it investigates gaze gesture performance using a high sample rate tracker. Facilitated by the sample rate, this is the first study to offer explanations of differences in gesture performance in terms of number and durations of fixations.

Background

During normal visual work, the eyes are constantly moving. Within this stream of normal eye movements, a pattern of deliberate eye movements that constitutes the gaze gesture has to be detected. Different types of gestures can be categorized then according to (1) how the start and end of the pattern is recognized (gesture segmentation), (2) how the pattern of eye movements is detected, and (3) how complex the pattern is.

(1) Recognizing the start and end of the pattern

Gesture segmentation refers to recognizing the start and end of a gesture pattern. This may use a starting fixation in a particular location and possibly a terminating fixation in the same location or zone. Here all gestures start and end in the same location, and the gesture consists of the sequence of locations visited in between. Another approach to segmentation is to use a dwell in a particular location to signal the start of a gesture. Alternatively, there is no explicit segmentation and the stream of zones or directions arising from normal eye movements is continuously parsed. A gesture is identified as soon as a valid sequence is detected.

(2) Gesture detection schemes

The pattern of eye movements that make up the gesture can be detected in several ways. These are boundary crossing, active zones, changes in saccade direction, shape-tracing, eye-based head gestures and smooth-pursuit. Each scheme has different properties that may be appropriate for different use scenarios and applications.

Boundary Crossing [9, 12, 14]. A gesture is registered by a fixation on one side of a boundary followed by another fixation on the other side of it. In order to prevent unintentional registrations of gestures, a gesture may require multiple crossings (over and back) to be completed in sequence. The boundary may be soft, i.e. that of an on-screen object, or hard such as the edge of the screen.

Active Zones [4, 15, 20, 26, 29]. Instead of simply crossing a boundary, a gesture is registered by hit testing whether fixations occur in specific designated areas of the screen or zones and in a particular sequence. Changing the number of zones, or the sequence in which zones are visited, produces a different gesture. An example of such a scheme in context of gaze control of games is shown in Figure 1.

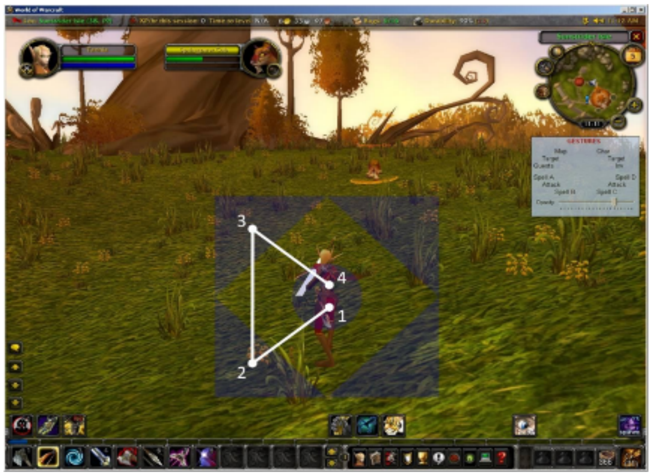

Figure 1: Active zones used for detecting gaze gestures while playing World of Warcraft located around the player character. The white lines indicate a 3-stroke gesture and not were visible during use [15].

Changes in Saccade Direction [7]. A gesture is a change in the direction of the gaze path without reference to any external features (such as a boundaries or zones). The gesture can be made anywhere and the pattern consists of a sequence of directions. A scheme may define a number of patterns, and the normal gaze path is monitored by a classifier to determine if any of the patterns have been generated. The patterns may be of varying levels of complexity, although the simpler the pattern the greater the risk of the classifier producing false positives.

Shape tracing [9, 26]. A gesture results from an attempt to trace the outline of a shape with the gaze path, for example two lines at right angles, a square or a circle. This may use the edges of an on-screen object as a visual guide, or not.

Eye-based head gestures [19, 22]. A head mounted eye tracker will detect a characteristic pattern of eye movements if the head is moved, with say a nod, while the viewer continues to look at an object. The rationale is that controlled, deliberate head movements are easier to make than deliberate eye movements.

Smooth pursuit gestures [17, 6]. Smooth pursuit gestures result from tracking the eyes while a person looks at a moving object in the world or on a display surface. Normal saccadic eye movement is suspended during smooth pursuits. The gestures would be characterized by the direction, velocity and possibly shape of the motion path.

(3) Pattern Complexity

The complexity of the pattern of movements is described by the number of deliberate eye movements or 'strokes'. In the literature, this complexity ranges from single stroke gestures [21] up to 6 stroke gestures [8]. High numbers of strokes are typically used to reduce the risk of false positives where no explicit segmentation event is used. However, higher numbers of strokes carry an overhead of both time to make the gesture, and effort. In a study [15] of eye movements made during game playing, there were between 3 to 5 times as many normal eye movements that would have been interpreted as 2-stroke horizontal and vertical gestures compared with 2-stroke diagonal gestures. Significantly there were no patterns of eye movements that would have been interpreted as 3-stroke gestures.

Gestures investigated in the current study

We chose to investigate an active zones scheme with explicit segmentation. The sequence of zones began with a fixation in the center zone followed by one or more fixations outside the center zone and continued until the next fixation back in the center. This was motivated by the prevalence of this approach in the literature and alignment with the previous investigation of gestures in games. As no naturally occurring patterns of eye movements that could be interpreted as 3-stroke gestures had previously been found, we considered it unnecessary to study any greater complexity. Thus, the current study considered only 2-stroke and 3-stroke gestures.

EXPERIMENT

The effect on gaze gesture performance of providing different amounts of visual information to aid making fixations was investigated in an experiment. Four zones were arranged around the center of the display. The zones extended to the edges of the display in each direction. The distance from the edge of the zone to the center of the display was chosen as a compromise between reducing the risk of unintended gestures and the effort required to make deliberate eye movements.

The information added to the display to enable the viewer to fixate within the zone was split into two types. One was the visualization of the size and location of the zone, and the other was the provision of small fixation targets within the zones.

In all conditions, the center zone was visible as a circular blue-grey semi-transparent overlay, shown in Figure 2a.

The visualization of the outer zones was one variable and was represented by 3 levels:

- no visualization of the outer zone (Figure 2a)
- showing only the *leading border* of the zone closest to the center zone (Figure 2b), and
- showing the size and location of the *zone area* by means of a blue semi-transparent overlay and a black border (Figure 2c).

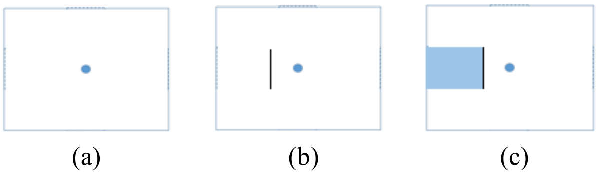

Figure 2: Three levels of visualizing an outer zone: (a) no visualization, center zone only is visible (b) leading border (c) zone area

The provision of fixation targets was a second variable and was represented by 2 levels:

- no dot, and
- small circular black dot in the center of zone.

These conditions are summarized in Figure 3.

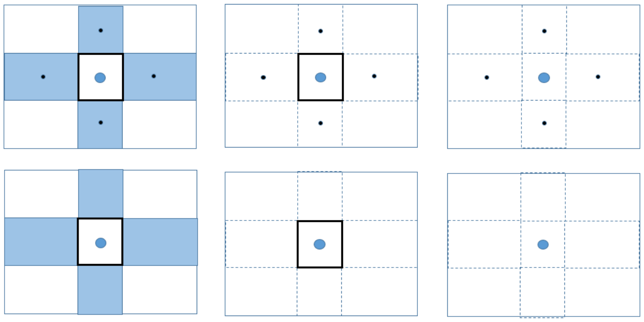

Figure 3. Six conditions arising from 3 levels of the visualization variable and 2 levels of the fixation target variable. The dotted lines show the extents of the active zones and were not visible on the display.

Previous work [15] had shown that 2-stroke gestures consisting of horizontal and vertical eye movements were completed more quickly than ones with diagonal movements, although the difference was small. In the case of 3-stroke gestures the difference was not significant. Other studies of gaze gesture performance have also reported different speeds of gesture completion depending on whether the gesture consisted of horizontal/vertical eye movements or diagonal movements [10]. The peak velocities for diagonal movements is less than those on the horizontal/vertical meridians [3], which may indicate that diagonal deliberate eye movements are more difficult to control.

We expected that the impact of adding information to the display to aid making fixations would be influenced by the difficulty of making the gesture. We considered the difficulty of making the gesture to be determined by the predominant direction of eye movement and by gesture complexity. We compared gestures consisting primarily of horizontal/vertical eye movements with those consisting primarily of diagonal eye movements. Gesture complexity was represented by 2 levels, 2-stroke and 3-stroke gestures. These variables are illustrated in Figure 4 below.

In summary, there are four independent variables in the study

- Zone visualization: 3 levels (no visualization, leading border only visible, zone area visible)
- Fixation targets: 2 levels (absent, present)
- Predominant direction of eye movement during gesture: 2 levels (horizontal/vertical, diagonal)
- Gesture complexity: 2 levels (2 strokes, 3 strokes)

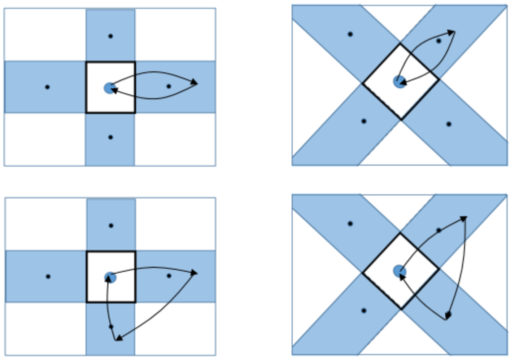

Figure 4: Two stroke gestures shown above and three stroke gestures shown below. Gestures consisting predominantly of horizontal/vertical eye movements are shown on the left and of diagonal eye movements are shown on the right.

Dependent Variables

The indicators of performance that were studied were speed and accuracy.

Speed is the time to complete an error-free gesture. The duration of the gesture was defined as being from the onset of the saccade that left the center zone, to include fixations and intermediate saccades in the zones that made up the gesture, and terminating with the end of the first saccade to land back in the center zone. The fixation prior to the gesture commencing and the first fixation on terminating are each specifically excluded from the duration of the gesture. This is important as the duration of either (or both) of these is large in comparison with the total duration of the gesture, particularly in the case of the 2-stroke gesture.

An *error* is defined as one of the following cases

- a pattern of eye movements where a fixation in the zone or zones that comprise the gesture is omitted,
- a pattern which includes fixations in zones that are not part of the gesture,
- a pattern in the case of 3 stroke gesture where the zones are visited in the wrong order, and
- trials where the gesture has not been completed within 3 seconds.

Fixations made outside any of the zones did not constitute an error, as long as fixations were made in the zones in the order prescribed by a particular pattern.

Experimental Design and Procedure

Three-stroke gestures clearly take longer to complete than 2-stroke gestures as they entail at least one extra saccade and one extra fixation. Consequently, we analyzed the data for each separately to reduce the complexity of the data model in the analysis of variance. The data for both 2- and 3-stroke gestures were collected in the same sessions.

A 3x2x2 repeated measures factorial design was used for each number of strokes. With 3 levels of visualization (none, leading borders, zone area) * 2 levels of targets (no targets, targets) * 2 levels of direction (horizontal/vertical, diagonal) this gives 12 combinations. For each combination, there were 4 initial directions of eye movement (either up, right, down, left, or upper left, upper right, lower right, lower left). Thus 48 trials are needed to provide all possible combinations. The initial direction was not considered as an independent variable but the trials were balanced with respect to this. The 48 trials were randomized into 2 sets of 24 trials for each participant. This gave a total of 4 sets of 24 trials, 2 for 2 stroke gestures and 2 for 3 stroke gestures. These 96 trials constituted one block. For each participant, this block was repeated in the same order once following a rest break. This was to give the opportunity of comparing individual gesture completion times between blocks knowing that each repetition would be separated by the same number of trials.

Participants

Twenty-four participants were recruited from staff and students at the University of Tampere. However, data for one participant was lost, leaving 23 complete data sets.

Fifteen of the participants were male, nine were female. Their ages varied from 19 to 57, with an average of 29. Fifteen had uncorrected vision.

In accordance with university ethics procedures, participants were each given and signed a written description of the purpose of the experiment. They were told that they could withdraw at any time without any negative consequence. The written description stated that all data collected during their trial would be scored securely and used anonymously.

Data collection

All data for each participant was collected in a single session. Each session lasted approximately 1 hour and 10 minutes. The session had 3 parts, practice, first block of trials, and second block of trials. The eye tracker was re-calibrated at the beginning of the first and second block with a procedure involving looking at a matrix of 9 dots on the screen. There was no re-calibration within blocks, but a drift check was made at the beginning of each trial.

The first part lasted 20 minutes and consisted of a spoken introduction to the experiment and to the eye tracking equipment, and the administration of the written consent. It included a demonstration of the different gestures and the different amounts of visual information provided. It included

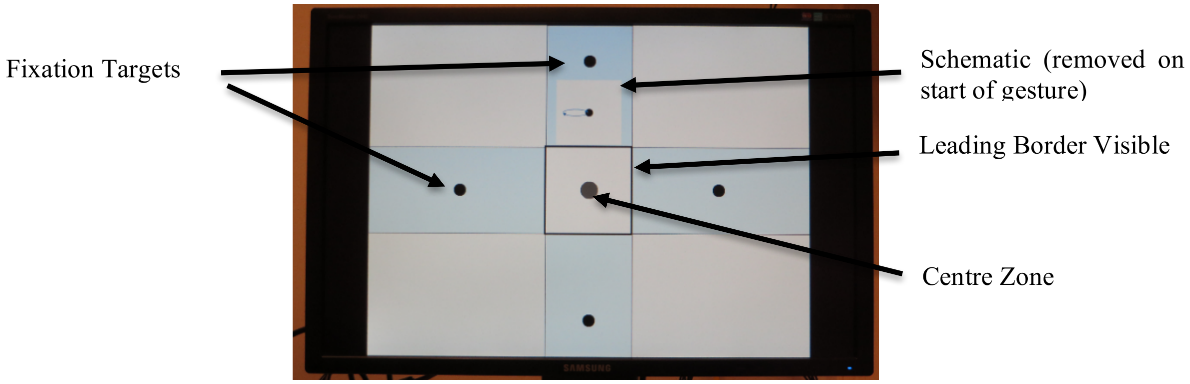

Figure 5: Appearance of screen at the start of a trial where both zone area and fixation targets are visible. A schematic of the gesture to be performed, here a 2-stroke horizontal gesture to the left, is overlaid on the upper part of the screen

a training session where the participant could practice making the gestures.

Following a break of a few minutes, the second part commenced. The first block of 96 trials was conducted as 4 sets of 24 trials. After each set the participant could rest and move. Each trial commenced with a mouse click under participant control.

At the start of a trial, the screen appeared with the particular combination of zone visualization, fixation targets and orientation (see Figure 5). The actual gesture to be performed was shown in a small schematic in the upper central part of the screen. All trials in a set were either 2 stroke or 3 stroke gestures. The participant memorized the action required and then looked at the center zone, which caused the schematic to be removed. After a 2 second delay, the color of the circular center zone changed to green, which together with a simultaneous beep signal informed the participant that they could make the gesture when they were ready. On completing the gesture, the center zone returned to the default blue-grey color. If an error occurred, feedback was given by an audio signal and the center zone turning red briefly after the terminating fixation before reverting to the default color. The next trial commenced when the participant clicked the mouse button. After the 96 trials were completed, the participant was asked to leave the lab and take a 20-minute break.

On their return, the third part commenced. The 96 trials were repeated in the same order as in the second session.

Equipment Used

An EyeLink 1000 eye tracker from SR Research was used together with a headrest, which provided a sample rate of 1000Hz. The tracker reports saccades made by the eyes. A fixation then is defined as any gap between the end of one saccade and the start of the next. Low sample rate trackers, of say 30Hz, are not able to do this accurately. A full discussion of fixation definition is given in Holmqvist [11]. The eye tracker was located on a desk and participants viewed the display seated at a distance of 985 mm.

The eye tracking systems that will be used for actual gesture detection in the situations described in the introduction will be far less accurate and have a far lower sample rate than the tracker we have used. The use of an accurate high sample rate tracker provided detailed information about eye movements and fixations made within a gesture.

The size of the components and the distances between these from Figure 5 are shown in Table 1.

Distance	Visual Angle
Center Zone diameter	1.0
Fixation target diameter	0.6
Center zone edge to leading border	1.9
Center zone edge to inside edge of nearest fixation target	6.5

Table 1: distances between on-screen objects in degrees of visual angle, viewing distance 985 mm

Hypotheses and expected results

We hypothesized that the more information provided, the better the gesture performance would be. We also hypothesized that the impact of providing visual support would have a more positive effect on performance as gestures became more difficult to make. In accordance with our previous work [15], we expected a learning effect and that the times and errors made during the second block of trials would be less than the first block of trials. We expected that 2-stroke horizontal/vertical gestures would be made more quickly than 2-stroke diagonal ones.

RESULTS

Presence of a learning effect between blocks of trials

There were 4 repeated trials within each combination of experimental conditions, balanced across the initial direction of eye movement. The average of the error-free gesture durations of these 4 trials was taken as the score for each participant. Error trials were excluded from the average. There were no combinations of conditions for any participant where all 4 trials were in error and using the average duration avoids the problem of missing data. The means and standard deviations of all participant scores is shown in Table 2.

		Block 1	Block 2
2 stroke	mean (ms)	704.8	617.4
	(std dev)	(174.6)	(176.7)
3 stroke	mean (ms)	1163.0	1058.4
	(std dev)	(293.8)	(249.0)

Table 2: Means and standard deviations of average gesture durations for error free trials for each participant, n =23

A 2x2 factorial analysis of variance (blocks and strokes) was conducted on the data summarized in Table 2. This showed that there was a significant main effect for blocks (F(1,22) = 10.4, p = 0.004) and a significant main effect for strokes (F(1,22) = 205.0, p < 0.001), but no significant interaction blocks x strokes interaction effect (F(1,22) = 0.308, p = 0.58). On the basis of this, we are justified in analyzing the data for the blocks and for the strokes separately.

The total numbers of trials in the study was 4416 (96 trials/block * 2 blocks * 23 participants). Of these 225 were error trials (5.1%). These were distributed across Blocks and Number of Strokes as shown in Table 3.

	Block 1	Block 2	Total
2 stroke	45	14	**59**
3 stroke	101	65	**166**
Total	**146**	**79**	**225**

Table 3: Distribution of 225 error trials between Blocks for 2 and 3 stroke gestures

Blocks and the number of strokes a gesture has are not independent in terms of the numbers of errors made (using Chi-Square(1), p = 0.033). Performance when making 2 stroke gestures improves more than 3 stroke gestures between the two blocks in terms of numbers of errors.

Two stroke gestures

Data from the second block of trials only was analyzed in order to study the impact of the independent variables on practiced performance. The means and standard deviations of the average completion times of the error-free gestures for each participant made in each treatment condition is shown in Table 4.

	Diagonal			Horizontal/Vertical		
	None	Leading Border	Zone Area	None	Leading Border	Zone Area
no targets	665.7 (227.9)	652.7 (189.8)	662.3 (217.9)	633.2 (223.4)	620.7 (144.0)	594.6 (182.3)
targets	617.7 (243.8)	650.1 (265.5)	569.3 (175.6)	616.6 (229.4)	575.3 (201.2)	560.0 (178.2

Table 4: Means and standard deviations of participant score (in milliseconds) for error free trials 2-stroke gestures, n = 23, Block 2 data only

The means, together with 95% within-participant confidence intervals, are shown in Figure 6. Here data from the individual trials is used and the between-participant variability is removed. This is done by normalizing the data between participants such that each has the same average [5]. This is only for visualization of the data and not for statistical comparisons.

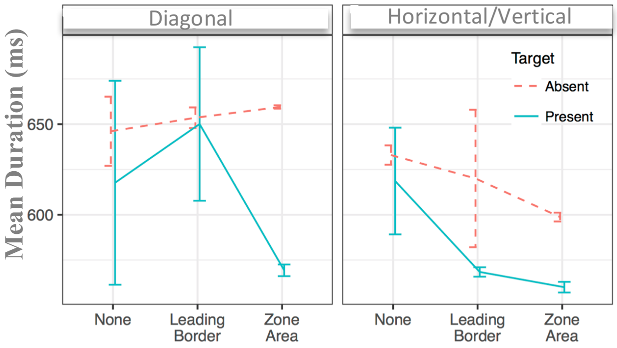

Figure 6: Mean values with 95% within-subject confidence intervals

An analysis of variance used a 3 factor within-participants repeated measures design with direction of eye movement (diagonal, horizontal/vertical), visualization of active zone (none, leading borders, zones), and presence of fixation targets (no targets, targets) as the factors.

Mauchly's Test indicated that the assumption of sphericity had been violated (Chi-Square (2) = 17.9, p < 0.001) therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (η = 0.635).

There were significant main effects of the direction of eye movement (F(1,22) = 12.4, p = 0.002), and the presence of fixation targets (F(1,22) = 5.9, p = 0.024), but not of visualization of zones (F (1.27,27.9) = 2.48, p = 0.12). There were no significant 2-way or 3-way interaction effects. The directions of these main effects were as predicted.

Gestures consisting of diagonal eye movements took significantly longer to make than those consisting of only horizontal/vertical eye movements.

Providing fixation targets enabled gestures to be made more quickly than when these were not provided.

There is no overall main effect for visualization of zones. As there is a main effect for direction of eye movement, we analyzed the data for horizontal/vertical gestures separately. We can assume sphericity (Mauchly's Test, Chi-Square (2) = 0.75, p =0.69) within this data. A 2 (no targets, targets) x 3 (no visualization, leading borders, zones) factorial analysis of variance showed that there is a main effect for visualization (F(2,22) = 4.2, p = 0.021) as well as for targets (F(1,22) = 4.77, p = 0.04). There was no significant interaction effect (F(2,44) = 0.24, p = 0.8). Examination of the significant effect of visualization by pairwise post-hoc comparisons, with the Bonferoni adjustment applied shows that only the *no visualization* and *zone area* conditions are significantly different from each other (p = 0.022).

So only in the case of horizontal/vertical 2-stroke gestures did visualization of the zone improve performance.

The total numbers of errors made summed across participants when attempting 2 stroke gestures is shown in Table 5.

	Practiced 2-stroke gestures			
	Diagonal		**Horizontal/Vertical**	
	no targets	targets	no targets	targets
None	4	0	1	1
Leading Border	1	0	1	1
Zone Area	2	0	3	0

Table 5: Absolute frequencies of errors, Block 2 data only, 23 participants made 4 gestures per condition = 92 trials per cell, a total of 14 errors in 1104 trials

While it should be remembered that the data refers to well-practiced behavior (Block 2 data only), the numbers of errors are very low. In the fixation target conditions, there are nearly no errors for either diagonal or horizontal/vertical eye movements.

Three stroke gestures

The corresponding data for the 3 stroke gestures is shown in Table 6, again from Block 2 trials only. The scores are the averages of the repeated error-free trials in each condition.

	Diagonal			**Horizontal/Vertical**		
	None	**Leading Border**	**Zone Area**	**None**	**Leading Border**	**Zone Area**
no targets	1122.8 (294.8)	1191.6 (299.7)	1195.6 (531.7)	1123.6 (292.4)	1123.6 (292.4)	1049.2 (277.6)
targets	958.2 (238.5)	942.0 (249.5)	1036.3 (329.3)	996.7 (302.5	1023.2 (330.3)	996.5 (260.5)

Table 6: Mean and standard deviations of participant scores (in milliseconds) for 3–stroke gestures per participant, Block 2 data only, n = 23

A similar visualization of the averages and 95% within-participant confidence intervals is shown in Figure 7.

Mauchly's Test indicated that the assumption of sphericity had been violated (Chi-Square (2) = 10.9, p = 0.004) therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (η = 0.711).

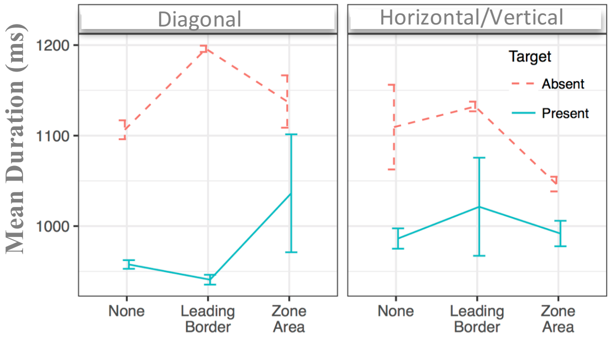

Figure 7: Mean values with 95% within-subject confidence intervals

There was a significant main effect of the presence of fixation targets (F(1,22) = 30.5, p < 0.001), but not of visualization (F (1.42,31.5) = 0.32, p = 0.65), nor of the predominant direction of eye movement (F(1,22) = 0.63, p = 0.44), and there were no significant 2-way or 3-way interaction effects. However, the interaction between direction of eye movement and fixation targets approaches, although does not reach, significance (F(1,22) = 3.94, p = 0.06).

The corresponding error data (Table 7) shows considerably more errors for 3 stroke gestures where there are no fixation targets in comparison with 2 stroke gestures. Most of these errors result from missing either the first or the second zone. These 'miss' errors are reduced by the extent of visualizing the zones. In the fixation target conditions, however, the error rate is almost zero, regardless of the direction of eye movement and the visualization of the zone.

Investigating causes of performance differences

We can examine the composition of gestures in terms of saccades and fixations to try to explain some of the observed differences in times to complete gestures. Why should gestures get faster with practice or with the presence of fixation targets? Are there fewer fixations, or shorter fixations, or both? The data provided by the eye tracker enables reliable segmentation of the time to complete a gesture into its component saccades and fixations. Saccades are ballistic and are generally very short in duration compared with fixations. We can examine possible reasons for the improvement in speed for 2 stroke gestures in terms of the fixations made during the gesture.

To recap, the gesture duration is from the start of the saccade leaving the center zone to the end of the saccade arriving back in the center zone. The minimum number of fixations is 1, so a *minimal gesture* consists a single saccade from the center zone followed by a single fixation in the outer zone and then a saccade back to the center zone.

	Practiced 3-stroke gestures			
	Predominantly diagonal eye movements		Predominantly horizontal/vertical eye movements	
	no targets	targets	no targets	targets
None	17 $(5+4^{a}+8^{b})$	1	12 $(4+3^{a}+5^{b})$	2
Leading Border	13 $(6+5^{a}+2^{b})$	1	5 $(1+2^{a}+2^{b})$	2
Zone Area	6 $(3+2^{a}+1^{b})$	0	4 $(2+0^{a}+2^{b})$	2

Table 7: Absolute frequencies of errors, Block 2 data only, 23 participants made 4 gestures per condition = 92 trials per cell, a total of 65 errors in 1104 trials, [a] missed 1st zone, [b] missed 2nd zone

Two findings involving improvement in speed of completion of 2 stroke gestures are examined here.

(i) 2 stroke gestures in Block 2 are made more quickly than those in Block 1

Table 8 shows the absolute numbers of 2 stroke gestures completed with 1, 2 and more than 2 fixations respectively. The duration values shown are the means of the single fixation durations or the means of the sum of the 2 fixations. The standard deviations of these distributions are shown in brackets. In Block 2, 87% of all gestures were completed at most 2 fixations. The difference between the distribution of frequencies between Block 1 and Block 2 is significant (Chi-Square (2), $p < 0.001$).

	Block 1			Block 2		
nr of fixations	n	cum %	duration (ms)	n	cum %	duration (ms)
1	384	36.3	436.9 (195.7)	544	49.9	413.3 (178.1)
2	487	82.2	563.5 (224.8)	404	87	481.1 (181.2)
3+	188	100		142	100	
total	1059			1090		

Table 8: Comparison of the numbers of fixations made in error-free 2 stroke gestures between Block 1 and Block 2, together with means and standard deviations

One reason then for the performance improvement with practice is that the proportion of minimal gestures increases from Block 1 (36.3%) to Block 2 (49.9%). It can be seen that the sum of the 2 fixations is far less than 2 x the duration of the single fixation. Breaking down the combined durations in row 2 of Table 8, Table 9 shows the mean durations of the first and second fixations respectively. This suggests that a *corrective saccade* is being made between the 2 fixations [28].

	n	1st fixation mean (ms)	2nd fixation mean (ms)
Block 1	487	200.7	362.7
Block 2	404	177.4	303.3

Table 9: For gestures completed with 2 fixations only, means of the first and second fixation durations respectively for Block 1 and for Block 2

If this is the case, then the effect of practice is to reduce the number of corrective saccades being made while making the gesture. Fully practiced performance could be tentatively assumed to be the case where all gestures are minimal gestures and made with 1 fixation only.

(ii) Gestures made with horizontal or vertical eye movements were made more quickly than those made with diagonal eye movements.

Table 10 offers a similar explanation for the difference found between the times to complete gestures with horizontal and vertical eye movements and diagonal eye movements respectively. The differences in the frequencies of gestures completed with 1 fixation, with 2 fixations and more than 2 fixations are significant (Chi-Square (2), p = 0.013). The probability of completing a horizontal-vertical gesture with one fixation is 0.538, while it is 0.461 for diagonal movement gestures. There is no apparent difference in the durations of the component fixations. The higher probability of making diagonal gestures with 2 or more fixations means that sample means of whole durations are higher.

nr of fixations	Horizontal /Vertical			Diagonal		
	n	cum %	duration	n	cum %	duration
1	293	53.8%	411 (165.8)	251	46.1%	416 (191.8)
2	194	89.4%	479 (185.4)	210	84.6%	483 (177.6)
3+	58	100%		84	100%	
total	545			545		

Table 10: Comparison between the numbers of fixations made while making error-free horizontal/vertical gestures and diagonal 2 stroke gestures respectively, Block 2 data only

DISCUSSION

The study shows that for practiced performance there is a clear relationship between the difficulty of making a gaze gesture and the amount of visual support provided on the display surface. This may be seen as self-evident. However, the results show the impact of different types of visual support, which in turn impacts upon how much information needs to be overlaid on the display with the attendant risk of interfering with the information being presented or viewed.

Two stroke gestures

The study shows that for 2-stroke gestures consisting of horizontal/vertical eye movements that almost error-free performance can be obtained without any additional visual support. These gestures can however be made more quickly with the addition of fixation targets, or by the visualization of the zones. There was no significant interaction effect between these 2 variables. Completion times of diagonal 2-stroke gestures are significantly longer. Here only the provision of fixation targets had a significant effect on completion time. Showing the area or the leading border outer zones around the center zone had no effect on completion time of diagonal gestures.

Considering practiced performance, the average completion time for horizontal/vertical gestures without fixation targets was 633.2ms, which dropped to 616.2ms when targets were provided. The corresponding times for diagonal gestures were 665.2ms and 617.2ms respectively.

Error rates for 2-stroke gestures with fixation targets were always low (<1%), and for horizontal/vertical gestures low even without targets.

Performance in terms of fixations

Measuring eye position with a very high sample rate eye tracker has enabled the compositions of gestures in terms of saccades and fixations to be studied. This enables us to offer explanations of why overall gesture completion time should be reduced under particular conditions. The notion of a *minimal gesture* is introduced to describe a gesture that is completed with a single fixation in a zone. The effect of practice was to increase the proportion of minimal gestures in relation to the proportion of gestures that were completed with 2 or more fixations. The average duration of a single fixation (in the region of 415ms) is much longer than those normally observed during other tasks. The average duration in silent reading is in the range 225 – 250ms, or for visual search in the range of 180 – 275ms [25]. As a 2-stroke gesture contains a reversal of direction of eye movement, then inhibition of return may contribute to the increased duration [23]. The combined durations of 2 fixations within a zone is much less than 2 times the duration of a single fixation, which suggests that gestures made with 2 fixations include a corrective saccade between these.

The difference between the times to complete horizontal/vertical gestures and diagonal gestures can also be explained in terms of the different proportions of minimal gestures. More horizontal/vertical gestures are completed with a single fixation than are diagonal gestures. The durations of the fixations are similar in each case.

Three stroke gestures

With 3-stroke gestures, very low error rates are obtained simply by adding fixation targets. This produces significantly faster completion times as well. Just by providing fixation targets with no other visualization, at least a 98% accuracy rate for 3 stroke gestures (2 or fewer errors in 92 trials) can be achieved, regardless of the predominant direction of eye movement. However, high error rates arise for 3 stroke gestures if fixation targets are not provided, as shown in Table 7. The majority of these arise from missing either the first zone or the second zone, particularly in the absence of any visual guidance about the location of the zone. While this might seem fairly obvious, it underlines the need for the designer to ensure that appropriate targets are always available when reliable gesture performance is expected.

Average completion times for 3-stroke gestures where only fixation targets are provided are below 1 second (958.2 ms where 2 of the strokes are diagonal and 996.7ms where 2 of the strokes are horizontal or vertical). This difference was not significant.

Comparison with other work

The results can usefully be compared with those reported by Rozado et al [26]. They reported a mean gesture completion time for a 3-stroke gesture of 1620 ms, and 2590 ms for the same gesture with a 500 ms dwell to signal the start of the gesture and a 500 ms dwell to stop it. Our 3-stroke gesture durations are shorter by a factor of a third. Also in that study, they reported error rates of 94% accuracy after practice when not using an explicit start and stop signal. They reported obtaining a 98% accuracy rate when gestures were made with the total extra time overhead of 1000 ms to start and stop the gesture. Rozado's study demonstrated that high accuracy and gesture durations of a comparable magnitude to those obtained in our study can be obtained using low-cost eye tracking equipment. As noted in the introduction, one of the main benefits of gaze gestures over other gaze interaction techniques is their robustness to less accurate eye position measurements and lower sample rates.

Implications for the design of a gesture scheme

Although simple horizontal/vertical 2-stroke gestures can be made quickly and reliably without any additional visual support, these can be expected to occur normally in the course of looking at the display. These would give rise to false positives (eye movements wrongly interpreted as gestures) if no explicit segmentation is used.

There are several ways to reduce the risk of false positives. One is to use more complex gestures. Different visual tasks will invoke different patterns of normal eye movements. We noted earlier that when using gaze gestures to interact with a computer game there were no unintentional 3-stroke gestures made during extended periods of normal game play [15]. Alternatively, a 'gesture' mode could be temporarily activated, say using a 3-stroke gesture or other device, where the center zone and fixation targets are made visible. Gestures would only be recognized in this mode. Two stroke gestures could be used in this mode as the user would know that their eye movements would be interpreted as gestures. This option carries an overhead of an action to enter and to leave the gesture mode. This is analogous to a pop-up menu appearing over an object in response to an activation event such as a right mouse button click.

Further work is needed to study how far the size of the fixation targets and the distance between the center zone and the targets can be scaled down without compromising the reliability and efficiency of making gestures to any significant extent. If the center zone was to be located over objects on the display, then context sensitive commands could be associated with gestures. However, the footprint of the gesture pattern would need to be reduced. The data from this study provides a baseline to study the size / performance trade-off.

Limitations of the current study and future work

The study described in the paper has not verified whether the results obtained hold true while performing the gestures over a variety of different information on a display. It can be argued that there will often be some feature on the display that could act as a fixation target. However, these will change as the displayed information changes and may be of inconsistent quality in terms of the ability to make reliable and efficient gesture patterns.

Another acknowledged limitation of the experiment reported here is that all of the data has been collected using a conventional desktop display. The results have not been verified using a near-eye display equipped with eye tracking capability. This remains as future work.

There is much further work in examining the saccades and fixations made within a gesture in order to understand more what affects performance while making them. This study has shown how improvements in completion time with practice and differences between diagonal and horizontal gestures can be attributed to the numbers of fixations made during gesture completion. The length of the fixations can be studied and in particular whether the inhibition of return mechanism [23] plays a significant role in this. If linear models of saccade duration in relation to the amplitude of eye movements made during gestures give sufficiently stable estimates, then modelling expert performance while making gaze gestures is possible. This could take the form of adding together estimates of the durations of saccades and fixations, and include the probability of a corrective saccade being made.

CONCLUSIONS

The study has investigated and shown how the speed and accuracy of making gaze gestures is related to the amount of information provided on the display surface to aid the user in making the fixations that form part of the gesture. The results show that in their simplest form, 2-stroke gestures using horizontal/vertical eye movements can be made reliably without additional support. Two stroke gestures made with diagonal eye movements and 3-stroke gestures only require the provision of small conspicuous fixation targets to be made reliably. The study has shown how differences in the speed of completing gestures due to practice and to the direction of eye movements can be explained on the basis of the number and duration of fixations made during the gesture. Understanding the composition of gaze gestures in terms of saccades and fixation may well provide a robust basis for modelling expert performance.

ACKNOWLEDGEMENTS

We would like to thank the participants for their time. Our thanks go also to Kari Räihä, Harri Siirtola, Jani Lylykangas, Jari Kangas and Oleg Spakov, all from the University of Tampere, for their input to this paper. Also, we wish to thank Sam Hutton of SR Research for his input. The work was supported by the Academy of Finland.

REFERENCES

1. Deepak Akkil, Andreas Lucero, Jari Kangas, Tero Jokela, Marja Salmimaa, and Roope Raisamo R. 2016. User Expectations of Everyday Gaze Interaction on Smartglasses. In Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16). ACM, New York, NY, USA, Article 24, 10 pages. DOI: https://doi.org/10.1145/2971485.2971496

2. Mihai Bâce, Teemu Leppänen, David Gil de Gomez, and Argenis Ramirez Gomez. 2016. ubiGaze: ubiquitous augmented reality messaging using gaze gestures. In SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications (SA '16). ACM, New York, NY, USA, , Article 11 , 5 pages. DOI: https://doi.org/10.1145/2999508.2999530

3. Wolfgang Becker. 1991. Saccades. In R. Carpenter, Vision and Visual Dysfunction, vol. 8: Eye Movements. Eye movements (pp. 95–137). London: Macmillan.

4. Nikolaus Bee, and Elisabeth Andre. 2008. Writing with Your Eye: A Dwell Time Free Writing System Adapted to the Nature of Human Eye Gaze. In E. André, L. Dybkjaer, W. Minker, H. Neumann, R. Pieraccini, & M. Weber, *Lecture Notes in Computer* Science (Vol. 5078, pp. 111-122). Berlin: Springer

5. Denis Cousineau. 2005. Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology,* 1(1), 4–45. http://www.tqmp.org/Content/vol01-1/p042/p042.pdf

6. Murtaza Dhuliawala, Juyong Lee, Junichi Shimizu, Andreas Bulling, Kai Kunze, Thad Starner, and Woontack Woo. 2016. Smooth eye movement interaction using EOG glasses. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (ICMI 2016). ACM, New York, NY, USA, 307-311. DOI: https://doi.org/10.1145/2993148.2993181

7. Heiko Drewes, and Albrecht Schmidt. (2007) Interacting with the Computer Using Gaze Gestures. In: Baranauskas C., Palanque P., Abascal J., Barbosa S.D.J. (eds) Human-Computer Interaction – INTERACT 2007. INTERACT 2007. *Lecture Notes in Computer Science*, vol 4663. Springer, Berlin, Heidelberg

8. Heiko Drewes, Alexander De Luca, and Albrecht Schmidt. 2007. Eye-gaze interaction for mobile phones. In *Proceedings of the 4th international conference on*

mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology (Mobility '07). ACM, New York, NY, USA, 364-371. DOI=http://dx.doi.org/10.1145/1378063.1378122

9. Henna Heikkilä, and Kari-Jouko Räihä. 2009. Speed and accuracy of gaze gestures. *Journal of Eye Movement Research,* 3(2), pp. 1-14. DOI: http://dx.doi.org/10.16910/jemr.3.2.1

10. Henna Heikkilä and Kari-Jouko Räihä. 2012. Simple gaze gestures and the closure of the eyes as an interaction technique. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (ETRA '12), Stephen N. Spencer (Ed.). ACM, New York, NY, USA, 147-154. DOI=http://dx.doi.org/10.1145/2168556.2168579

11. Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost van de Weijer. (2011). *Eye tracking: a comprehensive guide to methods and measures.* Oxford: Oxford University Press

12. Anke Huckauf and Mario H. Urbina. 2008. Gazing with pEYEs: towards a universal input for various applications. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (ETRA '08). ACM, New York, NY, USA, 51-54. DOI: https://doi.org/10.1145/1344471.1344483

13. Aulikki Hyrskykari, Howell Istance, and Stephen Vickers. 2012. Gaze gestures or dwell-based interaction? *In Proceedings of the Symposium on Eye Tracking Research and Applications* (ETRA '12), Stephen N. Spencer (Ed.). ACM, New York, NY, USA, 229-232. DOI=http://dx.doi.org/10.1145/2168556.2168602

14. Howell Istance, Richard Bates, Aulikki Hyrskykari, and Stephen Vickers. 2008. Snap clutch, a moded approach to solving the Midas touch problem. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (ETRA '08). ACM, New York, NY, USA, 221-228. DOI: https://doi.org/10.1145/1344471.1344523

15. Howell Istance, Aulikki Hyrskykari, Lauri Immonen, Santtu Mansikkamaa, and Stephen Vickers. 2010. Designing gaze gestures for gaming: an investigation of performance. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (ETRA '10). ACM, New York, NY, USA, 323-330. DOI=http://dx.doi.org/10.1145/1743666.1743740

16. Jari Kangas, Deepak Akkil, Jussi Rantala, Poika Isokoski, Päivi Majaranta, and Roope Raisamo. 2014. Gaze gestures and haptic feedback in mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA, 435-438. DOI: http://dx.doi.org/10.1145/2556288.2557040

17. Jari Kangas, Oleg Spakov, Poika Isokoski, Deepak Akkil, Jussi Rantala, and Roope Raisamo. 2016. Feedback for Smooth Pursuit Gaze Tracking Based Control. In *Proceedings of the 7th Augmented Human International Conference 2016* (AH '16). ACM, New York, NY, USA, Article 6, 8 pages. DOI: http://dx.doi.org/10.1145/2875194.2875209

18. Päivi Majaranta, & Andreas Bulling. 2014. Eye Tracking and Eye-Based Human–Computer Interaction. *Advances in Phys. Comp.,* 39-65. Springer, London.

19. Diako Mardanbegi, Dan Witzner Hansen, and Thomas Pederson. 2012. Eye-based head gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (ETRA '12), Stephen N. Spencer (Ed.). ACM, New York, NY, USA, 139-146. DOI=http://dx.doi.org/10.1145/2168556.2168578

20. Emilie Mollenbach, John Paulin Hansen, and Martin Lillholm. 2013. Eye Movements in Gaze Interaction. *Journal of Eye Movement Research*, 6(2), 1-15.

21. Emilie Mollenbach, John Paulin Hansen, Martin Lillholm, and Alastair G. Gale. 2009. Single stroke gaze gestures. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '09). ACM, New York, NY, USA, 4555-4560. DOI: https://doi.org/10.1145/1520340.1520699

22. Tomi Nukarinen, Jari Kangas, Oleg Špakov, Poika Isokoski, Deepak Akkil, Jussi Rantala, and Roope Raisamo. 2016. Evaluation of HeadTurn: An Interaction Technique Using the Gaze and Head Turns. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (NordiCHI '16). ACM, New York, NY, USA, , Article 43 , 8 pages. DOI: https://doi.org/10.1145/2971485.2971490

23. Michael Posner, and Yoav Cohen. 1984. Components of visual orienting. *Attention and Performance X: Control of Language Processes.* H. Bouma and D. Bonwhuis. Hillsdale, N. J., Erlbaum: 551-556

24. Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 85, 618-660.

25. Keith Rayner. 1995. Eye movements and cognitive processes in reading, visual search, and scene perception. In J. M. Findlay, R. Walker, & R.W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications* (pp. 3-22). Amsterdam: North Holland.

26. David Rozado, Javier S. Agustin, Francisco B. Rodriguez, and Pablo Varona. 2012. Gliding and saccadic gaze gesture recognition in real time. *ACM Trans. Interact. Intell. Syst.* 1, 2, Article 10 (January

2012), 27 pages. DOI=http://dx.doi.org/10.1145/2070719.2070723

27. SensoMotoric Instruments. SMI High Performance eye tracking hmd based on HTC Vive. http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/eye-tracking-htc-vive.html?gclid=CMC9g-zcndECFQcbaQoduK4Caw&cHash=bd18fc124ce9088b07a26a494fd2bed7

28. Jing Tian, Howard Ying, and David Zee. 2013. Revisiting corrective saccades: role of visual feedback. *Vision Research.* August 30; 89: pp. 54–64

29. Jacob O. Wobbrock, James Rubinstein, Michael W. Sawyer, and Andrew T. Duchowski. 2008. Longitudinal evaluation of discrete consecutive gaze gestures for text entry. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (ETRA '08). ACM, New York, NY, USA, 11-18. DOI: https://doi.org/10.1145/1344471.1344475

30. Xianjun Sam Zheng, Cedric Foucault, Patrik Matos da Silva, Siddharth Dasari, Tao Yang, and Stuart Goose. 2015. Eye-Wearable Technology for Machine Maintenance: Effects of Display Position and Hands-free Operation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 2125-2134. DOI: http://dx.doi.org/10.1145/2702123.2702305