Wrist-worn pervasive gaze interaction

Hansen, John Paulin; Lund, Haakon ; Biermann, Florian ; Møllenbach, Emilie; Sztuk, Sebastian ; San Augustin, Javier

Published in:
ETRA '16 : Proceedings of the symposium on eye tracking research and applications

Link to article, DOI:
10.1145/2857491.2857514

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Wrist-worn Pervasive Gaze Interaction

John Paulin Hansen*, Technical University of Denmark; Haakon Lund, University of Copenhagen; Florian Biermann & Emillie Møllenbach, IT University of Copenhagen; Sebastian Sztuk & Javier San Agustin, The Eye Tribe.

Abstract

This paper addresses gaze interaction for smart home control, conducted from a wrist-worn unit. First we asked ten people to enact the gaze movements they would propose for e.g. opening a door or adjusting the room temperature. On basis of their suggestions we built and tested different versions of a prototype applying *off-screen stroke* input. Command prompts were given to twenty participants by text or arrow displays. The success rate achieved by the end of their first encounter with the system was 46% in average; it took them 1.28 seconds to connect with the system and 1.29 seconds to make a correct selection. Their subjective evaluations were positive with regard to the speed of the interaction. We conclude that gaze gesture input seems feasible for fast and brief remote control of smart home technology provided that robustness of tracking is improved.

Keywords: Gaze tracking, input, mobility, pervasive technology, ubiquitous computing, smartwatch, smart home, hands-free interfaces, security and access systems.

Concepts: • **Human computer interaction (HCI)** ~ **Interaction techniques;** *Gestural input;*

1 Introduction

Imagine a cleaning operative looking at his gaze tracking smartwatch when approaching a door. Once eye contact has been established, the watch shows an arrow to the right. He looks to the right of his watch and back, which then opens the door. Inside the room, he increases the light by looking up and down from the watch to the dimmer to get the right illumination for his job. When finished cleaning, he updates the room status by a few gaze movements towards a wall monitor (c.f. Fig.1). For every such gaze command this person makes during his workday, a central security management system confirms his ID and handle permissions on basis of eye-feature recognition. Gazing the smartwatch confirms his intention to make a command. If he were just to use hand gestures for this, the risk of accidental activations would be high, especially for someone doing manual work in front of smart appliances.

*e-mail:jpha@dtu.dk

ETRA '16, March 14-17, 2016, Charleston, SC, USA © 2016 ACM.
ISBN 978-1-4503-4125-7/16/03…$15.00
DOI: http://dx.doi.org/10.1145/2857491.2857514

Our paper explores some of the preconditions for this future scenario. What are people's expectancies towards gaze control of smart homes that design should support? How may the watch guide gaze commands? What are the challenges for current gaze tracking technology to provide the accuracy, precision, responsiveness and stability required in a real work scenario?

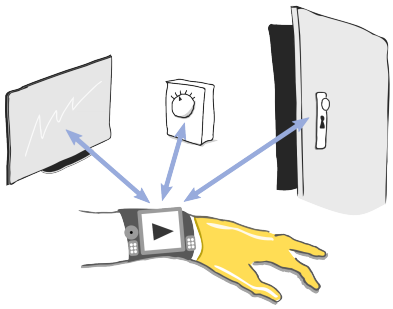


Figure 1: *Remote control by gaze strokes*

In regard to *how* the gaze tracking unit will eventually connect to the Internet of Things, we build on the feasibility of this happening through proximity sensing (e.g. Marquardt et al. [2012]). We also will not present research on eye-feature recognition from mobile devices. Iris recognition is a mature technology [Daugman 1993; Burge and Bowyer 2013] and major manufacturers recently introduced smartphones with user authentication by eye features. Instead, the main focus is on exploring usability and interaction issues of wrist-worn tracking. An effective system should provide almost 100% reliability in achieving and maintaining eye contact. This is a relevant issue, since we are designing a system where the combination of hand- and head movements creates additional noise to the tracking, not encountered by present-day head-mounted or stationary gaze tracking setups.

The proposed system is meant to support frequent operations, so it will have to be highly efficient. The operations addressed are commonly conducted with keys, key-cards, pin keypads, switches, dials and remote controls. In future intelligent environments, they may involve e.g. voice commands, hand-gestures, proximity sensing and smartphone-/smartglass- interfaces. Advantages of a gaze approach are that it imbeds the option of ID-confirmation; that it can be used hands-free (e.g. wearing gloves); and it works even in noisy or private locations, where voice-input would not be an option. But these advantages should not come at the cost of being more detaining than doing the same actions by conventional means.

So how fast can we expect a command by gaze to become – including the time it takes to position the arm, obtain eye contact, read the display, perform a gaze stroke and receive feedback?

Our approach entails one-directional gaze strokes. Gaze strokes are a simple form of gaze gestures [Møllenbach et al. 2013]. In our case they consist of a glance away from the unit in one of four directions, and back to the unit again. Some researchers (e.g. Esteves et al. [2015]) have warned that users may consider it unnatural to move the eyes in a particular pattern without visual support for doing so. Consequently, we will first study users spontaneous reactions to the idea of controlling a smart environment with gaze strokes. Subsequently, we will test users performance and consider their subjective evaluations of various prototypes. The experiment presented in this paper will compare three different prompts for a gaze stroke: single words ("up", "down", "left" or "right"), short text (e.g. "look up now") and arrow icons. Also, we will test if additional leading light mounted on and around the unit may serve as visual support for their strokes.

The paper offers 1) an elicitation study of users intuitions about pervasive gaze interaction with a smartwatch. 2) Designs for handheld gaze interaction based on visual feedback from the watch and from external LED lights. 3) A study of the basic user performance of a wrist-worn gaze input system in terms of errors, time-to-connect, selection time, accuracy and precision. 4) User evaluations of gaze stroke interaction with prototypes of a wrist-worn tracking device.

2 Related Work

Several studies, (e.g. Drewes et al. [2007]; Dybdal et al. [2012]; Rozado et al. [2015]), has found *on-screen gaze gestures* to be particularly efficient and robust for small screen gaze interaction, since no exact determination of fixations are required, only detection of eye movement directions. This is similar to finger swipe-gestures that can be done anywhere on a touchscreen when it is only the direction that matters.

The idea of utilizing *out-of-screen gaze gestures* was originally conceived by Isokoski [2000]. Four off-screen fix-points were placed on the frame of a monitor. All the letters in the alphabet could then be composed by different gaze strokes between the four points. Kangas et al. [2014] used off-screen gazing as input to a mobile phone. They found tasks to be completed faster and rated easier and more comfortable with a vibro-tactile confirmation on the gestures than without.

Akkil et al. [2015] presented the first study of gaze interaction with a smartwatch. A scene camera in a head mounted tracker recognized the position of the watch display. Participants rated haptic feedback significantly more comfortable than visual feedback for simple confirmations when a notification had been glanced. There were no clear preferences for gaze gesture feedback in a more complex menu navigation task. Esteves et al. [2015] presented a smartwatch interaction technique, Orbits, with targets moving circularly. Smooth pursuit eye movements, following the targets for 0.5 to 1.3 seconds, select the controls, with true-positive rates up to 0.96. The main difference to our research is that we presume all system components united in one smartwatch while Akkil et al. [2015] and Esteves et al. [2015] include both a smartwatch and a pair of smartglasses.

Hansen et al. [2015] measured precision and accuracy of a wrist-worn tracker to be around 3°, which makes it impossible to detect where on the smartwatch screen people are looking. Consequently, they recommended the use of off-screen gaze-gestures and presented promising user responses to this interaction principle. In this paper we extend some of the ideas introduced by Hansen et al. [2015] and Hansen et al. [2015a].

Shell et al. [2003] addressed gaze control of smart home devices through the design of Media EyePliance, a system that allowed appliances to be selected through gaze and subsequently controlled with a remote keyboard or voice. The tracker was fixed on the appliance and the goal of the interaction was not detailed gaze interaction with the functionality of the appliance, but gaze as object selection.

In summary, mobile gaze interaction is well studied. Smartwatch gaze interaction by a head-mounted setup recently gained research attention, while single-device wrist-mounted tracking is still unlit.

3 Elicitation Study: Enacting Gaze Control

We conducted our first user study to explore how people spontaneously would choose to interact with smart devices using a gaze controlled wrist unit and how they would explain their interactions.

We invited ten colleagues from our university who were all familiar with the basics of pervasive technology (6 males, 4 female, aged 30 to 43 years), to an individual, 30-minute interview. They were shown a mock-up of a smartwatch with a gaze tracking unit and explained that when fully developed, it would be able to record their eye movements if turned towards their face. We asked them to imagine that this could remotely control future appliances, for instance a TV or lamp.

They were then given a smartwatch and a pair of SMI Eye Tracking Glasses 2 to wear during the session. A one-point calibration with the tracker was performed. Twelve simple tasks were presented one by one, each written on a card attached on top of the smartwatch. The task questions concerned remote gaze control of the door, a TV, a thermostat, a smartphone and the smartwatch itself. All of these objects were physically present in the room. We asked them to perform an actual gaze pattern for every task presented. The questions should be read aloud before acting, for instance: "How would you move your eyes to close the door?"; "How would you move your eyes to decrease the temperature?"; "How would you move your eyes to activate your mobile phone?"; or "How would you move your eyes to go back in the menu on your smartwatch?". The order of questions where changed for each participant to minimize carry-over effects. When they enacted their answer, they were asked to explain the rationale behind the pattern they had just performed. Their eye movements were recorded for analysis, but did not have any actual effect on the smart devices or the watch.

3.1 Observations

In all cases, the participants would turn towards the controllable object and look at it, either before or during an input sequence. In 40% of the cases the object was gazed at before input was envisioned and in 60% the object was attended during the input sequence. When the object was not part of an input sequence, people would often look at something else in the preferred direction. Several participants told us that it could be difficult to just look in a certain direction, affirming the necessity of fix points. Having an object to look at solved the problem, especially if this object could be attended close to the tracking unit.

Only in 3 out of 120 cases would a participant just look into empty space without an apparent fixation point. In 14% of the cases the subject would prefer to look somewhere else than up, down, left or right; one third of these unusual movements would be a diagonal gesture. Thus, a field definition with four areas would cover 86% of the cases, c.f. Figure 2.

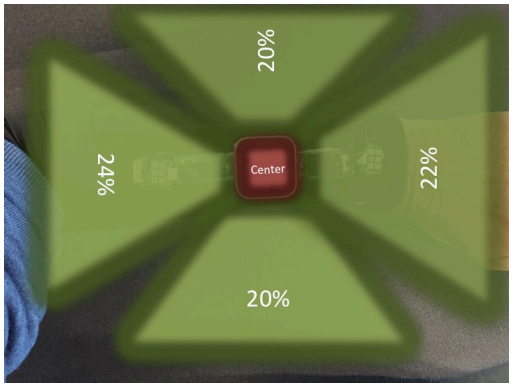


Figure 2: *Percentages indicate how often a particular field was included in a command. Center field was always attended because the tasks were presented here.*

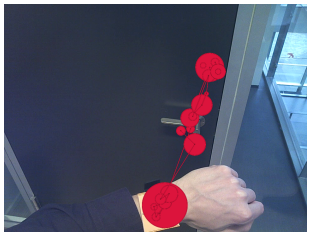
Wording of the questions had a strong impact on the preferred directions. When asked "How would you move your eyes to go **back** in the menu on your smartwatch?" eight of the participants went to the left, because - as they explained - "this is how you go back in a browser". When we changed the question to "Go **up** in the menu", 9 out of 10 participants would look up. Prepositions (up, down, on, below, above etc.) in the questions seem to have strong implications on eye movement preferences.

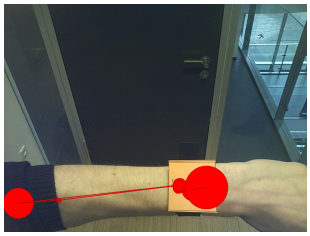
Other questions revealed mixed patterns. For instance, 4 participants would increase the temperature by looking up (c.f. Fig. 3), while two of them would look to the right, because "this is the direction you turn a thermostat to increase the heat". Apparently, in some cases the mental models of operation also primes eye movements. Finally, the physical affordance of objects may determine the preference for directions, for instance looking to the left to open a door, because it turned inwards. Figure 3 provides further examples of suggested gaze control patterns.

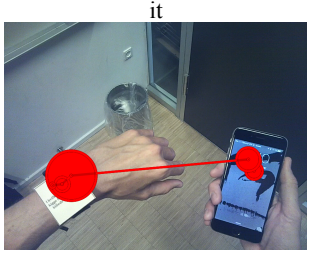
Most of the participants had difficulties deciding on a gaze pattern for turning off the TV, complaining that there is a large variety in spatial movements normally associated with this action. Some of the participants tended to suggest more advanced gestures for it, like swiping, or gestures involving diagonal movement and multi-strokes. One subject just wanted the TV to turn off when not attended; i.e. a "look-away" –command; similar to what Shell et al. [2003] suggested in their pioneering work on gaze responsive environments.

By the end of the tasks, we asked the participants if they would prefer a smartwatch with touch- or gaze- input. Seven out of ten would rather use their fingers, because they were familiar with this input, and because they felt it constraining to hold the hand perpendicular to the face for gaze commands. Some of them mentioned, though, that instant and reliable gaze interaction would be apt for hands-free interaction, for instance to open the house door when carrying shopping bags, and that it would be invaluable for people with motor disabilities.

The participants tended to suggest simple stroke patterns, guided by ad-hoc fixation-points. They would often position themselves so the object controlled could serve as a natural place to fixate. Four fields (up, down, left and right) covered a majority of the directions suggested. Except for the turn-off action, participants readily came up with gaze stroke actions that made sense to them. Most of their explanations could be related to either the wording of the task instructions, their mental models of operation, or the affordances of the objects controlled. The fact that some of the participants felt constrained by having to hold the hand towards the face emphasizes the importance of providing a tracking box size that is optimized for smartwatch gazing, e.g. with a wider angle towards the lower part.


Opening the door by looking at it


Opening the door by looking to the left


Activating mobile phone by looking at it

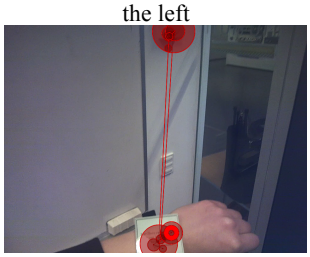

Increasing temperature by looking up

Figure 3: *Examples of enacted gaze control movements*

Our participants did not consistently look back at the watch after performing a stroke. However, if the smartwatch had provided a visual feedback confirming the input, we would probably have got more regular forth-and-back gaze stroke patterns.

On basis of these observations we built and tested a gaze tracking unit applying forth-and-back gaze stroke interaction in the four directions most commonly used. The unit prompts for a stroke in a particular direction, assuming that the user has only one option for action, e.g. to turn off the light or open the door. Later in the general discussion section we will address situations with more than one option.

4 User Test

The purpose of the user test was to examine the basic operations involved in a forth-and-back gaze stroke. How fast can users position their eyes in front of a tracking camera worn on the wrist? How may people best be prompted to move their eyes in a particular direction?

4.1 Participants

Twenty people from our university (14 male, 6 female, aged 20 to 55 years, mean = 31.1 years) volunteered to participate. Fifteen of the subjects had tried gaze interaction before. They all had normal or corrected-to-normal vision; 6 of the subjects were wearing glasses or contact lenses. One subject could not be tracked and all data from this person were excluded from the analysis.

4.2 Procedure

The first experiment measured the *contact time*, which is the delay between a notification and the user successfully making eye contact with the wrist-worn tracker. A laptop screen positioned in front of them displayed a countdown to focus their attention. At pseudo-random points during the countdown, the laptop display would show "Go!", which the participants had to respond to by looking down at their wrist-worn eye tracker.

When an eye contact had been established, i.e. when the tracker had detected the presence of the user and successfully achieved tracking of both eyes, the system would provide a short sound feedback, and a new countdown would start at the laptop monitor. If no contact had been established within a timeout period of 3000ms, a new countdown would start. Participants repeated this "time-to-connect" task ten times seated with their arm resting on a table.

We then conducted a second experiment to record fixation data to assess the *precision and accuracy* of the wrist-mounted tracker, asking each participant to look at an LED in each corner of the display, which would light up for 800ms in a pseudo-random order. In addition to the corner LEDs there was also a fixation marker displayed in the middle of the display itself. The LED were spanning a rectangle of 53mm(w) x 39mm(h). Each LED and the fixation marker in the middle of the display were turned on twice.

During the final experiment, *the interaction test*, the display would tell in which direction to look next. The prompts were given in one of three formats:

- *Arrow:* by an arrow-head pointing in one of four directions (up, down, right or left) c.f. Figure 4a and 4d;
- *Word:* by a single-word text command ("up", "down", "right" or "left") c.f. Figure 4b;
- *Text:* by a short sentence, displayed word by word in the so-called rapid serial visual presentation (RSVP)-format (e.g. Benedetto et al. [2015]; Hansen et al.[2015a]). The sentences consisted of three consecutive words: (1)"Look…. (2) up/down/left/right….. (3) now!" c.f. Figure 4c.

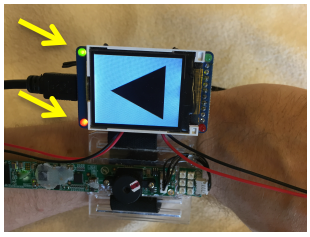
Each of the four directions would be prompted three times in all three of the above formats and under one of three conditions of visual aid:

- *No aid*
- *Corner LEDs:* When prompted to look up, two LED´s in the top corners would light up; when prompted to look left, two corner LED´s in the left side would light up – and so forth, c.f. Figure 4a and 4c. The subjects where told not to look at the corner LEDs themselves but in the general direction of the two LEDs. Only looking at the corner LEDs would not make an input.
- *External LEDs*: A LED was fixed on the ring finger and another LED was mounted up the arm, cf. Figure 4b and 4d. The Subject was told that they could use the external LEDs as a fixation point if the liked to.

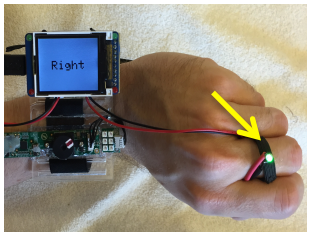
Consequently, each block had 36 prompts: 4 directions x 3 formats x 3 visual conditions. Each subject did 3 blocks with a 20 sec. pause between each block, producing a total of 108 observations. Prior to the experiment we had shuffled the order of directions, formats and visual conditions pseudo-randomly; this order was consistent across participants.

Participants were instructed to move their eyes as quickly as possible somewhere in the prompted direction and then back again, without us specifying a fixation-point to look at, except for the recommendation to use the external LED´s when they were available. We deliberately prioritized speed over robustness because we wanted to examine if stroke input would at all be fast enough for a work scenario. Our expectations are that this might have caused more errors for our novice participants, but would likely decrease with practice.

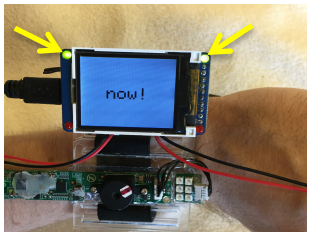
The three experiments finished with a short interview. A full session lasted approximately 30 minutes per subject.



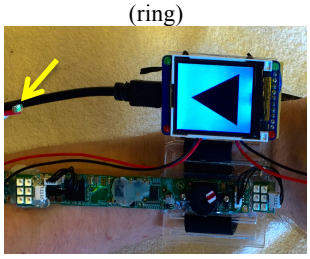
a) Arrow head with left-corner LEDs



b) Word with external LED´s (ring)



c) Text (last word of "Look up now!") and top-corner LED´s



d) Arrow head with external LED´s (arm)

Figure 4: *The 3 different prompt formats (arrow head, word and text) with corner LED´s (a and c) and external LED´s (b and d). Yellow arrows point at the LEDs.*

Saccades are fast eye movements with velocities above 100º/s and durations around 30 to 100ms (depending on amplitude). Due to the low frame rate of the eye tracker (30 frames per second, see next section), it is not possible to have reliable gaze estimates while the eye is performing a saccade, and therefore the direction of the saccade cannot be determined accurately. Instead, we configured four fields in a Maltese cross around the center (c.f. Fig. 2). When the gaze had fallen within one of the four areas for 150ms (i.e. five consecutive gaze data frames) and moved back to the center, this would count as selection in that direction, and a new prompt would be given after a brief delay. An audio click feedback would notify if the right area had been hit. Each prompt had a timeout of 3000ms.

When tracking was lost, the display background would immediately turn gray and users where then to try regaining contact.

4.3 Apparatus

A gaze tracker from The Eye Tribe was used for the experiment. We disassembled the tracker and mounted its circuit board (size LxHxD: 12 x 2 x 1,5 cm, weight 20g), with a camera and two near infrared light sources, below the display and LEDs (c.f. Fig. 4). This tracker records binocular gaze data at 30Hz, with a claimed accuracy of 0.5 to 1 degree (under stationary conditions). We adjusted the lens focus to 30cm - approximately the distance between the wrist and the eyes for adults. Images from the camera were transmitted through a USB 3.0 connection to a PC running a modified version of the tracking software that supports interaction without individual user calibration and tolerates head movements within a tracking box of X = 30 cm, Y = 20 cm and Z = 12 cm at 30 cm distance. The tracking method uses the relative differences between the pupil and the glints to determine at which location (with respect to the watch) the user is looking. This provides limited accuracy. A final implementation would most likely make use of an individual calibration process that could improve the accuracy and precision of the system. However, since we are applying very large input fields we decided to avoid the calibration in order to run the experiment more quickly, (c.f. Fig. 2). The PC also ran the software for the experiment and logged all user data.

A 1.8" Adafruit TFT screen with a resolution of 160 x 128 pixels, size 50mm (w) x 35mm (h), controlled by an Arduino Nano board, was connected to the PC via USB and mounted above the tracker. The display, LED and tracker board were assembled on a plastic board and mounted to the wrist with Velcro (Fig.4).

4.4 Results

The average time-to-connect was 1276ms, S.D. = 515ms. In 14% of the cases, the participants could not make connection within the 3000ms time-out period. The most successful subject got 100% connection, within an average of 640ms.

For every activation of each target (i.e. turning on one of the four LED in the frame corners or displaying the central marker on the screen, see Sect. 4.2) we computed the root mean square (RMS) of the visual angle θ between coordinates to measure precision:

$$\theta_{RMS} = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}\theta_i^2}$$

Furthermore, we computed the accuracy as the average angular offset [Holmqvist et al. 2011] (c.f. table 1).

$$\theta_{offset} = \frac{1}{n}\sum_{i=0}^{n-1}\theta_i$$

Table 1: *Contact-time, precision and accuracy (in visible degrees) for wrist-worn tracker.*

Contact time (ms)	Precision (θ_{RMS})	Accuracy (θ_{offset})
1276 ± 515	2.988 ± 0.122	2.966 ± 0.144

Before analyzing the results from the main interaction experiment we removed 61 outliers with a response time below 100ms. Figure 5 shows that there was a learning effect on the overall performance across the 3 blocks. Correct responses went up from 27 % in block A to 46% in Block C and errors decreased from 38% in block A to 25 % in block C. Timeout went from 35 % to 29%. There were no significant difference in errors between block B and C $\chi^2(1, N = 1043) = 0.007$, $p > .05$, so further data analysis will be based on data from these blocks, with more stability in performance. We use chi-square tests for comparing the frequencies, because each trial would terminate if an error were conducted - i.e. a movement in a wrong direction - or it would timeout, if no selections had been made within the 3000ms period, providing 3 possible event categories.

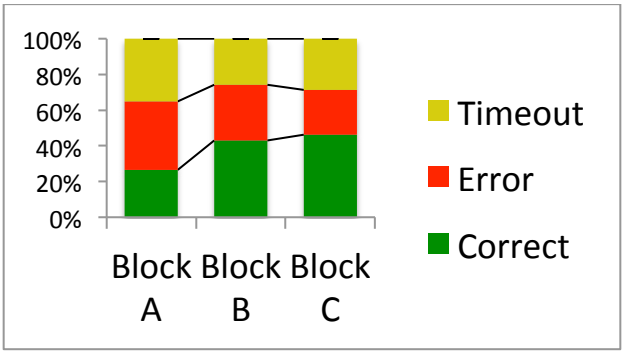


Figure 5: *Overall performance for interaction experiment with stroke gaze gestures, 19 subjects in three blocks.*

The grand mean error frequency was 32%; S.D. = 43%. There was a high variation in error rates among the participants (Fig. 6). One participant made no errors at all (subject 10), while two participants (6 and 18) committed more than 60% errors.

Errors for single words (43%, S.D. = 49%) were significantly more frequent, $\chi^2(2, N = 1043) = 31.86$, $p < 0.001$, than for text (25%, S.D. = 43%) and arrow prompts (26%, S.D. = 44%). Although errors were slightly less frequent for external LEDs (27%, S.D. = 44%), the difference was not significant, $\chi^2(2, N = 1043) = 3.59$, $p = 0.16$, compared to frame LEDs (32%, S.D. = 47%) and no LEDs (33%, S.D. = 47%).

The grand mean selection time for correct trails were 1.29 seconds; S.D. = 0.54 seconds. The individuals showed some variation; the fastest subject (6) performing with an average of 0.92 seconds and the slowest (1) at 1.52 seconds (c.f. Figure 7). The two fastest subjects (6 and 18) also had the highest error rate, indicating a speed-accuracy tradeoff.

We conducted a 2-way ANOVA with prompt and LED as independent variables. The ANOVA showed no effect of prompt types ($F_{2,1042} = 0.387$, $p > .05$), no effect of leading LED ($F_{2,1042} = 0.362$, $p > .05$ and no interaction effects ($F_{4,1042} = 0.357$, $p > .05$).

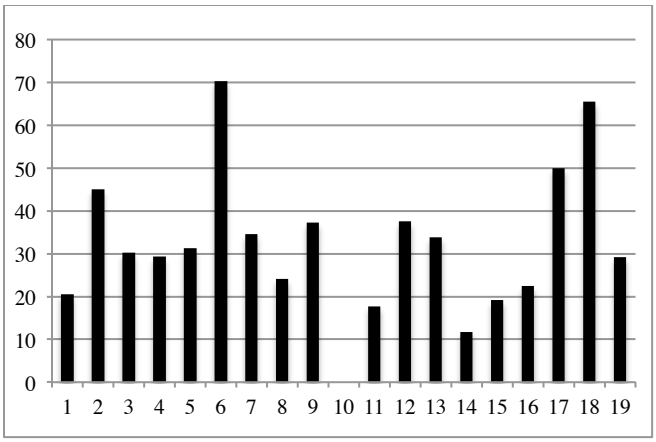


Figure 6: *Error percentages for 19 subjects, block B and C.*

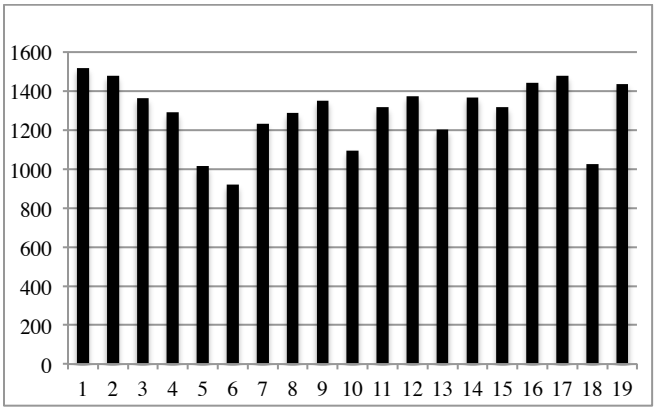


Figure 7: *Selection times in milliseconds for correct trials; 19 subjects, block B and C.*

The subjective evaluation asked the participants to rate on a scale from 1 to 7 how easy, fast and pleasant they deemed the system (7 being very easy/fast/pleasant). Speed was rated highest (= 4.2, S.D. = 1.6), with pleasantness (= 3.1; S.D. =1.4) and easiness (= 3.75; S.D. = 1.3) being average. A Wilcoxon Signed Rank test showed the difference between speed- and pleasant- rating to be significant, $Z = -2.227$; $p < 0.05$.

When asked about the preferred type of prompt, a large majority of 14 participants favored arrows, because they were "more suggestive", "very intuitive" and gave "no need to translate word into directions". Three people preferred single words and two people preferred the text.

Eleven participants preferred the LEDs in the frame because they were close to the display information they were looking at, even though they would not to use them as fixation points, only indicators of directions, while the external LEDs was perceived too far away to be effective. Five people preferred the LEDs at the sleeve and ring, because it provided them a fixation point. Three subjects would rather not have any LEDs at all because they were blinding, irrelevant or confusing.

Finally, we asked the participants to suggest future applications for wrist-worn gaze interaction. The following ideas came up (numbers indicate how often): Smart home control (5), smartwatch control (4), an ubiquitous input device (4), games (3), music player (3), assistive technology (2), typing (1), zooming (1), security/ID (1), and "don't know" (1).

In summary, the user test found time-to-connect to be around 1.3 seconds in average. The interaction test found the best average hit rate to be 46%. Selection time was 1.29 seconds which participants evaluated as fast. Single word prompts were more error-prone than arrows and text, while LED guiding did not have any significant impact on performance. Participants favored the arrow prompt and the frame LED´s most.

5 General Discussion

Off-screen horizontal and vertical gaze gestures on a wrist-worn remote controller seems interesting in light of two main findings: Most (i.e. 86 %) of the gaze commands that our participants generated spontaneously could be captured by four directions that they intuitively would suggest to use. Secondly, participants were able to make selections in the Maltese cross structure within 1.29 seconds and it would only take them 1.27 seconds to connect with the tracker. In total, this makes it possible to e.g. unlock a door in less than three seconds; our fastest subject might do it in just two seconds. For an informal comparison, we asked five of our subjects to enter a four-digit pin code on their smartphone five times. In average, this took them 6.6 seconds when they did it the first time and 4.4 when they did it the fifth time.

However, some critical issues require further research. The fact that only 71% of the entries could be made before timeout (in Block C) points to the importance of improving the tracker performance for real-life work applications. Also, the difference between our wrist-worn prototype (i.e. RMS = 3°/ offset = 3°) and the claimed performance of stationary trackers (i.e. RMS = 0.5 - 1°/ offset = 0.1°) is substantial. The high error in precision and accuracy is a result of using a gaze tracker in a wearable setting, where not only head but also hand movements interfere with the tracking performance. This implies that it will be challenging to use interface elements such as dwell-time activated buttons efficiently together with a wrist-worn tracker. Use of a lens with a larger field of view and motion sensors to counteract hand movements, could potentially improve performance. Use of smooth pursuits interaction (c.f. Esteves et al. [2015]) and larger-than-screen input areas may be other promising ways to counter the inherent noise in a smartwatch set-up.

Another way to improve performance is, of course, by training. All of our subjects were novices with regard to the task, and further longitudinal studies are needed to clarify, whether expertise will make interaction more reliable. One of our subjects had a 100% hit-rate, which makes it likely, that a wrist-worn system could work well for sufficiently trained professionals.

Even though the guiding LEDs had no significant effect, most of our participants seemed to like them. This suggests re-designing them, for instance by lightening up the full side frame and not just the corners. Comments from some of our subjects that the LEDs were blinding suggest that this should be done more saliently. Including additional vibro-tactile feedback, that previous research has shown effective [Akkil et al. 2015; Kangas et al. 2014], may also be considered.

Perhaps the most critical design issue is to help users direct and keep the camera pointing towards their eyes. In the current experiment we provided feedback when tracking was on by switching the display from gray to bright, but this may not be enough. Hansen et al. [2015] suggested using a moving cross-hair in the display to warn early when eye-head location approached

the border limits of the cameras tracking box, but this idea has yet to be evaluated.

The participants were most in favor of arrows pointing to the direction they should look. Single word prompts produced more errors, while RSVP-text prompts did just as well as the arrows. When there is only one action possible, (i.e. just one smart device in the proximity), and when only one input is needed, an arrow may be most relevant. If more devices are available, or if there are several interaction possibilities with one device, an elaborated display may be needed, either with a word-by-word explanation (cf. Hansen et al. [2015a]) or in a graphical user interface. Figure 8 suggests how to select the TV with a gesture to the left among 3 smart appliances currently in the proximity (1). A correspondence between the spatial location in the room and the display can be made if the smartwatch knows its orientation towards each appliance. Once the TV has been chosen, only options for this will be available (2). Volume adjustment is then selected with a gesture to the right (3).

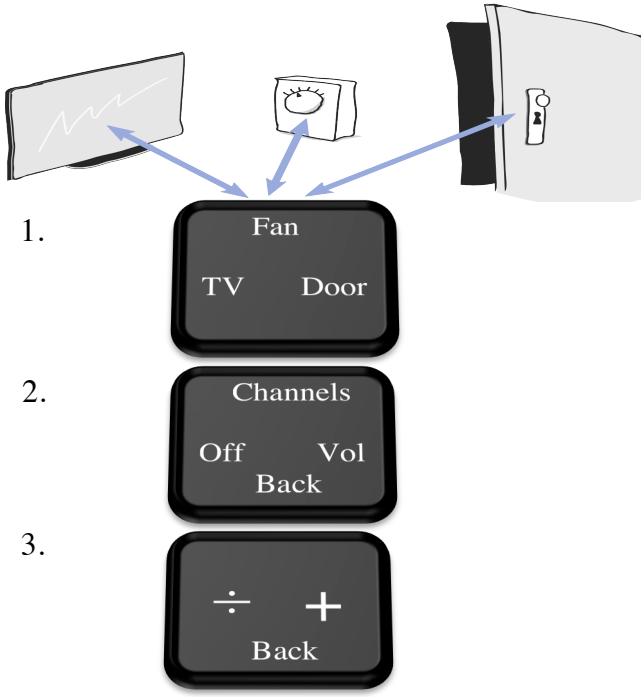


Figure 8: *The smartwatch connects to 3 smart appliances in the proximity and present them with a corresponding spatial mapping on the display (1). Selecting TV for adjustment with a gaze gesture to the left makes it possible to adjust the volume (2) with a gesture to the right (3).*

There will be a high risk of false selections on an interface with several options like this, especially for novice users; our experiment indicates that 30 to 40% of them may be wrong. This makes it mandatory to provide an easy back -option or allowing starting over by briefly braking the eye contact and then attending the smartwatch again. Future research will be needed to clarify how much of a problem this may be in daily operations and redesigns will have to be made accordingly.

Power consumption is highly important for mobile- and smartwatch systems [Rawassizadeh et al. 2014]. This issue has not been addressed in the present paper, but there may be potentials for savings by reducing the frame rate to e.g. 10Hz and adjusting the illumination of the IR-LEDs, e.g. by synchronizing them with the camera frame capture. Further research is needed to test how this affects tracking and power consumption. For professional use, like the cleaning operator scenario, it would be conceivable carrying extra sets of batteries, which might eventually keep the system charged for a full workday, since interaction only happens in short bursts.

There are numerous possibilities for combining gaze as an input to touch, voice, gesture and hand- tracking. For instance, a wrist-worn gaze tracker would likely suffer from accidental command completion: a potential overlap between natural search patterns and gaze input patterns. This risk might be reduced if an intention to initiate interaction is to be confirmed by always first looking directly at the device and with a short time-out for inputs thereafter. However, in any case the gaze camera would have to be in stand-by for a burst input. To save substantial amounts of power, the gaze camera should only turn on when acceleration and orientation sensors indicates that the unit has been moved quickly to a vertical position and then turn on the swatch display if gazed at. Finally, power might be saved if it turns off immediately when not attended.

The handheld gaze input method is particularly relevant to consider for use-cases where security plays a major role [De Luca et al. 2007]. Since it is the core of our concept to have continuous contact with the user's eye during engagements, it is possible to take advantage of the eye features that makes each and every person unique. Iris pattern recognition has been known for decades to be a very reliable biometric method [Burge and Bowyer 2013] and recent research [Holland and Komogortsev 2013] includes motion features as well. It is still an open research question, though, how to best get high-resolution images of the iris and eye when the user is holding the camera on the wrist.

Two recent papers (Akkil et al. [2015]; Esteves et al. [2015]) combine a smartwatch with head-mounted gaze tracking. However, we suggest embedding the eye tracker in the wrist-worn unit for two reasons: 1) only one device is then required and 2) wrist-worn units have the affordances of being non-intrusive, compared to head mounted displays, i.e. they don't cover the face and are therefore a more visually discreet technology. The disadvantage of a wrist-worn set-up is the low accuracy and precision.

The current work hinted us with other areas for future research. A wrist-worn tracker with acceleration and orientation sensors may identify a smooth motion of the hand. This opens up the possibility of making reliable hand-gesture inputs that will only be acknowledged in combination with gaze, avoiding false activations happening just because people move their arm. Secondly, another possibility lies in intelligent environments with several smart devices using gaze gestures as a metric for interaction: If the user looks consistently in one direction, the device would automatically lock on to the closest device in that direction for activation. The wrist-worn tracking unit would have to be placed in-between the locations of the smart devices in a way that would match its sensitivity for gaze directions. In our current version, we only support four directions (cf. Figure 2) but future systems could possible distinguish more. Most importantly, the smartwatch would also have to know is current orientation relative to the smart devices in the present proximity.

Finally, throughout this paper we have been referring to a work scenario. Evidently, the usefulness of a mobile remote controller

operated by gaze would be particularly welcome by many people with motor disabilities. The remote controller may be mounted on a wheelchair or placed in front of the user, occupying much less space than current screen-based gaze trackers for assistive systems.

5 Conclusion

We have explored off-screen gaze gestures for interaction with a wrist-worn unit. A four-region input field achieved a success rate of 46% with an average connection time of 1.28 seconds and a selection time of 1.29 seconds. The efficiency makes wrist-worn gaze interaction interesting for special use cases and motivates further research in improving robustness and comfort.

Acknowledgements

The Danish National Advanced Technology Foundation supported this research. Thanks to Lars Yndal Sørensen for system development and to the anonymous reviewers for their comments.

References

AKKIL, D., KANGAS, J., RANTALA, J., ISOKOSKI, P., SPAKOV, O. AND RAISAMO, R. 2015. Glance Awareness and Gaze Interaction in Smartwatches. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, ACM, 1271–1276.

BENEDETTO, S., CARBONE, A., PEDROTTI, M., LE FEVRE, K., BEY, L.A.Y. AND BACCINO, T. 2015. Rapid serial visual presentation in reading: The case of Spritz. *Computers in Human Behavior* 45, 352–358.

BURGE, M.J. AND BOWYER, K. 2013. *Handbook of iris recognition*. Springer.

DAUGMAN, J.G. 1993. High confidence visual recognition of persons by a test of statistical independence. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 15, 11, 1148–1161.

DE LUCA, A., WEISS, R. AND DREWES, H. 2007. Evaluation of eye-gaze interaction methods for security enhanced PIN-entry. *Proceedings of the 19th Australasian conference on computer-human interaction: Entertaining user interfaces*, ACM, 199–202.

DREWES, H., DE LUCA, A. AND SCHMIDT, A. 2007. Eye-gaze Interaction for Mobile Phones. *Proceedings of the 4th International Conference on Mobile Technology, Applications, and Systems and the 1st International Symposium on Computer Human Interaction in Mobile Technology*, ACM, 364–371.

DYBDAL, M.L., SAN AGUSTIN, J. AND HANSEN, J.P. 2012. Gaze Input for Mobile Devices by Dwell and Gestures. *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM, 225–228.

ESTEVES, A,. VELLOSO, E., BULLING, A. AND GELLERSEN, H. 2015. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, ACM, 457–466.

HANSEN, J.P., BIERMANN, F., MADSEN, J.A., JONASSEN, M., LUND, H., SAN AGUSTIN, J. AND SZTUK, S. 2015. A gaze interactive textual smartwatch interface. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, ACM, 839–847.

HANSEN, J.P., BIERMANN, F., MØLLENBACH, E., LUND, H., SAN AGUSTIN, J. AND SZTUK, S. 2015a. A GazeWatch Prototype. *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, ACM, 615–621.

HOLLAND, C.D. AND KOMOGORTSEV, O.V. 2013. Complex eye movement pattern biometrics: Analyzing fixations and saccades. *Biometrics (ICB), 2013 International Conference on*, IEEE, 1–8.

HOLMQVIST, K., NYSTRÖM, M., ANDERSSON, R., DEWHURST, R., JARODZKA, H. AND VAN DE WEIJER, J. 2011. *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

ISOKOSKI, P. 2000. Text input methods for eye trackers using off-screen targets. *Proceedings of the 2000 symposium on Eye tracking research & applications*, ACM, 15–21.

KANGAS, J., AKKIL, D., RANTALA, J. ISOKOSKI, P., MAJARANTA, P. AND RAISAMO, R. 2014. Gaze Gestures and Haptic Feedback in Mobile Devices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 435–438.

MARQUARDT, N., BALLENDAT, T., BORING, S., GREENBERG, S. AND HINCKLEY, K. 2012. Gradual engagement: facilitating information exchange between digital devices as a function of proximity. *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, ACM, 31–40.

MØLLENBACH, E., HANSEN, J.P. AND LILLHOLM, M. 2013. Eye movements in gaze interaction. *Journal of Eye Movement Research* 6, 2, 1–1.

RAWASSIZADEH, R., PRICE, B.A. AND PETRE, M. 2014. Wearables: has the age of smartwatches finally arrived? *Communications of the ACM* 58, 1, 45–47.

ROZADO, D., MORENO, T., SAN AGUSTIN, J., RODRIGUEZ, F.B. AND VARONA, P. 2015. Controlling a Smartphone Using Gaze Gestures As the Input Mechanism. *Hum.-Comput. Interact.* 30, 1, 34–63.

SHELL, J.S., VERTEGAAL, R. AND SKABURSKIS, A.W. 2003. EyePliances: Attention-seeking Devices That Respond to Visual Attention. *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, ACM, 770–771.