The use of Eye Gaze Gesture Interaction Artificial Intelligence Techniques for PIN Entry

Onyemauche U.C$^1$; Osundu, U.$^2$; Etumnu, R. C.$^2$; Nwosu, Q.N.$^3$

$^1$Department of Computer Science, Federal University of Technology Owerri, Imo State, Nigeria.
$^2$Department of Health Information Management, Imo State College of Health Technology and Management Sciences, Amaigbo. Imo State Nigeria.
$^3$Department of Physical & Health Education, University of Nigeria Nsukka, Enugu State, Nigeria.

Received: 01 Oct 2020; Received in revised form: 22 Oct 2020; Accepted: 01 Nov 2020; Available online: 10 Nov 2020

©2020 The Author(s). Published by Infogain Publication. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Abstract— With an increasing number of computer devices around us, and the increasing time we spend for interacting with such devices, we are vehemently motivated in finding new interaction methods which ease the use of computers or increase interaction efficiency. Eye tracking seems to be a promising technology to achieve this goal. This paper researches interaction methods based on eye gaze tracking technology with emphasis in PIN entry. Personal identification numbers (PINs) are one of the most common ways of electronic authentication these days and they are used in a wide variety of applications. The PIN-entry user study used three different gaze-based techniques for PIN entry. The first and second method used gaze pointing to enter the PIN on a number pad displayed on the screen. The first method used a dwell time of 800 milliseconds and the second method used a button, which had to be pressed when looking at the correct number on the number pad display. The second method was introduced as hardware key or gaze key, but called look & shoot method in the context of the user study as this name is self-explaining and got high acceptance by the participants. The third method used gaze gestures to enter the digits. The use of gaze gestures protects accidental input of wrong digits.

Keywords— Eye Tracker, Gaze Gesture, Human Computer Interaction, Personal Identification Number, Eye Movement.

1. INTRODUCTION

Gestures are a well-known concept for computer-human interaction. The idea behind gestures is the fact that we are used to employ movements of the body, mostly with the hands and the head, to communicate or to support communication. While such intuitive gestures are vague and culture dependent we are also able to perform well-defined and elaborated gestures. One example is gazing PIN entry tracking; other examples are the sign language for deaf mutes or the semaphore alphabet Wobbrock et al.[11]. Gestures for computer interaction are not very intuitive, as it requires learning a set of gestures and their semantics. For this reason the use of gestures for computer interaction has been seen as something for specialists or for very specific purposes. However, with the introduction of the iPhone and similar products, which have a touch sensitive surface as the only input modality, the use of gestures performed with the fingers became popular. In addition, the interaction with tabletop computers, which will appear on the mass market soon, will strongly rely on gesture input. An eye-gaze interface seems to be a promising candidate for a new interface technique, which may be more convenient than the ones we use. Traditionally, disabled people who cannot move anything except their eyes use eye gaze interaction. These systems are designed to direct the computer solely by the eyes. Such systems work well and are a great help for people who need them, but for others they are cumbersome and less efficient than keyboard and mouse. This contradicts the fact that looking is an easy task and that eye movements are fast.

Consequently, eye-gaze interfaces for the masses need a different design to bring benefit to the average user. An eye gaze interface may offer several potential benefits.
* Ease of Use

A benefit of eye tracking could be reduced stress for hand and arm muscles by transferring the computer input from the hand to the eyes. This need not necessarily put extra load on the eye muscles because for most interactions the eyes move anyway. For example, when clicking a button on the screen in most cases mouse and eyes move to the target.

*Interaction Speed up

Eye-tracking interfaces could speed up the interaction, as the eyes are quick. Although existing eye-typing systems are slower than traditional keyboard input, the combination of eye gaze with another input modality can provide fast interaction.

*Maintenance Free

Video-based eye tracking works contact free which means that no maintenance is necessary. There is no need to clean the device, which is a typical problem for keyboards and mouse devices. Placing the camera behind strong transparent material results in a vandalism-proofed interface, which is nearly impossible to realize for keyboards and mouse devices.

*Hygienic Interface

In environments with high hygienic demands, like an operation room for surgery, an eye-gaze interface would be useful because it allows interacting without anything to touch. Also for public interfaces, especially in times of pandemic threats, hygienic interaction is desirable.

*Remote Control

Another benefit resulting from eye tracking is possible remote control. Zoom lenses and high-resolution cameras make eye gaze detection possible over some meters of distance. Even the low-cost eye tracker used in the user studies of this paper offers one meter distance which is longer than an arm length.

*Safer interaction

Eye tracking not only guarantees the presence of a person but also his or her attention. A mobile phone, which requires eye contact for an outgoing call, will not call somebody because of accidentally pressed buttons while in a pocket. Eye tracking can ensure certain behaviour of the users. For example, a system can demand that a warning text is read before allowing the user to continue with other functions.

*More Information on User Activity

The eyes tell a lot about what somebody is doing. Tracking the eyes provides useful information for context-aware systems. In the simplest form an eye tracker tells where the attention is, which already has a big potential for the implementation of context-awareness. Simple analysis of the eye tracker data can detect activities like reading. Analysis that is more sophisticated could reveal the physical or emotional condition of a user, her or his age, and degree of literacy.

Of course, there are also possible problems.

*Ability to Control

The eyes perform unconscious movements and this might disturb their use as computer input. It is not clear to which degree people are able to control the movement of their eyes. The ability to control the eyes consists of both suppressing unintended movements and performing intended eye movements. It seems that we are at least able to control where we look because this is required by our social protocols. However, it is not clear whether we can train the motor skills of the eye muscles to the same extent as we can train the fingers for playing the piano.

*Conflict of Input and Vision

The primary function of the eyes is to enable vision. Using the eyes for computer input might result in conflicts. The well-known conflict is the Midas Touch problem – for the eye-gaze interface it is difficult to decide whether our gaze is on an object just for inspection or for invoking an action. Misinterpretation by the gaze interface can trigger unwanted actions wherever we look. The situation is similar when triggering actions by eye movements, i.e. gestures. The eye-gaze interface has to separate natural eye movements from intentional gaze gestures. Distraction by moving or blinking objects might also cause conflicts. The question how blinking advertisements on a web page interfere with eye gaze interaction is still a topic of research.

*Fatigue of the Eye Muscle

From other input devices we know that extensive use of particular muscles or muscle groups can cause physical problems called RSI (repetitive strain injury). There are fears that this might happen to the eye muscles too. The concern is justified and should be taken seriously, but as the eyes move constantly, even while we sleep, it might not turn out to be a problem.
II. RELATED LITERATURE

Personal identification numbers (PINs) are one of the most common ways of electronic authentication these days and they are used in a wide variety of applications. PINs have to preserve a personal secret as revealing the PIN can result in a personal loss. In the case of drawing cash from an Automated Teller Machine (ATM), which mostly happens in public space, there is a high danger that somebody observes the entry of the PIN. To avoid such shoulder surfing a less observable PIN entry method is desirable. Assuming that observing and interpreting eye movements is much more difficult than observing the finger movements on a number pad leads to the idea to use eye gaze for a more secure PIN entry. With this motivation Kumar et al.[5] did a user study with the title “Reducing Shoulder-surfing by Using Gaze-based Password Entry” The study presented here used the same gaze-based input techniques, but additionally researched the use of gaze gestures as a further input technique. The study was done within the project thesis of Roman Weiss and published in De Luca, et al. [1]

2.1 Eye Gaze as Context Information

Eye gaze as context in the sense of attention always means attention to an object and finally can be seen as pointing. It is definitely pointing whenever the focus lies on high accuracy of eye-gaze direction. However, there is no sharp border. Fono and Vertegaal presented a system where the gaze controls the focus selection of windows within a graphical user interface Fono & Vertegaal [2]. The reason to mention their work in this section and not in the related work on eye gaze as pointing is that windows normally have sizes above the accuracy of eye trackers and accuracy is not the main issue. They demonstrated that there is a big speed advantage for activation of windows by gaze compared to manual activation. In Vertegaal, et al. [2], they introduced Media EyePliances, where a single remote control can control several media devices. The selection of the device to control happens by looking at it. For this they augmented the devices with a simple form of an eye tracker, called Eye Contact Sensor ECS. The ECS uses the corneal reflection method and is calibration-free. To achieve this, a set of infrared LEDs is mounted on-axis around an infrared camera. When the camera delivers a picture with a glint in the pupil centre it means that the onlooker is looking directly towards the camera. Therefore, the ECS does not deliver coordinates, but signals eye contact only.

In Vertegaal, et al. [9], they present an attentive cell phone, which can detect whether the user is in a face-to-face communication with somebody else and use this information to employ social rules for interruption. The attentive cell phone uses speech detection and an ECS worn by the user. More so, Shell [8], went a step further and use attention to open sociable windows of interaction. They integrated the ECS into glasses to have a new wearable input device, which they call ECGlasses. The ECGlasses are an interesting hardware for discussing attentive user interfaces.

Most activities of humans involve eye movements related to the activity. Land [6] described eye movements for daily life activities such as reading, typing, looking at pictures, drawing, driving, playing table tennis, and making tea. Retrieving context information means to go the other way round and conclude from eye movements to the activity. Using such an approach Iqbal and Bailey measured eye-gaze patterns to identify user tasks Iqbal, & Bailey [4]. Their aim was to develop an attention manager to mitigate disruptive effects in the user’s task sequence by identifying the mental workload. The research showed that each task – reading, searching, object manipulation, and mathematical reasoning – has a unique signature of eye movement. The approach is interesting but the general problem is that the analysis of the eye movements can tell what the user’s task was in the past but not predict what the user is going to do. The period providing the data needed for the analysis causes latency. It is also not clear how reliable the task identification works.

Foo [3] opined that for “Eye-tracking to model and adapt to user meta-cognition in intelligent learning environments”, It is questionable how well such a model will work but the title of the paper shows how far the hopes go on analysis of eye movements. There are better chances for reliable activity detection by focusing on special activities like reading which is easier to define. However, the reliability for reading detection presented by Foo [3] is still below 90%.

III. RESULTS AND DISCUSSION

The PIN-entry user study used three different gaze-based techniques for PIN entry. The first and second method used gaze pointing to enter the PIN on a number pad displayed on the screen (see Figure 1). The first method used a dwell time of 800 milliseconds and the second method used a button, which had to be pressed when looking at the correct number on the number pad display. The second method was introduced as hardware key or gaze key, but called look & shoot method in the context of the user study as this name is self-explaining and got high acceptance by the participants. The prototype displays an
of them had already used an eye tracker before, but not on a regular basis. Figure 3 shows the completion time and error rate for the three different entry methods. The evaluation of the data using analysis of variance (ANOVA) showed no significant advantage regarding execution times for the look & shoot or dwell time method. Using the look & shoot method a four digit PIN entry took the subjects 12 seconds in average whereas a PIN entered using dwell time took 13 seconds. The error probability also showed no significant difference. Using dwell time, 15 of the entered 63 PINs were faulty (23.8%), using look & shoot, 13 entered PINs contained errors (20.6%).

Completion times and error rates for the 3 different methods

Dwell Time  Look & Shoot  Eye Gesture

The results of the user study show that eye gaze interaction is a suitable method for PIN entry and partially confirm the results found by Kumar et al. [5] regarding the dwell time and look & shoot techniques. Entering PIN numbers with the gaze gesture method took much longer than using the ‘classic’ methods (an average of 54 seconds per PIN entry was measured) but was also much more robust against errors than the methods described above. Only six of the entered PINs using gestures were erroneous (9.5%). Using a binomial test shows a significant enhancement of the error rate (p < 0.008).

The gaze gesture method is less intuitive than the classic methods as some subjects initially had problems to produce recognizable gestures. Furthermore, the gesture alphabet was unknown to all participants. This explains the big difference in time for completing the PIN entry task. The participants spent much time for looking at the sheet with the gestures for the single digits. As already shown in the previous user studies, a stroke within a gaze gesture needs about 400 to 500 milliseconds (a little bit more than 100 milliseconds for the saccade and around 300 milliseconds for the fixation) and entering a digit with four strokes takes about 2 seconds. A four-digit PIN with one second break between the inputs of the digits will last around 10 seconds. Indeed, there were participants in the study who entered the PIN correctly within 14 seconds. It needs a further study to find out whether all users can achieve this time once they are trained for gaze gesture input.

Fig.1: The user study setting (left) and the layout of the number pad for PIN entry

The third method used gaze gestures to enter the digits. As there is no need for remote entry of PINs it is not an obstacle to use a gesture key. With a gesture key it is not necessary to separate natural eye movements from gestures and this allows more freedom in the design of a gesture alphabet and consequently allows a more intuitive alphabet.

The alphabet used (see Figure 2) is the one introduced by Wobbrook et al. [11]. Without a gesture key the digit 9 and 5 would not be distinguishable.

Fig.2: The Digit Gesture Used for the prototype

In the user study, 21 volunteers completed the three different PIN entry tasks and afterwards answered a questionnaire. Seven of the participants were female and all participants were aged between 22 and 37 years. Five ATM number pad of an overall size of 730 × 450 pixels. Each button has a size of 180 × 90 pixels which is about 5° by 3° visual angle and hence clearly above the typical eye tracker accuracy of ± 0.5°. To retain the security benefit there was no feedback except asterisks to indicate successful entry of a digit.
In addition to the absence of a calibration process, the big advantage of the gaze gesture method is its robustness against input errors. Due to the abandonment of feedback for enhanced security, each wrong gaze leads to an incorrect PIN entry when using the dwell time or look & shoot method. This leads to high error rates for these methods. When using the gestures, a wrong gaze leads most probably to an unrecognizable gesture and not to an entry of a wrong digit. For gaze gestures the errors occur one level below the digit entry, i.e. at the gesture recognition level.

The main reason why a gesture performed by a user is not recognized by the system is a lack of exactness in the hand-eye coordination. As a button has to be pressed and held while performing the gesture, often an additional stroke was detected directly before or after the proper gesture. The reaction time for the finger, typically 300 ms is long compared to the time of a saccade, typically 100 ms. These unintended upstrokes or tails could be filtered out by the algorithm and improve the recognition rate.

IV. SUMMARY OF THE RESULTS FOR GAZE GESTURE

The user studies showed that people are able to perform gaze gestures. Help"ing lines are not helpful and should be replaced by helping points which displays naturally provide with the four corners. Gaze gesture detection works over some distance and can serve as a remote control. Gaze gestures can easily be separated from natural eye movements if they are performed on a big scale. Performing gaze gestures on a big scale is easy on large displays where the corners provide helping points. The situation is different for small displays like those of mobile phones, especially when buttons are arranged around the display. As it does not feel natural to perform a big gaze gesture with a small handheld device, the way to separate the gaze gestures from natural eye movements is the use of more complex gestures, for example with six strokes. The other solution for small displays is the use of gaze gestures only within a context. If the gesture detection is only active when the device expects an answer, an accidentally performed gaze gesture will not disturb general interaction.

One stroke of a gaze gesture needs about 500 milliseconds. With at least four strokes for the gesture, it is obvious that gaze gestures are not well suited for text entry as they are slower than other gaze-based methods. The argumentation that such form of text entry can make more fun or not correspond to the observations made during the user studies.

V. CONCLUSIONS

Gaze gestures are an alternative and interesting approach for eye gaze interaction. The absence of a calibration procedure makes it possible to build eye-gaze interfaces for instant use by different people.

Gestures in general are not very intuitive and the user has to acquire skills i.e. learn an alphabet and its meaning, before she or he can use it. It does not look like it is very difficult to learn the motor skills to perform gaze gestures. The number of participants in the study was not big enough to prove that everybody is able to perform gaze gestures but it clearly shows that there are people who can do it instantly and with ease. The eyes have one pair of muscles to control the x-direction and another for the y-direction. These suit well for the horizontal, vertical, and diagonal eye movements needed for the gaze gestures. In contrast, gestures for the hands expect straight lines while the natural movements are curved lines. The reason why it took some people much effort to get a gesture done, is partly because of a bad explanation using the term ‘line’ and partly because of nervousness of the participants. To look at the corners of a rectangle (corners of the display) in a certain order is not very difficult.

The analysis of natural eye movements recorded from different tasks and people indicates that gestures are reliably separable from natural eye movements when using reasonable parameters for grid size and timeout. Consequently, there is no need for an additional input modality like a gesture key. The corners of the display provide a perfect orientation for the gaze to perform a gesture. The only critical situation is a small display with buttons on the sides and below or above the display.

A main problem of eye tracking is the low accuracy. As the gaze gestures are movements on a big scale this problem dissolves. The accuracy is an angular value and for hitting a target with the gaze the situation gets worse on bigger distances. This is not the case for the gaze gestures because they are performed on an angular basis. The maximum distance of gaze gesture detection is a question of camera resolution and zoom factor of the lens. For this reason gaze gestures can serve as a remote control.

One field for application of gaze gestures is in highly hygienic environments (nothing to be touched) where an interface for instant use by different people (no calibration) is needed. Gaze gestures provide more complex control than possible with an eye contact sensor or by other touch free techniques such as a capacity sensor or a photoelectric barrier.

The question whether gaze gestures could serve as a remote control for the TV set and become an input

ISSN: 2456-2319
https://dx.doi.org/10.22161/eec.561
technology for the masses and not only for special purposes is interesting and still open. The big advantage of remote control by gaze gestures is the absence of a control device – nothing to look for and no batteries to recharge.

To answer the open questions it would be nice to have a small and cheap prototype of an eye-gaze gesture recognizer. This would make it possible to study the gaze gestures outside the lab in real situations.

Finally, the gaze gesture algorithm has the ability to translate eye movements into character strings. This could be useful for recognizing the user’s activity. The raw movement data allow the calculation of mean fixation time, mean saccade length or saccades per second, but make it hard to derive other meaningful numeric values. The representation of eye movements in form of a character string allows the application of string pattern matching algorithms. However recognizing the user’s activity from eye movements is the topic of the next chapter.

REFERENCES

[1] M.A. De Luca, D.W. Hansen, D.J. MacKay, J.P. Hansen, and L.M. Nielsen, Eye Tracking off the Shelf. In Proceedings of the 2004 Symposium on Eye Tracking Research & Applications. ETRA ’04. ACM Press (2004), pp 58 – 58.

[2] D. Fono, C., and C. Vertegaal, C. Touch-Typing With a Stylus. In Proceedings of the INTERACT and CHI Conference on Human Factors in Computing Systems CHI ’05. ACM Press (2005), pp 80 – 87.

[3] T. Foo Eye Gaze Based Reading Detection. In Conference on Convergent Technologies for Asia-Pacific Region, TENCON 2013. Volume 2, (2013) Pp 825 – 828.

[4] D. Iqbal, and C. Bariley, Eye-Gaze Tracking System with Free Head. In Proceedings of the INTERACT and CHI ’04 Conference on Human Factors in Computing Systems CHI ’04. ACM Press (2004), pp 180 – 189.

[5] M.Kumar, A. Paepcke, and T. Winograd, EyePoint: Practical Pointing and Selection Using Gaze and Keyboard. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI ’07. ACM Press (2007), pp 421 – 430.

[6] M.F. Land, Eye movements and the control of actions in everyday life. In Prog Retinal & Eye Res (2006) 25: pp 296 – 324. 

[7] A. Schmidt, A. Ubiquitous Computing in Context, PhD Thesis. Lancaster University, 11/2002.

[8] J.S. Shell, ECSGlasses and EyePliances: Using Attention to Open Sociable Windows of Interaction. In Proceedings of the 2004 Symposium on Eye Tracking Research & Applications. ETRA ’04. ACM Press (2004), pp 93 – 100.

[9] R Vertegaal, C. Dickie, C. Sohn, and M. Flickner, Designing Attentive Cell Phone Using Wearable EyeContact Sensors. In CHI ’15 Extended Abstracts on Human Factors in Computing Systems. CHI ’15. ACM Press (2015), 646 – 647.

[10] R. Vertegaal, A. Manuji, C. Sohn, and D.Cheng, Media EyePliances: Using Eye Tracking For Remote Control Focus Selection of Appliances. In CHI ’17 Extended Abstracts on Human Factors in Computing Systems. CHI ’17. ACM Press (2017), pp 1861 – 1864.

[11] J. O. Wobbrock, J. Rubinstein, M. Sawyer, W. Duchowski, Longitudinal Evaluation of Discrete Consecutive Gaze Gestures for Text Entry. In Proceedings of the 2007 Symposium on Eye Tracking Research & Applications. ETRA ’08. ACM Press (2007), pp 11 – 18.