Research on eye-gaze-based input for visual password authentication

イェサヤ, トミー, パウルス

KYUSHU UNIVERSITY

Graduate School of Design

Department of Human Science

Human Science International Course – Doctoral Program

Research on eye-gaze-based input for visual password authentication

視覚パスワード認証のための視線に基づく入力に関する研究

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF HUMAN SCIENCE
AND THE COMMITTEE ON GRADUATE STUDIES
OF KYUSHU UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Principal Advisor:

Dr. Gerard B. Remijn

Yesaya Tommy Paulus | 3DS16020N

February 2020

Acknowledgment

Abstract

The aim of this dissertation is to investigate the physical and cognitive aspects of using low-cost eye-tracking devices for visual password authentication. By using eye tracking, users can select objects on a display by using their eye gaze. In total, seven experiments were performed, in which low-cost eye trackers were used.

The physical aspects of eye-gaze-based input of objects for password authentication concern the measurement of the maximum and minimum viewing distances, highest and lowest viewing angles, and the ideal viewing angle of the user. In four experiments, measurements were performed in conditions both with and without glasses, at different viewing angles, under different lighting conditions. The results showed that even low-cost eye trackers worked in a stable manner in registering users' eyes at various viewing distances and viewing angles under different conditions of illuminance and luminance (Experiments 1 and 2). However, the results showed that the use of glasses indeed influenced user registration into low-cost eye-tracker interfaces (Experiment 3). Nevertheless, even with glasses, when the eye-tracking device and the display were set at a certain angle adjusted to the viewing height of the user, good and fast calibration and authentication could be achieved (Experiment 4). Low-cost eye-tracking devices thus can be considered for implementation into interactive-based interface systems that require eye-gaze-based authentication, such as visual password systems, under various lighting conditions in public spaces (e.g., personal computer use or Automatic Teller Machines – ATMs) or semi-public spaces (e.g., vehicles).

The cognitive aspects investigated in this dissertation related to the users' abilities in selecting visual object sequences from a screen using eye-gaze-based input. In three experiments, the users needed to select a sequence of visual objects from the grid-based interface screen by using their eye gaze, to enter a single password. Three different types of visual objects were used: alphanumeric characters, patterns of dots, and visual icons. These objects are commonly used in recognition-based visual password systems. A variety of grid densities and formations were considered and made in 16 ways, in between 3×3 and 6×6 object keys (Experiment 5), to indicate the positions of the objects on the screen. The results showed that password authentication with eye-gaze-based input is best performed on horizontal grids with relatively few cells, in the alphanumeric format (Experiment 6). Furthermore, a dwell time of 500 ms per object was easiest to use for selecting a sequence of visual objects on a screen using eye-gaze-based input (Experiment 7). Generally, the results of the dissertation suggest that eye-gaze-based input can be a suitable option to support the different necessities of users in performing user-interaction tasks involving object selection from a display, e.g., password authentication tasks in public settings.

Research Output

The following Chapters of this dissertation have been presented in the following conferences/symposiums and published in academic peer-reviewed journals.

Chapter 3 was presented at the 2017 2nd International Conference on Automation, Cognitive Science, Optics, Micro Electro--Mechanical System, and Information Technology (ICACOMIT), Balai Kartini Convention Center, Jakarta, Indonesia.

Paulus, Y. T., Hiramatsu, C., Syn, Y. K. H., & Remijn, G. B. (2017). Measurement of viewing distances and angles for eye tracking under different lighting conditions. In *2017 2nd International Conference on Automation, Cognitive Science, Optics, Micro Electro--Mechanical System, and Information Technology* (ICACOMIT) (pp. 54–58). IEEE. https://doi.org/10.1109/ICACOMIT.2017.8253386.

Chapter 4 was presented at the 2017 2nd International Conference on Automation, Cognitive Science, Optics, Micro Electro--Mechanical System, and Information Technology (ICACOMIT), Balai Kartini Convention Center, Jakarta, Indonesia.

Paulus, Y. T., Remijn, G. B., Syn, Y. K. H., & Hiramatsu, C. (2017). The use of glasses during registration into a low- cost eye tracking device under different lighting conditions. In *2017 2nd International Conference on Automation, Cognitive Science, Optics, Micro Electro--Mechanical System, and Information Technology* (ICACOMIT) (pp. 59–64). IEEE. https://doi.org/10.1109/ICACOMIT.2017.8253387.

Chapter 5 was presented at the 2018 9th International Conference on Awareness Science and Technology (iCAST), Kyushu University, Fukuoka, Japan.

Paulus, Y. T., Herlina, Leni, K. Z., Hiramatsu, C., & Remijn, G. B. (2018). A

Preliminary Experiment on Grid Densities for Visual Password Formats. In *2018 9th International Conference on Awareness Science and Technology (iCAST)* (pp. 122–127). IEEE. https://doi.org/10.1109/ICAwST.2018.8517236.

Chapter 6 was published in the peer-reviewed journal "*Journal of Ergonomics*".

Paulus, Y. T., & Remijn, G. B. (2019). What Kind of Grid Formations and Password Formats are Useful for Password Authentication with Eye-Gaze-Based Input? *Journal of Ergonomics*, *09*(01), 1–11. https://doi.org/10.35248/2165-7556.19.9.249.

Chapter 7 was presented at the 15th Asia-Pacific Conference on Vision (APCV) 2019.

Paulus, Y.T., Remijn, G.B. (2019). Suitable dwell time for eye-gaze-based object selection with eye tracking. In the *APCV2019 15th Asia-Pacific Conference on Vision*, Ritsumeikan University, Osaka, Japan.

Chapter 7 is being prepared as an article to be submitted to a journal.

Paulus, Y.T., Remijn, G.B. (in preparation). Usability of various dwell times for eye-gaze-based object selection with eye tracking.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

In this dissertation, research is described in which the user's eye gaze is utilized as an input method for performing user-interaction tasks that involve object selection from a display, e.g., eye-gaze-based authentication. An eye-tracking device enables the user to select a sequence of visual objects, e.g., alphanumeric characters, patterns of dots, or visual icons, from various grid-based screen interfaces, by focusing the user's eye gaze on the object. These visual objects were intended for three visual password formats. Alphanumeric characters, similar to those used in the eye-typing task for English words (e.g., see Majaranta, MacKenzie, Aula, & Räihä, 2003), were used in an alphanumeric password format. Sequences consisting of patterns of dots were designated for a pattern password format, whereas visual icons were used in a picture password format. These types of visual objects are commonly used in recognition-based password systems (Biddle, Chiasson, & van Oorschot, 2012), for example to manually unlock smartphones.

Recognition-based visual password systems were often considered as easier to memorize (Renaud, 2005; Nelson & Vu, 2010), and that systems with a denser grid potentially allow more secure password formation (Thorpe & van Oorschot, 2004; Alam, 2016). Furthermore, it has been suggested that eye-gaze-based input could be suitable against password theft ("shoulder surfing"), especially in public spaces (Dunphy, Fitch, & Olivier, 2008). Eye-gaze-based input is also considered as an easy and natural means of human-computer interaction (Majaranta & Räihä, 2002), which only needs slight practice (Stampe & Reingold, 1995). Finally, an eye-tracking device allows the user to actively interact with systems in which the user can select an object

from a display only by looking (Duchowski, 2018). It has further been reported that eye-gaze-based input is regarded faster than other input devices, e.g., a mouse or stylus, if the target object on the display is large enough (Ware & Mikaelian, 1987; Sibert & Jacob, 2000).

In eye-tracking research, however, some challenges still need to be faced, specifically in optimizing eye-tracking interfaces to the physical and cognitive abilities of users when selecting a specific object (e.g., a character/word, a menu item, or a password object) from a visual display. For example, in order to register the user's eyes into screen interfaces while standing, a suitable viewing distance and angle need to be established for every user. Furthermore, calibration for the same users with and without glasses standing in front of a display under different lighting conditions is also necessary to test the use of eye-tracking devices in real-life, practical situations. In today's world, an increasing variety of (public or semi-public) devices with displays require a variety of passwords. Therefore, research is necessary with regard to the type of password formats, grid formations, and dwell time durations that are useful for selecting a sequence of visual objects as a password from a display using eye-gaze-based input.

1.1   The aim of this dissertation

The aim of this dissertation is to investigate a number of physical and cognitive aspects of using low-cost eye-tracking devices for visual password authentication. The user used his/her eye gaze to select an object from a visual display. Seven experiments

were performed, with low-cost eye trackers (Tobii EyeX[©], Eye Tribe[©], and Tobii Eye Tracker 4C[©]).

The physical aspects of eye-gaze-based input of objects for password authentication concern the measurement of the maximum and minimum viewing distances, highest and lowest viewing angles, and ideal viewing angle of the user. The measurements were performed in conditions both with and without glasses, at different viewing angles, under different lighting conditions. The first objective of this dissertation was to obtain viewing distances and viewing angles at which the participant's eyes could be registered under three different lighting conditions. I also investigated whether lighting conditions had an influence on the maximum and minimum viewing distances and viewing angles. The second objective of this dissertation was to investigate the effect of wearing glasses on the calibration process into a low-cost eye-tracking device, and to investigate the ideal viewing angle of participants viewing from different heights, with and without glasses at different display angles.

With regard to the cognitive aspects of visual password authentication, the user used his/her eye gaze to select a sequence of visual objects on a grid-based screen interface. The visual objects that were used consisted of alphanumeric characters, dots, and visual icons, from which a single visual password with a different number of objects needed to be selected. An interactive interface with multiple objects shown on the screen typically employs a grid to organize objects based on sequenced columns and rows. Therefore, the third objective of this dissertation was to investigate what grid densities potentially are suitable for authenticating a visual password with actual

eye tracking. Furthermore, the first step towards a safer password system is to employ eye tracking to investigate which types of password format and grid formation are suitable for password authentication using eye-gaze-based input. Finally, in order to avoid that the user potentially selects non-target objects among multiple objects on a screen unintentionally (Jacob, 1991; Velichkovsky, Rumyantsev, & Morozov, 2014), the fourth objective of this dissertation was to establish what dwell time duration is useful for selecting a variety of visual objects from a grid-based interface with eye-gaze-based input.

1.2   Contents of this dissertation

In Chapter 2, more research background of eye-gaze-based input is described, e.g., an overview of eye tracking and its research, the basics of eye-gaze-based input, an overview of the interaction tasks with eye-gaze-based input, and grid-based interfaces for eye tracking. In Chapter 3, based on two experiments, the viewing distances and angles of two low-cost eye trackers under different lighting conditions are described. The eye-tracking devices used were the Tobii EyeX$^{©}$ and the Eye Tribe$^{©}$. In Experiment 1, the boundaries of maximum and minimum viewing distances and the ranges of highest and lowest viewing angles of the Tobii EyeX$^{©}$ eye-tracking device were obtained. In Experiment 2, the boundaries of maximum and minimum viewing distances and the ranges of highest and lowest viewing angles were also obtained for the Eye Tribe$^{©}$ eye-tracking device. In both experiments, three different lighting conditions were used, in order to check whether lighting conditions had an

influence on users' eyes registration at the maximum and minimum viewing distances and viewing angles.

In Chapter 4, two experiments are described that investigated whether the use of glasses during registration to the Eye Tribe$^©$ eye-tracking device could influence the calibration quality and calibration time. In Experiment 3, the effect of wearing glasses on the calibration process into an eye tracker was investigated under three different lighting conditions. In Experiment 4, the ideal viewing angle of participants viewing from different heights (standing, sitting), with and without (replica) glasses at different display angles was investigated under two different lighting conditions. For this experiment, different display angles were used.

One preliminary experiment (Experiment 5) is described in Chapter 5, the aim of which was to investigate what grid densities potentially are suitable for a follow-up experiment with actual eye tracking. In Chapter 6, an experiment is described (Experiment 6) which aimed to investigate what type of password format and grid formation would be suitable for password authentication using eye-gaze-based input. In order to perform password authentication, the participant was asked to identify and select a sequence of visual objects (e.g., alphanumeric characters, a pattern of dots, or visual icons) on 16 grid formations (ranging from 3×3 to 6×6 cells; columns × rows) by using eye-gaze-based input. An experiment on various dwell time durations for eye-gaze-based object selection (Experiment 7) is described in Chapter 7. In Chapter 8, the general discussion and conclusions of this dissertation are described.

Chapters 3 and 4 were intended to prepare and check whether the eye tracking devices used here were suitable for performing user-interaction tasks. Chapters 5 to 7

concerned the main topic of this dissertation, describing password authentication on a grid-based screen interface using eye-gaze-based input. For this dissertation, the procedures in all experiments described in Chapters 3-7 were approved by the Ethical Committee of the Faculty of Design, Kyushu University, Japan (131-3).

Chapter 2. Research background of eye-gaze-based input

2.1   Eye Tracking

The eye is one of the main human organs that can be used to learn and understand things in our environment. Some authors also said that the eye is the window to the soul through the brain (Ellis et al., 1998; Brigham et al., 2001). The actions of humans are mainly based on their understanding of the information obtained through the eye. Behaviorally, the user can collect relevant information and/or neglect inessential information (Chun & Wolfe, 2005). Thus, it can be assumed that knowing the user's eye gaze points may give insight into what exactly drew his/her attention, and even likely give a hint on how he/she perceived the things in the visual field.

Eye tracking refers to the activity of recording and measuring eye movements to establish where the point of the user' gaze would be, what he/she is staring at, and how long his/her eye gaze is in a certain place in the visual field. Eye tracking has become increasingly popular in human-machine interaction, potentially creating communication between humans and machines (Majaranta & Bulling, 2014). In one of the earliest applications, Fitts and his colleagues recorded aircraft pilots' eye movements with a motion-picture camera and analyzed the eye movements in every single frame (Fitts, Jones, & Milton, 1950). They concluded that the eye movements were different between flight instruments (e.g., air-speed indicator, altimeter, etc.), based on their measurements of the fixation frequency and required fixation duration. They also discovered that the more experienced pilots made shorter fixations. Ever since, eye-tracking technologies have been used in a broad range of application areas.

Eye tracking has been used as a usability research tool for researchers who would like to study and analyze human performances or behaviors. It also has been used as an eye-gaze-based input medium that can be operated in real-time (Jacob & Karn, 2003).

2.2   Research on eye tracking

As a usability research tool, eye tracking has been used to study human attention direction. The user's eye movements can give information as regards the direction of autistic children's attention and eye tracking can assess autistic children's intentions and performances while they use a device system. For example, a review study regarding autism by Boraston and Blakemore (2007) showed that the attention of autistic children was more focused at the mouth region of someone's face than the eye region. Measuring infants' gaze thus may help early detection of the risk of autism (Navab, Gillespie-Lynch, Johnson, Sigman, & Hutman, 2012). Other studies on eye tracking in pediatrics showed that when infants were looking at target pictures on a display to learn new words, infants with late speech development have different representations of new words than their peers (Ellis, Borovsky, Elman, & Evans, 2015). Moreover, children's attention based on first fixations was less occupied on target faces for 3-month-olds as compared to 6-month-olds and adults (Di Giorgio, Turati, Altoè, & Simion, 2012). Other studies have used eye tracking for the recognition of emotion in human faces (Shechner et al., 2013), or to investigate the gaze time of customers on nutrition labels of food (Graham & Jeffery, 2012). Finally, eye tracking has been used to identify the gaze duration of users when performing password enrollment with manual input (Mihajlov & Jerman-Blazic, 2018). All the

above studies indeed demonstrated that eye tracking is highly useful for analyzing the user's behavior or habits based on his/her attention direction. It can even help to identify problems related to cognitive functioning, which can help the user in daily life.

One of the main functions of eye tracking is as an eye-gaze-based input medium to support users who perform interaction or communication with other devices (e.g., computers, smartphones, tablets, etc.), in real-time. At first, this technology was mainly intended for users with disability issues, who are only able to use their eyes in order to perform an action through a visual display. Eye tracking, for example, can help persons with Amyotrophic Lateral Sclerosis (ALS). As demonstrated by Hutchinson, White, Martin, Reichert, and Frey (1989), disabled users could operate system interfaces by gazing at proper menu options on the display without moving their heads. Following this, gaze-based interaction for selecting static objects, smooth-moving objects and menu options was developed for any user as a natural and simple means to control the system interface (Jacob, 1991). For example, a webpage system that incorporated eye tracking was developed as a language translator to users after reading pages for long fixations (Hyrskykari, Majaranta, Aaltonen, & Räihä, 2000). In recent years, eye gaze has been employed as an interactive interface for typing sentences (Majaranta, Ahola, & Špakov, 2009), communicating with a virtual character (Bee et al., 2010), and authenticating a visual password (Dunphy et al., 2008; De Luca, Denzel, & Hussmann, 2009; Forget, Chiasson, & Biddle, 2010). Many applications that incorporate eye-gaze-based input nowadays may support the activities of all users, disabled or not, in daily life. Taken

together, eye tracking is not only a highly robust tool for usability research but also is a very promising technology that facilitates interaction or communication between users and devices.

2.3   Basics of eye-gaze-based input

2.3.1   Characteristics of eye movements

In general, eye movements are not stable over the visual field. To see an object accurately, the user needs to scan every object in the visual field with rapid eye movements, so-called saccades. It means that the eye gaze jumps quickly from an object to another object, and it typically lasts approximately 30-120 ms (Jacob, 1995). Because it is so rapid, as soon as a saccade is started, the eye movements cannot be interrupted. However, its track orientation can be changed. Naturally, a saccade is followed by fixation: a time duration that the user needs to hold his/her gaze steadily on an object. It should be relatively long, so that the user's brain can understand the features of the object. Typically, a fixation holds on for 200 to 600 ms after each saccade occurs. Therefore, regular eye movements consist of fixations on objects connected by rapid saccades between those objects.

A user can look at a relatively narrow area of the visual field with high acuity during fixations. It is because the fovea in the retina of the eye gives accurate vision with one arc-min (1/6 of a degree) of visual angle. Thus, the user is not able to see an object in detail outside the fovea (the peripheral vision), but inside the fovea area (the narrow vision) the user can see an object accurately. The narrow vision induces the need to move the eyes rapidly around the visual field, while the peripheral vision gives

hints on where the next object can be seen in the visual field. Because of this, the direction of the user's eye gaze can be easily traced by using an eye tracking device (for more details on eye movements and visual perception, e.g., see Haber & Hershenson, 1973).

2.3.2 Calibration of an eye-tracking device

For every user, the eye-tracking device must be calibrated in order to specify the gaze point of a user's eye precisely at every spot where objects exist on the screen. Simply put, calibration is the process of tracking a user's eye movements for an accurate gaze point calculation. The eye-tracking software normally presents 9 points that are equally spaced on the screen, and the user is required to gaze at each point, one by one. During this process, several images of a user's eyes can be collected and then analyzed. As an output, each image is mapped and converted into eye-gaze points (as x, y coordinates) on the screen (for more information about this see Majaranta & Bulling, 2014). By doing a successful calibration, the quality of eye tracking can become rather accurate - it is around 0.5 deg in visual angle, to see the area of about 15 pixels on a 17-in display (a screen resolution of 1024×768 pixels) from a viewing distance of 70 cm. In practice, however, the registered eye-gaze point often drifts far off from the real eye-gaze point. This occurs likely as a result of head movements, changes in pupil size and in lighting, or the use of glasses. This can be solved by recalibrating the registered eye-gaze points using eye-tracking software. For example, Tobii eye-tracking systems utilize data of both eyes to prevent and minimize the drifting effects (for details, see Tobii, 2006) – this supports a user to avoid continuous recalibration. Furthermore, the system even still works with data from one eye in case

another eye is off track. It is necessary to note that the eye gaze registration and calibration may not be done if the user is sitting or standing too close to or too far from eye-tracking devices. Therefore, a suitable viewing distance and angle need to be established for every user.

It is known that the use of glasses can reduce the accuracy and speed of user registration and calibration into eye-tracking devices (Nyström, Andersson, Holmqvist, & van de Weijer, 2013; Stawicki, Gembler, Rezeika, & Volosyak, 2017). This has been investigated (Funke et al., 2016) by comparing two low-cost eye trackers (Tobii EyeX[©] and Eye Tribe[©]) with a medium-cost (Smart Eye Aurora[©]) and two high-cost eye trackers (Seeing Machines faceLAB[©] and Smart Eye Pro[©]). The performance evaluation of the devices showed that the percentage of viewers that could be calibrated with the Tobii EyeX[©] was just 50% when the users wore glasses. However, it is unclear whether the calibrations in this study were performed between users with prescription glasses and a group with normal eyesight, or among the same users with and without glasses. Comparison between two different user groups may not have provided valid information; it is conceivable that a group with prescription glasses has more difficulty performing calibration (tracking) tasks because their eyesight is worse than that of users who do not need glasses (Nyström et al., 2013). Moreover, in this study the calibration was obtained with users sitting in front of a display under a single room lighting condition (Funke et al., 2016). Hence, calibration for the same users with and without glasses standing in front of a display under different lighting conditions is also necessary to test the use of eye-tracking devices in real-life, practical situations.

2.3.3 Dwell-time-based object selection

When using eye-tracking devices, the user can select an object from a display by focusing his/her eye gaze on the object, such as a letter, a menu item or a password object. Before selecting the object, the user is required to dwell his/her gaze on the object for a certain amount of time, a so-called "dwell time", in order to trigger an action. For systems that only use eye-gaze-based input, dwell time is the most common and easy way to select a static object from a display (Sibert & Jacob, 2000). The use of dwell time makes other actions obsolete. Using dwell time, however, requires some practice by the user. On a display with multiple objects, the user must first reliably identify the target object, before performing an action on it (Land and Furneaux, 1997). Accordingly, when eye gaze is used to identify a target object, the user may unintentionally and inattentively dwell his/her gaze on the wrong object. As a result, this object may even become selected as the target object – a problem that is known as the Midas-Touch problem (Jacob, 1991; Velichkovsky et al., 2014). In order to counteract the Midas-Touch problem, developers of eye-gaze-based object selection interfaces typically use a fixed duration of dwell time.

Previous studies have shown that various dwell time durations have been used that ranged from 300 to 1100 ms. In eye-typing systems, for example, eye-gaze-based input is used to select characters on an on-screen keyboard with a fixed duration of dwell time (e.g., see, Miniotas, Spakov, & Evreinov, 2003; Hansen, Johansen, Hansen, Itoh, & Mashino, 2003; Majaranta et al., 2003; Bee & André, 2008; Kurauchi, Feng, Joshi, Morimoto, & Betke, 2016; Spakov & Miniotas, 2004). Furthermore, in visual password systems, users are asked to select a sequence of characters and faces as a

password on a screen-lock interface by maintaining their eye gaze for a fixed dwell time (e.g., see, Maeder, Fookes, & Sridharan, 2005; Kumar, Garfinkel, Boneh, & Winograd, 2007; De Luca, Weiss, & Drewes, 2007; Dunphy et al., 2008). Other object selection techniques than dwell time have been described in detail by Majaranta (2009). To my knowledge, however, in the above-mentioned visual password systems dwell time was fixed at one single value without concern for user preferences. Besides that, comparative research on object dwell times has not yet been performed systematically.

2.4   Interactive interfaces with eye-gaze-based input

By using eye-gaze-based input, a user can perform interactions with interface systems, such as typing sentences. In the "Symbol-Creator" system, for example, a user was able to type sentences by gazing at a symbol key on an on-screen keyboard (Miniotas et al., 2003). In the Japanese version of the "GazeTalk" system, a user could select a Kana character by his/her gaze to type sentences (Hansen et al., 2003). In the same manner, eye-gaze-based input has also been utilized by some studies for typing English sentences (Majaranta et al., 2003; Bee & André, 2008; Kurauchi et al., 2016). Other than typing sentences, authenticating passwords with eye-gaze-based input have been carried out as well (Dunphy et al., 2008; De Luca et al., 2009; Forget et al., 2010). Since password authentication with eye-gaze-based input is the topic of this dissertation, I will explain past research in more detail below.

Typically, the user authenticates him-/herself into a system by something he/she is (biometrics) or by something he/she knows or recognizes (cognometrics).

Cognometrics concern human cognitive abilities, such as the ability to recognize a sequence of scenes, faces or visual objects. In the case of biometrics, the system captures digital information of a user's physiology at the enrollment phase, which is then verified at the authentication phase. Still, biometric data, such as fingerprints or pupil recognition, might pose great dangers to the user – thieves may opt for heavy bodily harm. In the case of cognometrics, the user and the system share an agreed secret at the enrollment phase, and the system then determines whether the user being authenticated has the pre-agreed secret. If the user proves knowledge of the secret – known as a "password" – the system will authenticate him/her (Bishop, 2005).

Over the past few years, it has been put forward that visual password systems mediated by manual input could be an alternative to text-based password systems (Biddle et al., 2012). These visual password systems typically can be formed by drawing a figure on a grid, by indicating marker points on an image, or by selecting a sequence of symbols, patterns, or pictures from a display (De Angeli, Coventry, Johnson, & Renaud, 2005). The last form is often named visual-based recognition, which can facilitate retention. Visual passwords based on recognition are considered more straightforward to memorize since humans have a vast memory for searching visual information (Renaud, 2005; Nelson & Vu, 2010). Furthermore, visual password systems that incorporated eye-gaze-based input have also been explored.

Visual password authentication with eye-gaze-based input, or a combination of manual input and eye-gaze-based input, can be done in various ways depending on the type of system. For example, a password can be made by drawing strokes using eye tracking and a keyboard sequentially in a system named "EyePassShapes" (De Luca

et al., 2009). Next, the user can press the spacebar on a keyboard for a few seconds to record his/her gaze after selecting points on a sequence of images in the "Cued Gaze-Points" system (Forget et al., 2010). As compared with manual input only, the combination of a keyboard and eye-gaze-based input is potentially safer against "shoulder-surfing" in public spaces, i.e., password stealing by a third party who observes from a certain distance and then copies a user's manual input of digits or text. The use of eye-gaze-based input only (with eye tracking) has been tested in the recognition based password system "PassFaces" (Dunphy et al., 2008). The results suggest that this input would be a suitable and safe option for authentication, amongst others, on Automated Teller Machines (ATMs). However, this system only used a single password format. Furthermore, the above password systems have been tested only under a single room lighting condition for users sitting in front of a personal computer. So far, to my knowledge, comparative research about visual password formats has not yet been performed in a systematic way.

2.5   Grid-based interface

One of the easiest ways to organize objects on the screen interface is to apply a grid as a layout. A grid can hint to object position and identification which is also an important factor in password systems (Jermyn, Mayer, Monrose, Reiter, & Rubin, 1999; Tao & Adams, 2008). The grid aids the aligning of objects based on sequenced columns and rows. Therefore, the user may easily find the target objects on a display when identifying familiar object positions. At first, the grid was used to set handwriting on paper and then on printed pages in publishing. Given that printed and

electronic pages have much in common, it should not come as a surprise that a grid is also used in designs of the interface of webpages and applications. For example, a grid is typically used for screen-lock interfaces of personal computers or smartphones.

In visual password systems, besides the common 3×3 (+1) grid for digits, grids with higher densities (more object keys) and different formations have been considered and tested. For example, 3×4 and 4×4 grids (columns×rows) were used for the recognition-based system "Visual Identification Protocol" (De Angeli et al., 2005). A 4×4 grid has been used for "ImagePass" (Mihajlov, Trpkova, & Arsenovski, 2013), and a 5×5 grid for "Draw a Secret (DAS)" (Jermyn et al., 1999) and "Déjà Vu" (Dhamija & Perrig, 2000). Four equal numbered grid densities of 5×5, 6×6, 7×7, and even 10×10 cells have been tested with DAS (Thorpe & van Oorschot, 2004) and "Signature-based User Identification System (SUIS)" (Alam, 2016). All these systems used manual input.

The reason why the above-mentioned studies have explored the use of higher grid densities for password authentication is that a higher number of object keys (columns × rows) enables safer passwords. If a user has more object keys to choose from, he/she can form more complex passwords. In the systems "DAS" and "SUIS", increasing the grid density increased the password space (Thorpe & van Oorschot, 2004; Alam, 2016), which is an indicator of security strength as specified by the total number of possible passwords (2^n, where n is the number of grid cells). Furthermore, in the case a user prefers relatively short passwords, a higher grid density lowers the chance that the correct sequence of object keys can be copied or discovered by third parties, for example, through shoulder-surfing. Research on the relation between grid

density and password complexity, however, has shown mixed results. Research with a system that used manual input has shown that the use of grid densities of more than 4×4 cells had minimal influence overall on the complexity of passwords (Aviv, Budzitowski, & Kuber, 2015). Moreover, no systematic, comparative research about grid formations has been performed.

Chapter 3. The measurement of viewing distances and viewing angles for eye tracking under different lighting conditions

3.1 General Purpose

In order to investigate whether eye trackers can be used for visual password registration in public spaces, such as in settings with personal computers or ATMs, it is first necessary to establish whether low-cost eye-trackers perform efficiently under various lighting conditions. The following two experiments were conducted to measure the limits of viewing distances and viewing angles of low-cost eye trackers under different conditions of luminance and illuminance. The participant was asked to stand in front of a display at various distances and under various viewing angles, and he/she was instructed to register his/her eyes into the eye-tracker system. The eye-tracking devices used in these two experiments were among the most cost-effective devices, that is, the Tobii EyeX$^{©}$ and the Eye Tribe$^{©}$.

3.2 Experiment 1. Measurements with the Tobii EyeX$^{©}$

3.2.1 Purpose

The first purpose of Experiment 1 was to obtain the maximum and minimum viewing distance at which the participant's eyes could be registered. The second purpose was to obtain the highest and lowest viewing angle at which the eye tracker worked at a viewing distance of 40 cm. Both viewing distance and viewing angle data were obtained under three different lighting conditions since illuminance (in a room) and luminance (on a display) vary 24 hours a day, especially if the weather changes.

The third purpose was to investigate whether lighting conditions had an influence on the maximum and minimum viewing distances and viewing angles. The eye-tracking device used in this experiment was the Tobii EyeX$^©$.

3.2.2 Method

Participants

Twenty-five participants (14 males and 11 females) were invited to participate in this experiment. Their ages ranged from 20 to 61 years ($M = 27.7$ years, $SD \pm 9.3$ years). Eight participants wore glasses, 4 participants wore contact lenses, and 13 participants had uncorrected vision. The height of participants was between 151 and 205 cm ($M = 169.5$ cm, $SD \pm 10.8$ cm). Twenty participants were Asian (Japanese, Chinese, Indonesian), and 5 participants were Caucasian. The participants participated on a voluntary basis and provided written, informed consent as to their participation after the purpose and procedure of the experiment were explained to them (Appendix A).

Apparatus

A touch-screen display (Microsoft Surface 4, 12-in) was used and equipped with Tobii EyeX$^©$ software. The eye-tracking device was placed just below the display, at the height of 133 cm from the ground. The eye tracker and the display were fixed at a viewing angle of 90 degrees. Viewing distances and angles were measured in a room under three lighting conditions: 'natural light' from a window, 'room light', and 'full light'. The 'full light' condition combined the 'natural light', the 'room light', and an additional light spot on the display to which the eye tracker was connected (Appendix

B). For these three conditions, illuminance was measured at the point where the participant was viewing the display by a lux meter (i1 Pro XRITE). Next to illuminance, the display's luminance was measured under the three lighting conditions as well (TOPCON Luminance Meter BM-9). The display luminance was first measured for a dark display, and then for a bright display (with white paper placed on the display to get brightness levels from a white surface when the display was off), at seven different points in time during a day in between 9:00 a.m. and 19:00 p.m. The illuminance in the natural, room, and full light conditions was 6.32 ± 5.18 lux, 203.32 ± 9.57 lux, and 210.71 ± 8.63 lux on average, respectively. The display luminance in the natural, room and full light conditions for the dark display was 0.03 ± 0.02 cd/m^2, 1.08 ± 0.31 cd/m^2, and 6.72 ± 1.35 cd/m^2 on average, respectively. For the bright display it was 0.15 ± 0.09 cd/m^2, 3.95 ± 0.30 cd/m^2, and 28.11 ± 2.06 cd/m^2 on average, respectively.

Procedure

After receiving informed consent, each participant was asked to stand in front of the display and instructed to focus on a fixation point (a yellow star) in the geometrical middle of the display. From a viewing distance of 40 cm, the participant was registered into the system, and his/her face photo was taken with his/her permission (Appendix C). The viewing distance was measured from the tip of the nose of the participant to the middle of the three near-infrared lights of the eye tracker (Figure 3.1).

Following this, the participant was asked to slowly walk forward or backward while focusing on the fixation point in order to measure the minimum and maximum

viewing distance at which his/her eyes were registered (Figure. 3.2a). The Tobii EyeX© software indicates successful registration of the viewer's eyes by means of two white dots on the display. Registration was considered unsuccessful when the dots disappeared from the screen, i.e., when the viewer was standing too close to or too far away from the display.

Figure 3.1. Procedure of Experiment 1 and Experiment 2. Measurements of viewing distances and angles, as obtained from the tip of the participant's nose to the middle of the three near-infra-red lights of the eye tracker used in Experiment 1, or to the midpoint in between the two near-infra-red lights of the eye tracker used in Experiment 2.

The lowest and highest viewing angle possible were also measured at which the viewer's eyes were registered at a viewing distance of 40 cm (Figure. 3.2b). The participant was asked to move upward (by stretching his/her legs) until he/she reached the lowest viewing angle at which the eyes were still registered. Also for this measurement, the participant had to keep looking at the fixation point at the middle of the display at the viewing distance of 40 cm. The angle was measured from the tip of the participant's nose to the near-infrared light in the middle of the eye tracker. The

same procedure was used to measure the highest viewing angle possible at a viewing distance of 40 cm. All four parameters (minimum viewing distance, maximum viewing distance, lowest viewing angle, and highest viewing angle) were assessed under the three lighting conditions. This took about 20 minutes for each participant. The procedure for Experiment 1 was approved by the Ethical Committee of the Faculty of Design, Kyushu University, Japan (131-3).
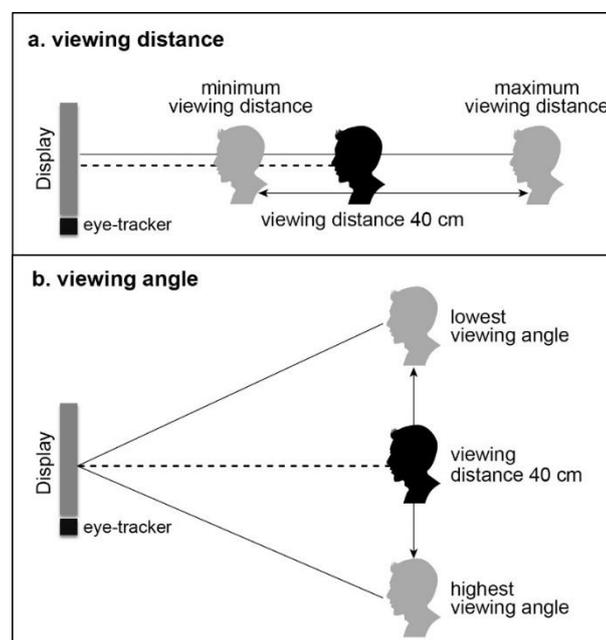
Figure 3.2. Illustration of measurements performed in Experiment 1 and 2 (a. viewing distance, and b. viewing angle).

3.2.3   Results of Experiment 1

The results generally show that the Tobii EyeX$^©$ device could register the participant's eyes at relatively the same viewing distances and angles under three lighting conditions. Figure 3.3 shows the maximum and minimum viewing distance at which the participant's eyes (n = 25) could be registered under the three lighting

conditions. The average maximum viewing distance (black bars) in the natural light condition was 71.2 ± 4.6 cm. In the room light condition, it was 70.3 ± 5.5 cm, and in the full light condition, it was 69.4 ± 4.9 cm. The average minimum viewing distance (white bars) at which the participant's eyes could be registered was 36.2 ± 1.7 cm in the natural light condition, 36.2 ± 1.6 cm in the room light condition, and 36.3 ± 1.3 cm in the full light condition.
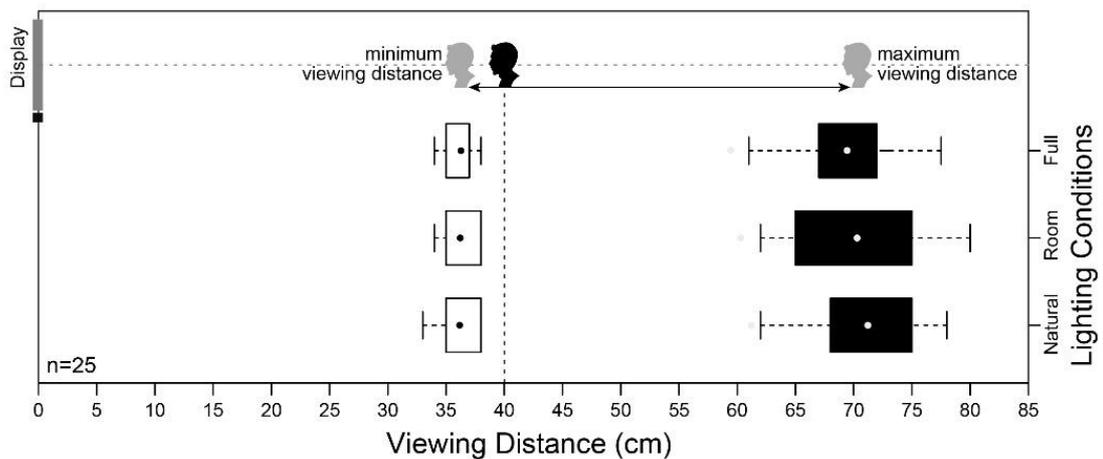
Figure 3.3. Results of Experiment 1. The maximum and minimum viewing distances for which the participant's eyes could be registered under the three lighting conditions. Error bars indicate ± 1 *SD*.

Shapiro-Wilk tests for the data obtained at the minimum viewing distances showed that they were not normally distributed (*df* = 25, natural lighting: $W = 0.87$, *p* = 0.004; room lighting: $W = 0.86$, *p* = 0.002; full lighting: $W = 0.90$, *p* = 0.019). Statistical analyses were therefore performed using non-parametric Friedman tests. The minimum viewing distance data (*df* = 2, n = 25) showed no significant difference between lighting conditions ($\chi^2 = 0.78$, *p* = 0.679). A significant difference between

the maximum viewing distances under the three lighting conditions was also not found

($\chi^2 = 2.36$, $p = 0.307$).

Figure 3.4 shows the lowest and highest viewing angle at which the participant's eyes (n = 25) could be registered under the three lighting conditions. The average lowest viewing angle (white bars) in the natural light condition was 71.3 ± 2.6 degrees. In the room light condition, it was 71.4 ± 2.8 degrees, and in the full light condition, it was 71.4 ± 3.5 degrees. The average highest viewing angle (black bars) at which the participant's eyes could be registered was 100.7 ± 1.2 degrees in the natural light condition, 100.9 ± 1.3 degrees in the room light condition, and 101.3 ± 1.0 degrees in the full light condition.
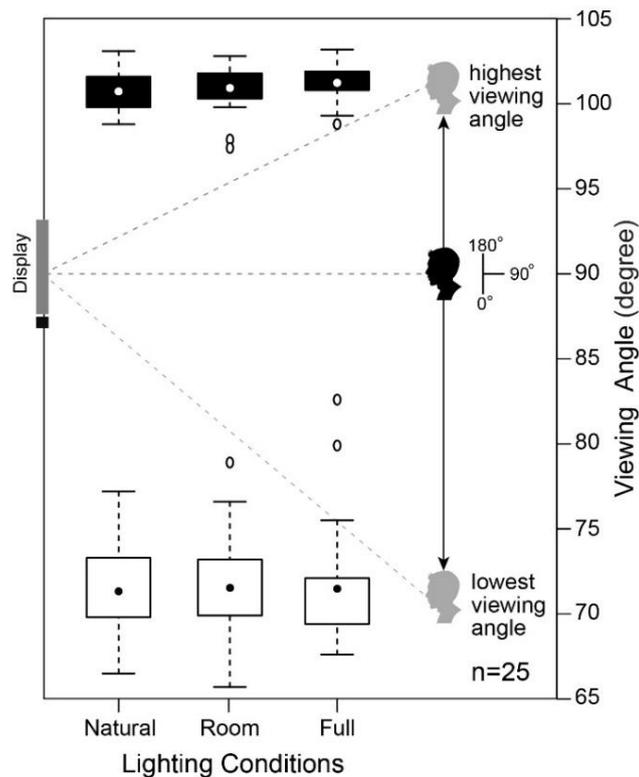
Figure 3.4. Results of Experiment 1. The lowest and highest viewing angles for which the participant's eyes could be registered under the three lighting conditions. Error bars indicate ± 1 *SD*.

Shapiro-Wilk tests for the highest viewing-angle data showed that the data obtained under room lighting were not normally distributed ($df = 25$, $W = 0.90$, $p = 0.020$). Similarly, the lowest viewing-angle data obtained under full lighting were also not normally distributed ($df = 25$, $W = 0.81$, $p < 0.001$). Statistical analyses performed using Friedman tests ($df = 2$, n $= 25$) showed no significant difference between lighting conditions for the highest viewing-angle data ($\chi^2 = 3.08$, $p = 0.214$) and for the lowest viewing-angle data ($\chi^2 = 0.06$, $p = 0.970$). Details about the statistical analysis of the data obtained in Experiment 1 are in Appendix D.

3.3   Experiment 2. Measurements with the Eye Tribe[©]

3.3.1   Purpose

In Experiment 2, the goal was to obtain the same parameters (the maximum and minimum viewing distance at which the user's eyes were still registered, and the highest and lowest viewing angle) under three different lighting conditions using a different low-cost eye-tracking device, from Eye Tribe[©].

3.3.2   Method

Participants

Twenty-eight participants (18 males and 10 females) were invited to participate in this experiment. Their ages ranged from 22 to 42 years ($M = 25.6$ years, $SD \pm 4.6$ years). Ten participants wore glasses, 5 participants wore contact lenses, and 13 participants had uncorrected vision. The height of the participants was between 153 and 183 cm ($M = 166.8$ cm, $SD \pm 7.8$ cm). All participants were Asian (Chinese,

Indonesian, Malaysian and Japanese). The participants participated on a voluntary basis and provided written, informed consent as to their participation after the purpose and procedure of the experiment was explained to them (Appendix A).

Apparatus

The Eye Tribe© software was installed on a touch-screen display (Microsoft Surface 4, 12-in). Viewing distances and angles were measured in a room under three lighting conditions with the same settings as used during Experiment 1 (Appendix E). Illuminance and luminance were measured in the same way and with the same equipment as in Experiment 1. The average of the illuminance in the natural, room and full light conditions was 6.68 ± 5.63 lux, 202.82 ± 10.19 lux, and 211.86 ± 10.11 lux, respectively. The average of the display luminance in the natural, room and full light conditions for the dark display was 0.03 ± 0.02 cd/m^2, 0.97 ± 0.32 cd/m^2, and 6.53 ± 1.44 cd/m^2, respectively, and for the bright display it was 0.14 ± 0.09 cd/m^2, 3.88 ± 0.29 cd/m^2, and 27.46 ± 1.80 cd/m^2, respectively.

Procedure

All procedures in Experiment 2 were the same as in Experiment 1, except for the procedure to measure the calibration quality. In the default procedure used in the Eye Tribe© software, the participant was asked to follow the movement of a circle by using his/her eye gaze in order to measure the calibration quality at which his/her eyes were registered. The Eye Tribe© software indicates acceptable results when the result is one of the following: 'Perfect', 'Good' or 'Moderate'. Calibration was not acceptable when the result was 'Poor' or 'Redo'. The calibration measurement was repeated when

the result was not acceptable, with a maximum of three repetitions for each participant. The limit of three repetitions was set in order to avoid eye fatigue. The Eye Tribe[©] has two near-infra-red lights. The procedure for Experiment 2 was approved by the Ethical Committee of the Faculty of Design, Kyushu University, Japan (131-3).

3.3.3   Results of Experiment 2

The results show that, for the maximum viewing distance, the Eye Tribe[©] eye-tracking device could register the participant's eyes at the larger distance under the full lighting condition than under the room lighting condition. Also, for the lowest viewing angle, participants' eyes could be registered at a lower angle under the room lighting condition than under the full lighting condition. Figure 3.5 shows the maximum and minimum viewing distance at which the participant's eyes (n = 28) could be registered under the three lighting conditions. The average maximum viewing distance (black bars) in the natural light condition was 77.9 ± 5.1 cm. In the room light condition, it was 76.9 ± 5.1 cm, and in the full light condition, it was 77.9 ± 4.8 cm. The average minimum viewing distance (white bars) at which the participant's eyes could be registered was 28.9 ± 1.6 cm in the natural light condition, 29.0 ± 1.7 cm in the room light condition, and 29.3 ± 1.7 cm in the full light condition.

Shapiro-Wilk tests for the data obtained at the minimum viewing distances showed that data obtained in the full light condition were not normally distributed (df = 28, W = 0.92, p = 0.026). Statistical analyses were therefore performed using non-parametric Friedman tests. The minimum viewing distance data (df = 2, n = 28) showed no significant difference between lighting conditions (χ^2 = 2.58, p = 0.275). The

Friedman test for the maximum viewing distances under the three lighting conditions, however, showed a significant difference ($\chi^2 = 9.41$, $p = 0.009$).
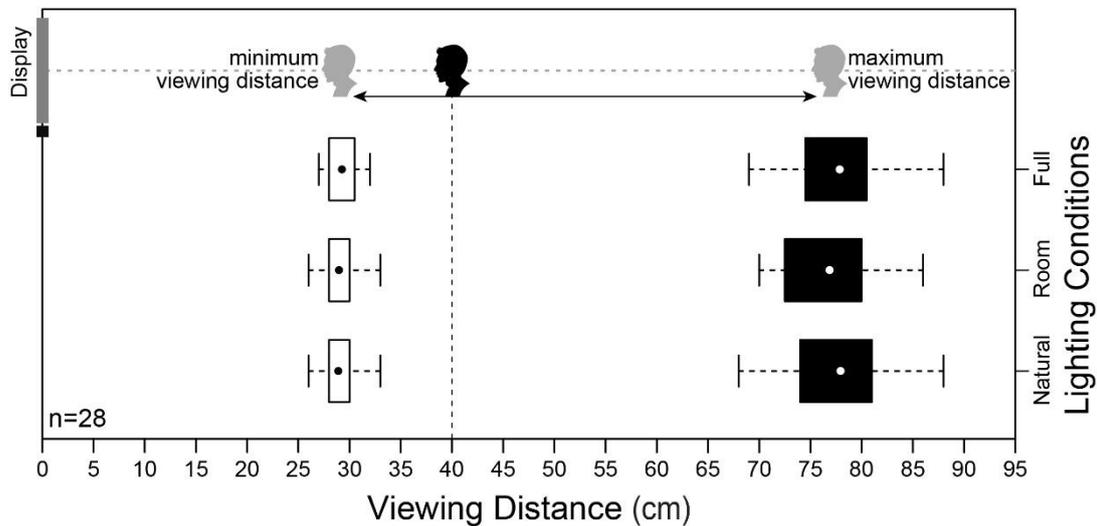
Figure 3.5. Results of Experiment 2. The maximum and minimum viewing distances for which the participant's eyes could be registered under the three lighting conditions. Error bars indicate ± 1 *SD*.

Paired comparisons were performed with Wilcoxon signed-rank tests to see which pair(s) showed a significant difference. Because of multiple comparisons, the alpha level was Bonferroni-adjusted to $0.05/3 = 0.017$. The maximum viewing distance obtained under the natural lighting condition neither differed from that obtained under the room lighting condition ($Z = -1.92$, $p = 0.055$), nor from that obtained under the full lighting condition ($Z = -0.38$, $p = 0.706$). The maximum viewing distance under the full lighting condition, however, was significantly larger than that measured under the room lighting condition ($Z = -2.82$, $p = 0.005$).

Figure 3.6 shows the highest and lowest viewing angle at which the participant's eyes (n = 28) could be registered under the three lighting conditions. The

average lowest viewing angle (white bars) in the natural light condition was 81.7 ± 1.3 degrees. In the room light condition, it was 81.3 ± 1.6 degrees, and in the full light condition, it was 81.9 ± 1.6 degrees. The average highest viewing angle (black bars) at which the participant's eyes could be registered was 113.9 ± 3.4 degrees in the natural light condition, 113.7 ± 3.2 degrees in the room light condition, and 113.4 ± 4.2 degrees in the full light condition.
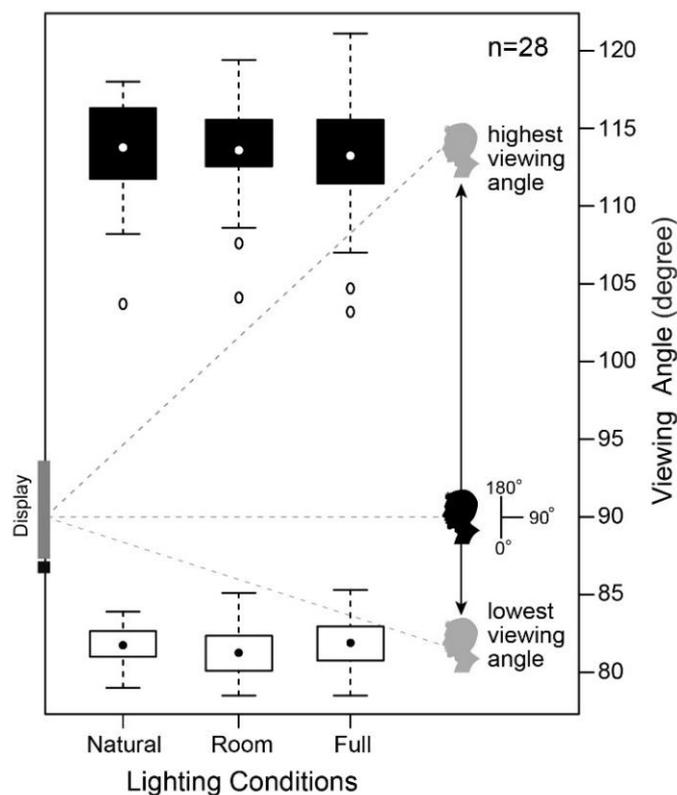
Figure 3.6. Results of Experiment 2. The lowest and highest viewing angles for which the participant's eyes could be registered under the three lighting conditions. Error bars indicate $\pm$ 1 *SD*.

Shapiro-Wilk tests showed that the highest viewing angle data measured under the natural lighting condition were not normally distributed ($df = 28$, $W = 0.91$, $p = 0.025$). The Friedman test ($df = 2$, n = 28) showed that there were no significant

differences between the highest viewing angles obtained under the three lighting conditions ($\chi^2 = 0.49$, $p = 0.782$). The Friedman test for the lowest viewing angles, however, showed significance ($\chi^2 = 9.14$, $p = 0.010$). Paired comparisons performed with Wilcoxon signed-rank tests and a Bonferroni-adjusted alpha level ($0.05/3 = 0.017$), taking multiple comparisons into account, showed that the lowest viewing angle obtained under the natural lighting conditions did not differ from that obtained under the room lighting ($Z = -1.66$, $p = 0.096$) and the full lighting conditions ($Z = -0.79$, $p = 0.428$). The lowest viewing angle under full lighting, however, was significantly higher than the lowest viewing angle under room lighting ($Z = -2.82$, $p = 0.005$). Details about the statistical analysis of the data obtained in Experiment 2 are in Appendix F.

3.4   Discussion

The first objective of Experiments 1 and 2 was to obtain maximum and minimum viewing distances at which the participant's eyes could be registered. The average maximum viewing distance and the average minimum viewing distance obtained with the Tobii EyeX© (Experiment 1; Figure 3.3) and the Eye Tribe© (Experiment 2; Figure 3.5) are relatively different. Experiment 1 showed that the participant's eyes (n=25) could be registered under three lighting conditions at the average maximum viewing distance of 70.3 cm and the average minimum viewing distance of 36.2 cm. Experiment 2 showed that participants (n=28) could register their eyes under three lighting conditions at the average maximum viewing distance of 77.5 cm and the average minimum viewing distance of 29.0 cm. The results showed that

Experiment 2 with the Eye Tribe© allows larger and closer viewing distances than as found in Experiment 1 with the Tobii EyeX©.

The second objective of Experiments 1 and 2 was to obtain the highest and lowest viewing angle at which eye-tracker devices worked at a viewing distance of 40 cm. There are obvious differences between the average lowest viewing angle and the average highest viewing angle obtained with the Tobii EyeX© (Experiment 1; Figure 3.4) and the Eye Tribe© (Experiment 2; Figure 3.6). This is most likely caused by the eyes' position on the Eye Tribe© device as used in Experiment 2, which was slightly above the middle of the display. When the participants were standing in front of the display and looked at the fixation point at a viewing angle of 90 degrees during the measurements, the Eye Tribe© system placed the participant's eye slightly upwards from the midline of the display. From a viewing distance of 40 cm, the Tobii EyeX© could register participants' eyes at the average lowest viewing angle of 71.4 degrees and the average highest viewing angle of 101.0 degrees (Experiment 1). The Eye Tribe© could register participants' eyes at the average lowest viewing angle of 81.6 degrees and the average highest viewing angle of 113.7 degrees (Experiment 2). In general, the highest viewing angle for the Eye Tribe© was wider than for the Tobii EyeX©, while the lowest viewing angle for the Eye Tribe© was narrower than that for the Tobii EyeX©.

The third objective of Experiments 1 and 2 was to investigate whether lighting conditions had an influence on the maximum and minimum viewing distances and viewing angles. Experiment 1, in which the Tobii X© eye-tracking device was used, showed that the maximum and minimum viewing distance and highest and lowest

viewing angle were not influenced by lighting conditions. No significant differences were found between the three lighting conditions for the maximum and minimum viewing distance and highest and lowest viewing angle. In other words, the limit of viewing distances and viewing angles of the Tobii X$^{©}$ eye-tracking device was relatively equal under different lighting conditions.

However, measurements with the Eye Tribe$^{©}$ in Experiment 2 were different. Experiment 2 showed that the maximum viewing distance under room lighting was smaller than that under full lighting. The average difference was small - only 1 cm, but nevertheless significant. Also when measuring with Eye Tribe$^{©}$, the lowest viewing angle was significantly higher when viewing under full light as compared to viewing under room light. Here too the average difference was small, just 0.6 degrees. It needs to be mentioned, though, that when measuring with the Eye Tribe$^{©}$ there were 'poor' calibration results, especially for participants who wore glasses. In Experiment 2, ten participants wore glasses. The poor calibration was obtained for 6 participants in natural light, 5 participants in room light, and 7 participants in full light. Research has shown that the calibration quality for eye-tracking devices was indeed poorer for participants with glasses as compared to those without glasses (Funke et al., 2016). Because of this, the effect of wearing glasses on calibration accuracy for low-cost eye tracking devices is investigated in Chapter 3. In conclusion, the results of Experiments 1 and 2 show that maximum and minimum viewing distances and angles can be stable under different lighting conditions at least for the Tobii X$^{©}$ eye-tracking device. For the Eye Tribe$^{©}$ device, however, small significant differences were found in some cases and the use of glasses may affect calibration quality.

Chapter 4. The use of glasses during registration into a low-cost eye tracking device under different lighting conditions

4.1   General Purpose

One of the issues that emerged from the findings in Chapter 3 was that there were some 'poor' calibration results, especially for participants who wore glasses when measuring with the Eye Tribe$^©$ eye-tracking device. In order for a participant to register onto an eye-tracking device, a calibration needs to be performed so that the eye-tracking software can generate a model to accurately estimate the viewer's gaze (Janthanasub & Meesad, 2015). Calibration is the process by which the characteristics of a viewer's eyes are assessed as the base for an accurate gaze point calculation (see Chapter 2 for details on eye-tracking calibration). The general purpose of the following Experiment 3 is to investigate the effect of wearing glasses on the calibration process into a low-cost eye-tracking device (Eye Tribe$^©$). The purpose of Experiment 4 is to investigate the ideal viewing angle of participants viewing from different heights (standing, sitting), with and without (replica) glasses at different display angles. In Experiment 3, the participant was asked to perform the calibration by following the movement of the circle on the display both with and without glasses under the three different lighting conditions. In Experiment 4, two different lighting conditions were used. The participants performed calibration with various display angles to obtain good accuracy and to mimic the viewing position for the use of an eye tracker in public spaces from different viewing angles of users. The full light condition was not used in

Experiment 4 since the reflection of the light spot appeared on the display at certain angles.

4.2 Experiment 3. The effect of wearing glasses on the calibration quality and calibration time

4.2.1 Purpose

In this experiment, the goal was to systematically test the influence of the use of glasses on the calibration quality and time for a low-cost eye-tracking device (Eye Tribe©) under three different lighting conditions. For participants with glasses, it is reported that the calibration quality of eye-tracking devices was indeed poorer (Funke et al., 2016). However, it is unclear whether the calibrations in this study were performed between users with prescription glasses and a group with normal eyesight, or among the same users with and without glasses. Moreover, in this study the calibration was obtained with users sitting in front of a display under a single room lighting condition (Funke et al., 2016). It is also understandable that a group with prescription glasses has more difficulty performing calibration (tracking) tasks because their eyesight is worse than that of users who do not need glasses (Nyström et al., 2013). In Experiment 3, therefore, the same participants with prescription glasses were asked to perform the calibration both with and without glasses, if they were able to do so, while the same participants with uncorrected vision were also asked to perform the calibration without glasses and with non-prescription, clear replica glasses. The measurements were conducted under three different lighting conditions, at a viewing distance of 40 cm.

4.2.2 Method

Participants

Sixteen participants (9 males and 7 females) were invited to participate in this experiment. Their ages ranged from 22 to 34 years ($M = 25.4$ years, $SD \pm 3.2$ years). Five participants had uncorrected vision, 4 participants wore contact lenses, and 7 participants wore glasses with an average thickness of 1.6 mm, $SD \pm 1.3$ mm. The height of the participants was in between 157 and 182 cm ($M = 168.2$ cm, $SD \pm 6.9$ cm). All participants were Asian (Chinese, Indonesian, Malaysian and Japanese). The participants participated on a voluntary basis and provided written, informed consent as to their participation after the purpose and procedure of the experiment was explained to them (Appendix G).

Apparatus

A touch-screen display (Microsoft Surface 4, 12-in) was equipped with the Eye Tribe© software. The calibration qualities were measured at a viewing distance of 40 cm under three lighting conditions: 'natural light' from a window, 'room light', and 'full light'. The 'full light' condition combined the 'natural light', the 'room light', and an additional light spot on the display to which the eye tracker was connected (Appendix E). For these three conditions, illuminance was measured at the point where the participant was viewing the display by a lux meter (i1 Pro XRITE). Next to illuminance, the display's luminance was measured under the three lighting conditions as well (TOPCON Luminance Meter BM-9). The display luminance was first measured for a dark display, and then for a bright display (with white paper placed on

the display to get brightness levels from a white surface when the display was off), at seven different points in time during a day in between 9:00 a.m. and 19:00 p.m. The illuminance in the natural, room and full light conditions was 10.91 ± 6.59 lux, 199.40 ± 9.96 lux, and 209.29 ± 9.22 lux on average, respectively. The display luminance in the natural, room and full light conditions for the dark display was 0.03 ± 0.02 cd/m^2, 0.86 ± 0.25 cd/m^2, and 6.15 ± 1.50 cd/m^2 on average, respectively. For the bright display it was 0.21 ± 0.13 cd/m^2, 4.13 ± 0.36 cd/m^2, and 30.90 ± 3.12 cd/m^2 on average, respectively. The participants without glasses were asked to wear replica glasses, and the participants with glasses wore their own. The replica glasses were non-prescription, round, and clear glasses with an average thickness of 1.2 mm, $SD \pm 0.2$ mm (Appendix H). Since the low-cost Eye Tribe$^©$ software could not record the calibration time, each calibration process was measured with a stopwatch.

Procedure

Written informed consent was obtained from each participant. After a face photo was taken (Appendix H), each participant was asked to follow the movement of a circle by using his/her eye-gaze in order to measure the calibration quality at a viewing distance of 40 cm (Appendix I). For each participant, the calibration quality was obtained based on the scaling values indicated by the Eye Tribe$^©$ software. The scaling value was 4 for 'Perfect', 3 for 'Good', 2 for 'Moderate', and 1 for 'Poor' including 'Redo' or 'Uncalibrated'.

First, the calibration qualities were measured for the participants who wore glasses. Depending on their eyesight, if they could see the calibration circle on the display without glasses, the participants were asked to follow the movement of the

circle on the display both with and without glasses under the three different lighting conditions. If they could not see the circle without their prescription glasses, the participants were asked to follow the circle only with their glasses. Second, the calibration qualities were also measured for the participants who did not wear glasses, including participants who wore contact lenses. Each participant was asked to follow the movement of a circle on the display by using his/her eye gaze with and without (replica) glasses under the three different lighting conditions. Each participant performed calibration at a viewing angle of 90 degrees without a head-and-chin rest.

Calibrations were obtained with counterbalance in the order of measurement with or without glasses. Thus half of the group with the non-prescription glasses performed the circle-following task first with replica glasses and then without. The other half performed the task first without and then with replica glasses. The order was also counterbalanced for the participants in the prescription-glasses group who were able to follow the circle without glasses. The time for each calibration under each lighting condition was measured, and the calibration was repeated when the result was 'Poor', with a maximum of three repetitions for each participant under each lighting condition. The measurements took about 40 minutes for each participant. The procedure for Experiment 3 and the following Experiment 4 was approved by the Ethical Committee of the Faculty of Design, Kyushu University, Japan (131-3).

4.2.3   Results of Experiment 3

The results show that the calibration quality was significantly poorer and calibration time was significantly longer for participants with glasses in the room and full lighting conditions than the same participants without glasses. Figure 4.1 shows

the calibration quality for participants (n = 16) with glasses and without glasses under the three lighting conditions. The calibration quality scaling based on the indication by the Eye Tribe© software was 4 for 'Perfect', 3 for 'Good', 2 for 'Moderate', and 1 for 'Poor' including 'Redo' or 'Uncalibrated'. The average calibration quality for participants with glasses (white bars) in the natural light condition was 2.06 ± 1.06. In the room light condition, it was 1.63 ± 1.03, and in the full light condition, it was 1.75 ± 1.13. The average calibration quality for participants without glasses (gray bars) was 2.88 ± 1.15 in the natural light condition, 2.50 ± 1.21 in the room light condition, and 2.88 ± 1.15 in the full light condition.
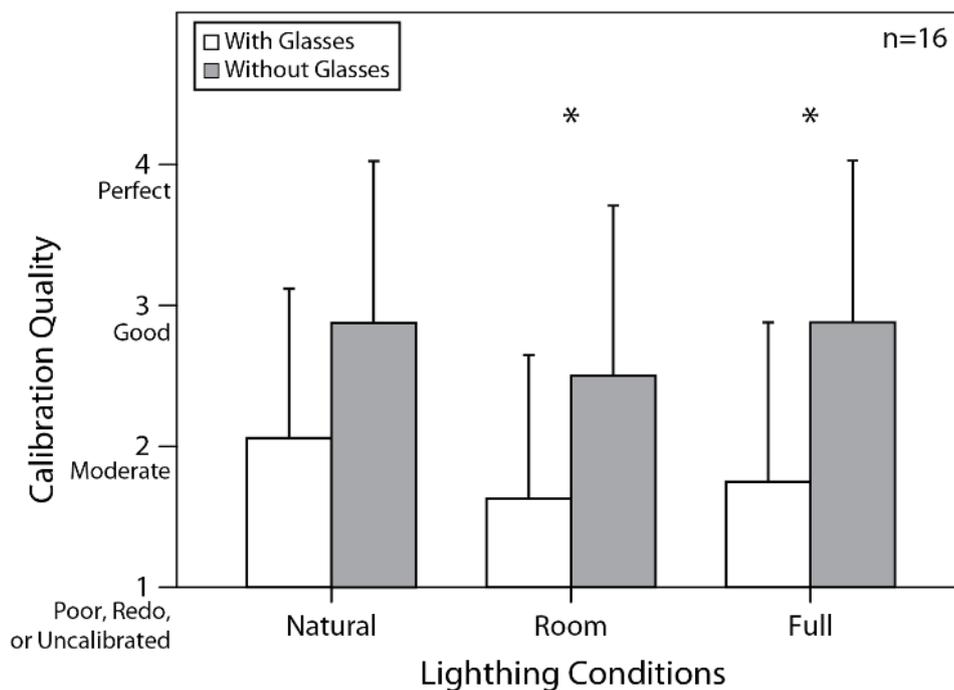
Figure 4.1. Results of Experiment 3. Eye-tracking calibration quality obtained with participants wearing prescription glasses or replica glasses (white bars) and the same participants without glasses (gray bars) under the three lighting conditions. Error bars indicate ± 1 *SD*. Asterisks indicate a significant difference between conditions ($p < 0.05$).

Shapiro-Wilk tests for calibration quality data obtained with participants who wore prescription glasses or replica glasses, and for calibration quality data gained with the participants who did not wear glasses showed that they were not normally distributed ($p < 0.05$) in three lighting conditions. Statistical analyses were therefore performed using a non-parametric Friedman test. The calibration quality data for the participants who wore glasses with repeated measures ($df = 2$, n = 16) showed no significant difference between lighting conditions ($\chi^2 = 5.83$, $p = 0.054$), although this bordered on significance. A significant difference in the calibration quality for the participants without glasses under three lighting conditions ($\chi^2 = 0.91$, $p = 0.636$) was also not found.

Paired comparisons were performed within each lighting condition for participants with and without glasses using Wilcoxon signed-rank tests in order to find out whether the use of glasses influenced calibration quality between groups (Appendix J). The calibration quality data showed no significant difference for the same participants with glasses and without glasses under natural lighting condition ($Z = -1.94$, $p = 0.052$), although this difference bordered on significance. However, the calibration quality was significantly poorer for the participants with glasses in the room lighting condition ($Z = -2.17$, $p = 0.030$) and in the full lighting condition ($Z = -2.47$, $p = 0.013$) than for the same participants without glasses.

Figure 4.2 shows the calibration time for participants (n = 16) with glasses and without glasses under the three lighting conditions. The average calibration time for participants with glasses (white bars) in the natural light condition was 17.64 ± 2.48 sec. In the room light condition, it was 18.75 ± 3.26 sec, and in the full light condition,

it was 19.52 ± 4.36 sec. The average time of calibration for participants without glasses (gray bars) was 16.38 ± 1.45 sec in the natural light condition, 16.75 ± 2.66 sec in the room light condition, and 16.01 ± 1.46 sec in the full light condition.
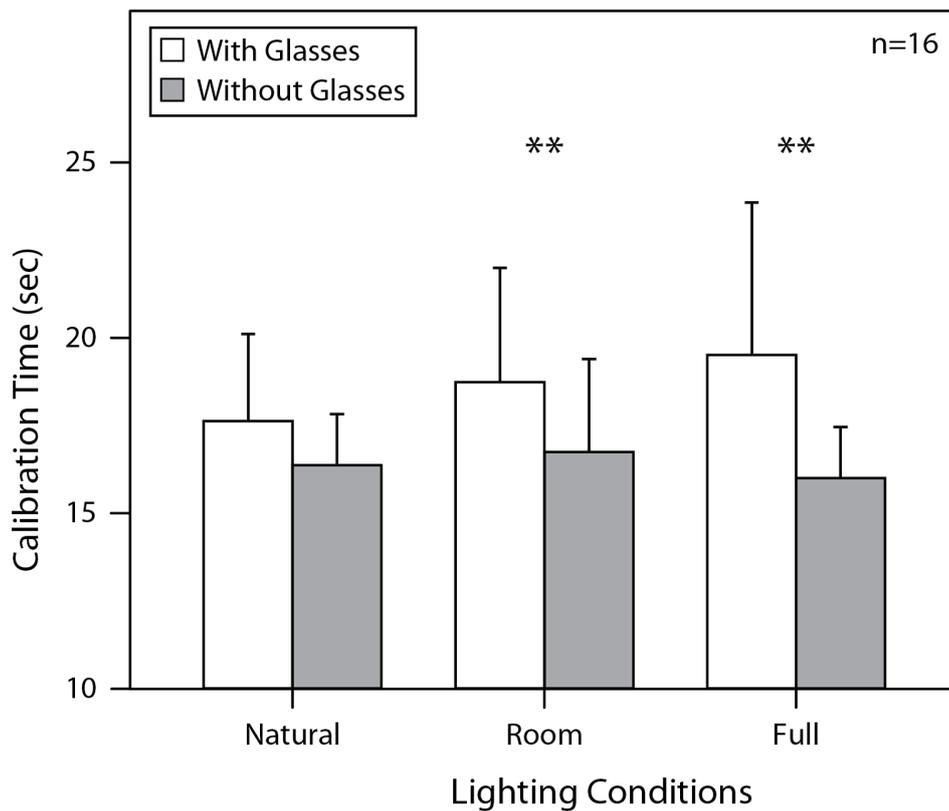
Figure 4.2. Results of Experiment 3. Eye-tracking calibration time for participants wearing prescription glasses or replica glasses (white bars) and the same participants without glasses (gray bars) under the three lighting conditions. Error bars indicate ± 1 *SD*. Asterisks indicate a significant difference between conditions ($p < 0.01$).

Shapiro-Wilk tests for the calibration-time data from participants who wore prescription glasses or replica glasses showed that the data were not normally distributed ($p < 0.05$) in three lighting conditions. For the calibration-time data from the participants without glasses showed that the data were not normally distributed ($p < 0.05$) in room and full lighting conditions. Friedman tests ($df = 2$, n = 16) showed

that for the participants who wore glasses there were no significant differences in calibration time under the three lighting conditions ($\chi^2 = 5.38$, $p = 0.068$). A significant difference between the calibration times for the participants who did not wear glasses under three lighting conditions ($\chi^2 = 2.42$, $p = 0.298$) was also not found (Appendix J).

To assess whether the use of glasses influenced the time to calibrate the eye-tracking device, paired comparisons were performed within each lighting condition for participants with and without glasses using Wilcoxon signed-rank tests (Appendix J). The calibration time data showed no significant difference between participants with glasses and without glasses under natural lighting ($Z = -1.70$, $p = 0.088$). However, the calibration time was significant longer for participants with glasses under room lighting ($Z = -2.99$, $p = 0.003$), and under full lighting ($Z = -3.26$, $p = 0.001$) than for the same participants without glasses.

4.3   Experiment 4. The ideal viewing angle of participants viewing from different heights, with and without glasses at different display angles

4.3.1   Purpose

Experiment 3 indicated that the calibration quality of the Eye Tribe© eye-tracking device was poorer and calibration time was longer for participants who wore glasses as compared to the same participants without glasses under the three different lighting conditions. It can be assumed that the result was not attributable to poor eyesight in the group with corrected vision; the participants with prescription glasses should have had difficulty performing the calibration task without their glasses, while

the participants without glasses (or with contact lenses) should have had no increased difficulty performing the calibration with replica glasses. One obvious problem with the use of (replica) glasses is the reflection or glare from (room) light on the glasses, making calibration difficult. However, some participants noticed another problem when viewing at 90 degrees with glasses. Tracking the calibration circle was especially difficult when the circle was on top of the display, because when viewing the display from a fixed viewing position, their eyes were occasionally occluded by the thick frame of the glasses. In Experiments 1 to 3, the participants used the eye-tracker device at a fixed viewing angle of 90 degrees.

In order to obtain good accuracy of the Eye Tribe$^{©}$ device, and in order to better mimic the viewing position for the use of an eye tracker in public spaces by users viewing from different angles, the purpose of Experiment 4 was to perform the calibration without a fixed viewing position. In Experiment 4, the participant was either asked to stand at a natural viewing position in front of the eye tracking device – as if he/she was using an ATM machine, while some participants were also asked to sit in front of the eye tracker. The Eye Tribe$^{©}$ is a low-speed system (sampling rate: 30 or 60 Hz), so free head movement does not affect the calibration quality (Ooms, Lapon, Dupont, & Popelka, 2015). The angle of the display and the eye tracker were systematically varied, and the ideal angle was investigated for participants of different viewing heights for registration into the system. Furthermore, these measurements were performed for participants with and without (replica) glasses at the different display angles under two different lighting conditions. In public settings, lighting will vary depending on weather conditions and artificial lighting. In order to avoid the

reflection of the light spot appearing on the display at certain angles in the full light condition, only the natural light and the room light conditions were used in Experiment 4. The measurements were conducted with counterbalance between the with/without glasses conditions and the two lighting conditions.

4.3.2 Method

Participants

Thirty participants (16 males and 14 females) were invited to participate in this experiment. Their ages ranged from 21 to 47 years ($M = 27.6$ years, $SD \pm 6.8$ years). Twelve participants had uncorrected vision, 7 participants wore contact lenses, and 11 participants wore glasses with an average thickness of 2.1 mm, $SD \pm 0.6$ mm. The height of the participants was between 145 and 181 cm ($M = 167.7$ cm, $SD \pm 9.5$ cm). Because I also wanted to obtain data from persons viewing from high positions (>185 cm) and from persons viewing from low positions (<140 cm), i.e., when sitting, some participants' viewing heights were manipulated.

To obtain data for persons viewing from below 140 cm, two participants were asked to sit on a chair (with a height of 46 cm from seat to floor). While they sat on the chair, their height from the top of their head to the floor were measured. To obtain data from persons taller than 185 cm, and thus viewing from a relatively high position, three participants with a height of 170 cm, 178 cm, and 181 cm, respectively, were asked to stand on an elevation at the height of 26 cm from the floor. Twenty-four participants were Asian (Japanese, Chinese, Indonesian), 1 participant was Caucasian, and 5 participants were Hispanic/Latinos. After the procedures were explained to the

participants, they signed an informed consent form (Appendix K). Each participant was paid for their participation.

Apparatus

The Eye Tribe© software was installed on a touch-screen display (Microsoft Surface 4, 12-in). The eye tracker was placed below the display, and together they were fixed on a tripod at the height of 133 cm from the ground. The tripod was used to facilitate the angle adjustment of the eye tracker and display. The different angles of the eye tracker and the display screen were measured under two lighting conditions: natural lighting and room lighting, as used in Experiment 3 (Appendix L). The full lighting condition was not used in this experiment because sometimes there was a reflection of spot light on the display when the display angles were changed. Each calibration process was timed with a stopwatch. Illuminance and luminance were measured at every display angle under two different lighting conditions in the same way and with the same equipment as in Experiment 3 (Table 4.1).

Table 4.1. Calibration of Experiment 4. The average of the luminance and illuminance on Experiment 4 in the two lighting conditions.

Illuminace (lux)						
Display angles	60	45	30	15	0	-15
Natural light	14.19 ± 7.46	13.60 ± 8.00	13.93 ± 7.57	12.47 ± 4.29	11.83 ± 3.89	15.74 ± 7.78
Room light	223.37 ± 18.52	224.10 ± 19.43	224.56 ± 14.67	216.41 ± 14.21	206.67 ± 7.82	209.73 ± 11.26
Luminance (cd/m^2) in the natural light						
Display angles	60	45	30	15	0	-15
Dark display	0.34 ± 0.29	0.37 ± 0.33	0.44 ± 0.39	0.46 ± 0.39	0.30 ± 0.27	0.16 ± 0.18
Bright display	2.83 ± 2.44	2.42 ± 2.01	2.33 ± 1.98	1.78 ± 1.51	1.69 ± 1.43	1.02 ± 0.95
Luminance (cd/m^2) in the room light						
Display angles	60	45	30	15	0	-15
Dark display	0.38 ± 0.31	0.42 ± 0.41	0.46 ± 0.37	0.57 ± 0.48	0.37 ± 0.29	0.24 ± 0.22
Bright display	3.39 ± 2.56	2.70 ± 2.07	2.58 ± 2.03	2.13 ± 1.61	1.61 ± 1.28	0.99 ± 0.88

Procedure

        After written informed consent was obtained and a face photo (Appendix M) was taken at a viewing distance of 40 cm, the participant was asked to follow the instructions. First, the participant was asked to stand in the middle in front of the screen without crossing a floor mark set at a viewing distance of 40 cm. While standing, the participant was asked to relax and to take a natural viewing position. Each participant was asked to maintain this natural viewing position during calibration, as if he/she was using an ATM machine. Two participants were instructed to sit and to maintain a natural viewing position throughout the calibration. All participants were instructed to make no head movements even when the angle of the display was changed. Second, the participant was asked to perform calibration under different display and eye tracker angles, in order to assess their ideal viewing angle. The following 9 display angles were used: 60, 45, 30, 15, 0, -15, -30, -45, and -60 degrees (Figure 4.3).
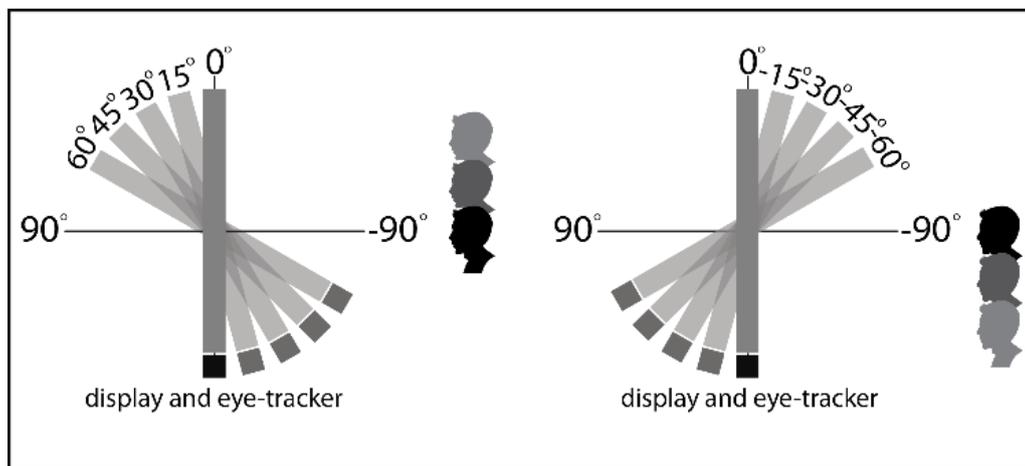
Figure 4.3. The angles of the display and eye tracker used in Experiment 4. Participants were asked to maintain a natural viewing position while they tried to register into the eye-tracking system under various display and eye-tracker angles in random order.

Third, with regard to the participants' eyesight (with or without glasses), the same procedure as in Experiment 3 were performed. Thus, the calibration quality was obtained by asking each participant to follow the movement of a circle with his/her eye-gaze. For each participant, the calibration quality was obtained based on scaling values, that is, 4 for 'Perfect', 3 for 'Good', 2 for 'Moderate', and 1 for 'Poor' including 'Redo' or 'Uncalibrated'. First, the calibration qualities were measured for the participants who wore glasses. Depending on their eyesight, if they could see the calibration circle on the display without glasses, the participants were asked to follow the movement of the circle on the display both with and without glasses under the three different lighting conditions. If they could not see the circle without their prescription glasses, the participants were asked to follow the circle only with their glasses. Second, the calibration qualities were also measured for the participants who did not wear glasses, including participants who wore contact lenses. Each participant was asked to follow the movement of a circle on the display by using his/her eye gaze with and without (replica) glasses under the three different lighting conditions.

The participant performed the calibration both with and without glasses under the 9 display angles, under the two different lighting conditions. I did not measure the calibration time when the eyes of the participants did not appear on the calibration display, indicating that the eye tracker could not detect the participant's eyes (invalid tracking). Each participant performed the calibration with randomized display angles and without a head-and-chin rest. The order of wearing glasses and lighting conditions was counterbalanced among participants. The measurement was not repeated when the

calibration result was 'Poor', 'Redo' or 'Uncalibrated'. The measurements took about 60 minutes for each participant.

4.3.3 Results of Experiment 4

The data show that 16 participants could register into the eye-tracking device at two display angles, and 14 participants could register at one display angle. The display angles depended on the participant's viewing height. From here on, the display angles will be called the "first angle" (n = 30) and "second angle" (n = 16). A Pearson's correlation analysis was performed to determine the relationship between the viewing height of the participants and the angles of the eye tracker and the display that allowed registration (Appendix N). There was a significant correlation between the viewing height of the participants and the first angle of the display they could register themselves with ($r = 0.94$, n = 30, $p < 0.001$), as well as a significant correlation between the viewing height of the participants and the second angle of the display they could register with ($r = 0.97$, n = 16, $p < 0.001$).

Figure 4.4 shows the box plot of the data, showing a linear relation between the angles of the display and the height of the participants under which they could register themselves into the eye-tracking system. Roughly summarized over the data of both angles combined, the participants with a height of 200-220 cm could register into the eye-tracking device at a display angle of 60 or 45 degrees. The participants of 180-200 cm could register at 45 or 30 degrees, while participants of 160-180 cm could register at display angles of 30 and 15 degrees, and those of 140-160 cm could register at display angles of 15 and 0 degrees. The participants who were sitting and viewing from below 140 cm could register at 0 and -15 degrees. The results convincingly show

that in order to use eye trackers in a public setting, e.g., for ATM machines, users need to be able to adjust the display angle according to their height. Especially people viewing from below 140 cm, such as wheel-chair users, require a very different angle than people standing up.
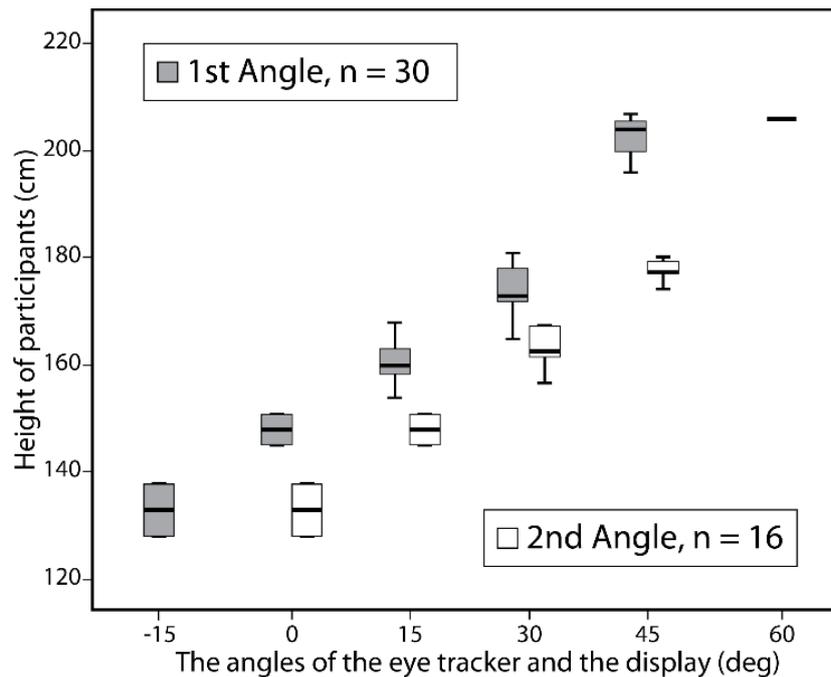
Figure 4.4. Results of Experiment 4. The gray boxes and white boxes show the first and second angles, respectively, of the eye tracker and the display under which participants could register into the eye-tracking device. The angles are plotted against the participant's (viewing) heights. Error bars indicate ± 1 *SD*.

Similar to the results of Experiment 3, the quality of the registration into the eye-tracking device is better without glasses than with glasses as shown in Figure 4.5. The calibration quality for participants with glasses (white bars) at the first angle was 2.80 ± 1.35 under natural lighting, and 2.93 ± 1.26 under room lighting. At the second angle, it was 2.63 ± 1.26 and 2.88 ± 1.41 for natural and room lighting, respectively. The calibration quality for the same participants without glasses (gray bars) at both

angles was higher; at the first angle it was 3.30 ± 1.03 under natural lighting and 3.04 ± 1.20 under room lighting. At the second angle, it was 3.20 ± 0.98 under natural lighting and 3.44 ± 1.03 under room lighting.
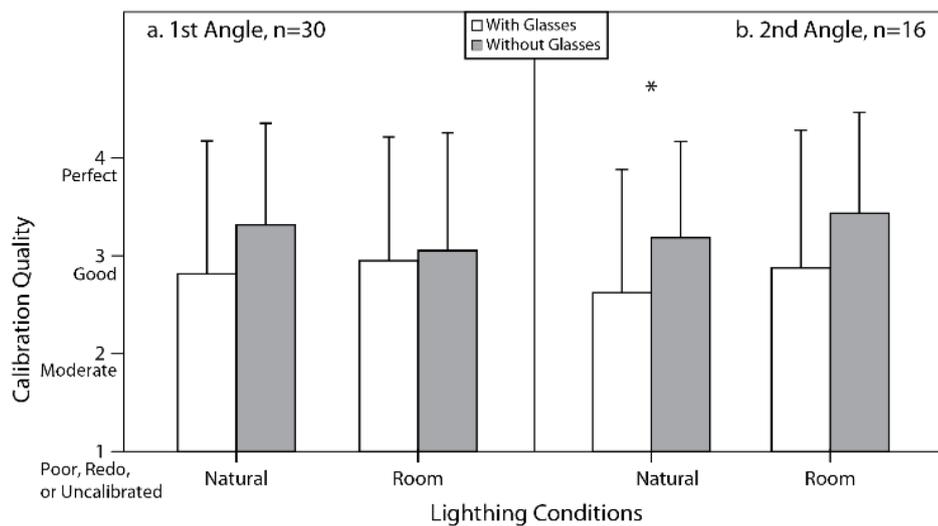
Figure 4.5. Results of Experiment 4. Eye-tracking calibration quality obtained at the first and second angle with participants wearing prescription glasses or replica glasses (white bars) and the same participants without glasses (gray bars) obtained under two lighting conditions. Error bars indicate ± 1 *SD*. Asterisks indicate a significant difference between conditions ($p < 0.05$).

Paired comparisons were performed to determine whether the use of glasses influenced calibration quality within each lighting condition for participants with and without glasses using Wilcoxon signed-rank tests. Three participants with prescription glasses were not able to perform the calibration without glasses at their first angle. Their data were excluded from the tests (Appendix N). The difference in the first-angle calibration quality between participants with and without glasses under natural lighting ($Z = -1.92$, $p = 0.056$) and room lighting ($Z = -0.32$, $p = 0.975$) was not significant, although the difference under natural lighting bordered on significance. Similarly, the second-angle calibration quality was not significantly different between participants

with glasses and without glasses under room lighting ($Z = -1.71$, $p = 0.088$). However, calibration quality at the second angle was significantly higher under natural lighting ($Z = -2.12$, $p = 0.034$).

Figure 4.6 shows that the registration time was also shorter overall without glasses. The calibration time needed by participants without glasses (gray bars) at the first angle (natural lighting: 15.16 ± 0.53 sec; room lighting: 15.27 ± 1.21 sec) and at the second angle (natural lighting: 15.43 ± 1.57 sec; room lighting: 15.19 ± 1.00 sec) was shorter than the calibration time needed by the same participants with glasses (white bars) at the first angle (natural lighting: 15.85 ± 1.60 sec; room lighting: 15.98 ± 2.36 sec) and at the second angle (natural lighting: 15.97 ± 1.70 sec; room lighting: 15.60 ± 1.37 sec).
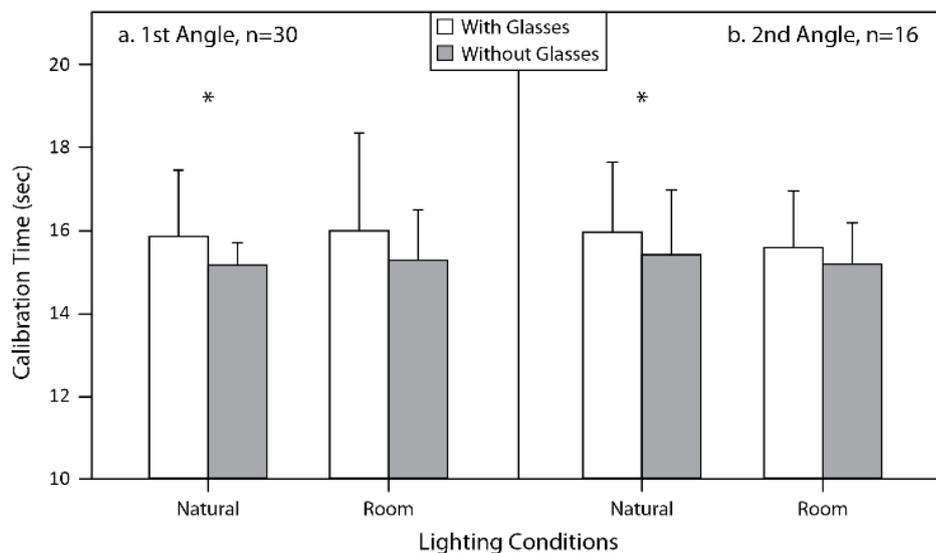
Figure 4.6. Results of Experiment 4. Eye-tracking calibration time obtained at the first and second angle for participants wearing prescription glasses or replica glasses (white bars) and the same participants without glasses (gray bars) obtained under two lighting conditions. Error bars indicate ± 1 *SD*. Asterisks indicate a significant difference between conditions ($p < 0.05$).

Paired comparisons were also performed using Wilcoxon signed-rank tests to determine whether the use of glasses influenced calibration time within each lighting condition for participants with and without glasses (Appendix N). A significant difference was found for the first-angle data ($Z = -2.42$, $p = 0.015$) and the second-angle data ($Z = -2.44$, $p = 0.015$) between participants with glasses and without glasses under natural lighting. No significant difference at the first angle ($Z = -1.83$, $p = 0.068$) and at the second angle ($Z = -0.68$, $p = 0.495$) was found under room lighting.

4.4   Discussion

Experiment 3 aimed to systematically examine the influence of the use of glasses on the calibration quality and time for the Eye Tribe$^{©}$ eye-tracking device under three different lighting conditions. The results of Experiment 3 showed that the calibration quality of participants who wore prescription glasses or replica glasses was 'Moderate' (average 2.06 on a scale of 4) under natural lighting. The calibration under natural lighting with the same participants without glasses, though, was closer to 'Good' (average 2.88 on a scale of 4). In the room lighting and the full lighting conditions, the calibration quality for participants with glasses was similarly 'Moderate' (room lighting: 1.63; full lighting: 1.75), and almost 'Good' without glasses (room lighting: 2.50; full lighting: 2.88). Experiment 3 also showed that the calibration time required by participants was about 1-3 seconds longer on average when they wore (replica or prescription) glasses than when they did not. The negative effect of glasses on the calibration quality and time as found here could be due to the presence of light

reflection and glare, or due to the fact that when tracking an object high in the visual field, the participants' eyes were occluded by the frame of the glasses.

The first goal of Experiment 4 was to investigate the ideal viewing angle of participants viewing from different heights (standing, sitting) for registration of users' eyes. The results of Experiment 4 showed a significant correlation between the height of the participants and the angles of the eye tracker and the display under which they could register themselves. In short, if the participant was tall or viewing from a high position looking down on the display, the display should be angled upwards in a more horizontal position for registration to occur. By contrast, people sitting down on chairs or in wheelchairs would benefit from a vertical display position or a downward angle. From this result, it is clear that users are able to correctly register their eyes when display angles are according to their viewing position (standing or sitting).

The second goal of Experiment 4 was to perform calibration for participants with and without (replica) glasses at the different display angles under two different lighting conditions. Overall, the results of the calibration quality and the calibration time obtained in Experiment 4 were better than in Experiment 3. Although no statistics were performed due to differences in group sizes, the average quality for participants with glasses under natural and room lighting in Experiment 3, from a 90-degrees viewing angle, was 2.06 and 1.63, respectively. In Experiment 4 the average quality markedly improved: 2.80 and 2.93 at the first angle, respectively, and 2.63 and 2.88 at the second angle, respectively. The calibration time obtained for participants with glasses in Experiment 4 under natural and room lighting (at the first angle 15.85 sec and 15.98 sec, and at the second angle 15.97 sec and 15.60 sec, respectively) was also

faster than that obtained in Experiment 3 (17.64 sec and 18.75 sec, respectively). For achieving good and fast calibration, therefore, the eye-tracking device and the display need to be set at a certain angle. Nevertheless, more research on the efficiency of the low-cost eye-tracker with regard to the thickness of the prescription glasses still needs to be done and checked in the future.

It is known that eye-tracking devices allow the user to select a specific object from the screen interface only by gazing on it. An interactive interface with multiple objects shown on the screen typically employs a grid to organize objects based on sequenced columns and rows. For example, a grid is typically used for screen-lock interfaces of personal computers or smartphones. In visual password systems, grids with higher densities (more object keys) and different formations have been considered and tested (for more information about grid, see Chapter 2). The use of higher grid densities potentially enables complex passwords, since higher grids have a higher number of object keys. Research on the relation between grid density and password complexity, however, has shown mixed results. It was reported that the use of grid densities of more than 4×4 cells had minimal influence overall on the complexity of passwords (Aviv et al., 2015). In contrast, increasing the grid density increased password complexity (Thorpe & van Oorschot, 2004; Alam, 2016). In these studies, the grids were tested with manual input. Moreover, no systematic, comparative research about grid formations has been performed. Therefore, it is necessary to investigate what type of grid densities and formations is suitable to use with low-cost eye trackers. This matter is further investigated in Chapter 5.

Chapter 5. A preliminary experiment on grid densities for visual password formats

5.1 Experiment 5. Suitable grid densities for use with a low-cost eye-tracker

5.1.1 Purpose

In order to select objects on a display with eye tracking, often some kinds of grid densities and formations are used. Depending on the size and the number of objects, the grid density and formation changes. In earlier research grid densities and formations of 3×3 to 10 × 10 cells have been used in interactive tasks with manual input (see Chapter 2). However, the use of grid formations in previous studies just focused on increasing the password complexity with denser grids. In addition, comparative research about grid formations has not yet been performed systematically. The goal of Experiment 5 was to identify which grid densities potentially are suitable for use with a low-cost eye-tracking device. To achieve this purpose, it is important to obtain judgments related to how easy to use and how safe the user thinks each grid density would be, when authenticating a password.

In Experiment 5, sixteen different grid formations were used in between 2×2 to 7×7 cells (see Figure 5.1). For convenience, the participant was asked to sit in front of a computer and to create an ideal visual password by selecting objects on the display using manual input with a mouse. The ideal visual password was an imaginary password that consisted of four to eight objects (alphanumeric characters, dots, or

visual icons). Another reason why the participant was sit down is that the actual eye tracking had not yet been used for making a password in this experiment. Following this, the participant had to give assessments for three visual password formats on grid densities, in relation to how easily and safely he/she thought each grid density could be used for making a visual password in a real situation with eye tracking. The range of possible grid densities was meant for a follow-up experiment with actual eye tracking (see Experiment 6, Chapter 6). A 7–point rating scale was used to obtain participant judgments.

5.1.2 Method

Participants

Twenty-seven participants (11 males and 16 females) with normal or corrected-to-normal vision participated in this experiment on a voluntary basis. They were 21-47 years of age ($M = 27.0$ years, $SD = \pm 6.5$ years). Twenty-four participants were Asian (Japanese, Chinese, Malaysian, and Indonesian), 2 participants were Caucasian, and 1 participant was Latino/Hispanic. All participants gave informed consent as to their participation after the purpose and procedure of the experiment was explained to them (Appendix O).

Apparatus

The interface for the grids was presented on a monitor (Lenovo ThinkVision, 20-in, refresh rate of 60 Hz) with a resolution of 1600×1200 pixels. The monitor was placed on the desk with a height of 70 cm from the floor and at a viewing distance of approximately 62 cm where the participant was sat. The interface for the questions

about the grid was presented on a Laptop (Lenovo Z40, 15-in) with a resolution of 1366×768 pixels and a refresh rate of 60 Hz. This laptop was placed below the monitor (Appendix P). The interface was programmed in Visual Studio C# (2015), and the data gathered from the participants were saved in a MySQL database. The experiment was performed under a room lighting condition at an illuminance of 106.42 ± 11.36 lux, as measured using a TOPCON Illuminance Spectro Meter IM-1000 at the location where the participant was seated. Stimulus luminance on the monitor was measured using a TOPCON Luminance Meter BM-9. The luminance and illuminance values used here were similar to those used with the same experimental set-up as used in Experiments 1 to 4 (Chapters 3 and 4). The measurements were performed ten minutes before the start of the experiment.

Stimuli

For three visual password formats, sixteen different grid densities were made with 2×2 to 7×7 cells (columns × rows, see Figure 5.1). The first format was an alphanumeric visual password format. For the alphanumeric format, numbers, letters, and special characters were presented on each grid. One cell of a grid was always reserved for a 'Clr' (Clear) key. This key could be used by the participant to clear his/her input. In the case of a grid density of four cells, i.e., a 2×2 grid, the digits 1, 2, 3, and 'Clr' were presented starting from the top left to the bottom right cell. If the grid density exceeded ten cells, letters (in alphabetical order) and special characters were added to the numbers. For example, on a 7×7 grid with 49 cells, the numbers 0 to 9, the letters A to Z, and the special characters -, +, *, /, =, :, ;, ?, !, (, \,), along with 'Clr' were presented starting from the top left to the bottom right cell. The alphanumeric

characters in the grid cells were black with a luminance of 0.73 ± 0.21 cd/m^2, on a white background (8.45 ± 0.65 cd/m^2).

The second format was a pattern format. The pattern format consisted of dots, which could be selected by the participant to create a visual pattern as a password. The dots were black (0.80 ± 0.26 cd/m^2) and white (8.12 ± 0.92 cd/m^2). That is, a white dot with a radius of 47 pixels was placed in the middle of a black dot with a radius of 128 pixels, and both dots were presented together as a key.
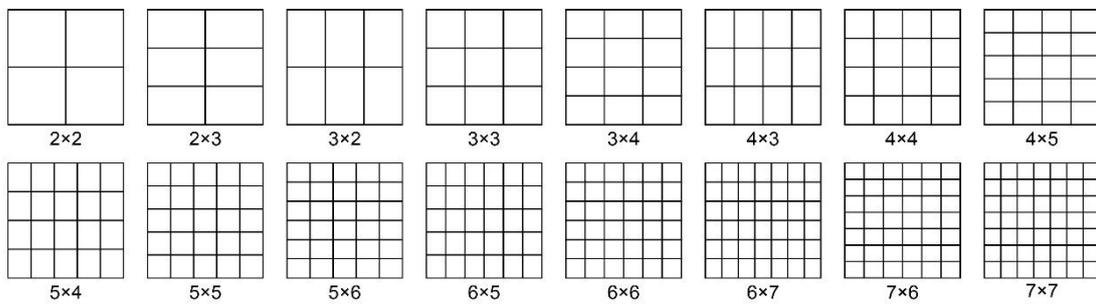
Figure 5.1. Sixteen different grid densities for three visual password formats were used in Experiment 5.

The third format was a picture format, which consisted of visual icons in a fixed order on a grid, with the number of icons depending on the grid density. Similar to the alphanumeric format, a 'Clr' key with the same function occupied one cell of a grid in this format. The visual objects were a single picture in gray with a luminance of 3.10 ± 1.04 cd/m^2, against a white background (8.29 ± 0.94 cd/m^2). Fruits, animals, and familiar objects are examples of visual objects used in this experiment and the two following experiments, Experiments 6 and 7 (Appendix Q).

Each key of a password (i.e., alphanumeric character, dot, or visual icon) was put in a grid cell with a size of 128×128 pixels, which was 3.29 degrees in visual angle

and every pixel within a key was 0.027 degrees in visual angle. A stimulus pane was made with a size of 1600×1200 pixels for each grid density set against a gray background (3.64 ± 0.09 cd/m^2) for each of the three visual password formats. A text box was displayed on the upper part of the pane (1600×125 pixels), while a grid was displayed for a given password format on the main part of the pane, with a size of 1600×1075 pixels (see Figure 5.2). When the participant selected a key (an alphanumeric character, a dot, or a visual icon), an asterisk would be displayed on the text box above the grid, and a chime sound would be played (1538 ms; Appendix R), to indicate that the participant had made a selection.
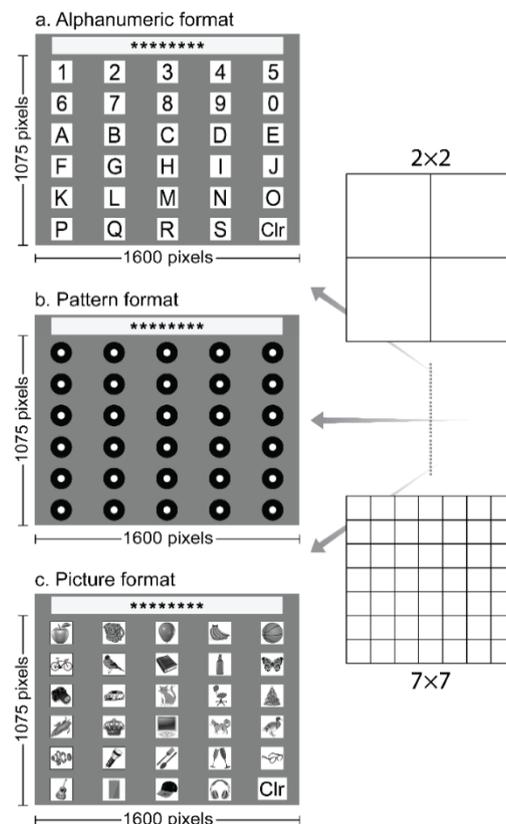
Figure 5.2. Example of stimuli used in Experiment 5. The three visual password formats [a. The alphanumeric format with a grid of 5×6 cells (columns × rows), b. The pattern format with a grid of 5×6 cells (columns × rows), and c. The picture format with a grid of 5×6 cells (columns × rows)] were presented in each of the 16 different grids depicted in Figure 5.1.

Procedure

The participant was asked to sit in front of a desk on which the laptop and monitor were placed. The participant was asked to make an "ideal" password for each grid density on three visual password formats. Each grid was displayed on the monitor in a random order when the participant made a password. He/she was instructed never to create a real password used in his/her daily life. Furthermore, although each participant was using a mouse to make passwords, he/she was asked to imagine using the eye-tracking system during the experiment. The participant was instructed that the password should be an imagined password, with a minimum length of four and a maximum length of eight objects. For each format, the imagined password was created by clicking objects on the monitor using a mouse. For the alphanumeric format, the participant had to select a sequence of characters such as numbers, letters or special characters, as his/her password. For the pattern format, he/she had to draw an assembly of dots as a password, while for the picture format, the participant had to select icons/images as his/her password.

After making a password, the participant was asked whether he/she thought that a particular grid density would be easy to use for making a visual password, on a scale between 1 (not easy) and 7 (very easy). Next, the participant was instructed to evaluate the potential safety of the grid for visual password input with eye tracking, between 1 (not safe) and 7 (very safe). The participant was not given any instructions as regards the meaning of "easy to use" or "potential safety" of the visual password format and grid. These evaluations were performed with a program on the laptop. Each evaluation for each grid was saved in a database.

In order to get used to the program interface, the participant was asked to run a practice program before starting the actual experiment. In the practice program, the participant had to practice making a password twice for each format, on a grid randomly chosen from the 16 grids. After that, the participant made the evaluations. After practice, the participant performed the experiment with counterbalance in the order of the three visual password formats. That is, nine participants first made passwords in the alphanumeric format, then in the pattern format, and finally in the picture format, for each of the grids. Another nine participants started with the pattern format, followed by the picture format, and ending with the alphanumeric format. The remaining nine participants started with the picture format and ended with the pattern format. The experiment for all formats took about 60 minutes for each participant. The procedure was approved by the Ethical Committee of the Faculty of Design, Kyushu University, Japan (131-3).

5.2   Results of Experiment 5

For every visual password format, with each increase in the number of grid cells, a grid was evaluated difficult to use but it was considered potentially safety for password formation in an imagined situation using eye tracking. Figure 5.3 shows the participant judgments (n = 27) on whether a grid was potentially easy to use for three visual password formats. As the number of grid cells increased, the participants judged the grid as increasingly less easy to use for the alphanumeric format [white circles, ranging from the 2×2 grid (6.30 ± 0.54), to the 7×7 grid (3.52 ± 0.84)]. Participants

also assessed the grid progressively not easy to use for the pattern format (gray circles) and the picture format (black circles) when the number of cells in the grids increased.
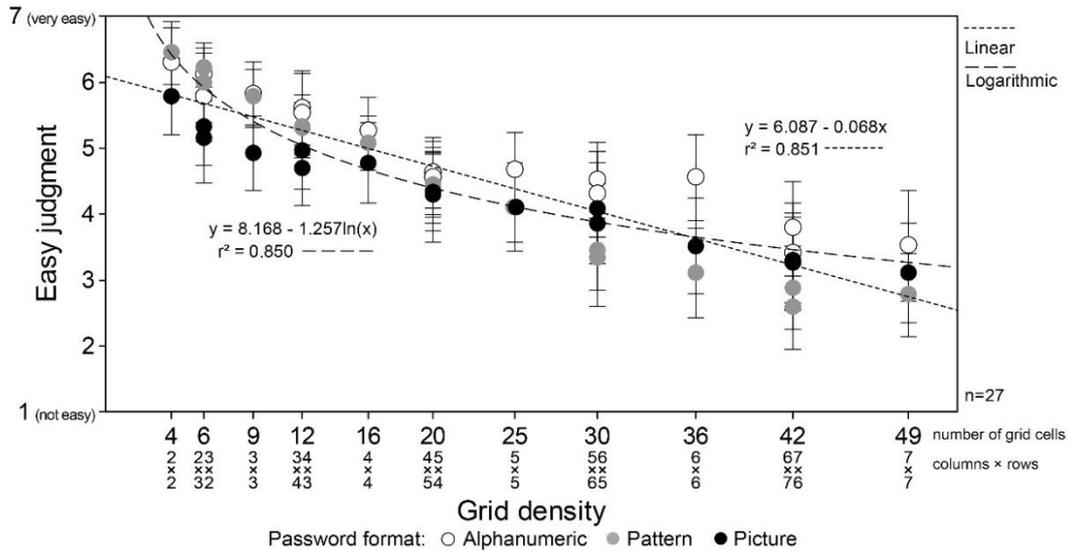
Figure 5.3. Results of Experiment 5. The participant judgments (n = 27) on whether a grid was potentially easy to use for three visual password formats. The circles show the average grid evaluations for alphanumeric passwords (white), pattern passwords (gray), and picture passwords (black). The lines show the linear (short dashes) and logarithmic (long dashes) functions for the relation between the judgment and the grid density for the three visual password formats. Error bars indicate ± 95% confidence intervals around the means.

In order to examine the the correlation between the "easy" judgment and grid density, a regression analysis with a linear (5.1) and a logarithmic (5.2) function was performed.

$$y = a + bx \tag{5.1}$$

$$y = a + b\ln(x) \tag{5.2}$$

where x is the number of grid cells (from 4 to 49 cells), x must be greater than zero for Equation 5.2, y is the participant judgment (from 1 to 7), a is the y-intercept, and b indicates the slope value. The linear ($y = 6.087 - 0.068x$) and logarithmic ($y = 8.168 -$

1.257ln(x)) regression equations with r^2 values of 0.851 and 0.850, respectively, show that the average "easy" judgment significantly decreased ($p < 0.001$) with each increase in the number of grid cells (Appendix S). Both the linear and the logarithmic function fitted the data about equally well.

Table 5.1 shows the 95% confidence intervals (CIs) around the means (n = 27) of the "easy" judgment for three visual password formats. In order to get an indication of which grid density would be useful for each format, the interval ranges were compared with the midpoint of the rating scale, i.e., "4". For the alphanumeric format, the means and the 95% CIs for the 2×2, 2×3, 3×2, 3×3, 3×4, 4×3, 4×4, 4×5, and 5×5 grids exceeded the scale midpoint, suggesting that these grid densities tended to be easy to use with this format. For the pattern and picture formats, the means and the 95% CIs for the 2×2, 2×3, 3×2, 3×3, 3×4, 4×3, and 4×4 grids were greater than the scale midpoint ("4"), suggesting that these grid densities were thought to be easy to use with both of these password formats.

Table 5.1. Results of Experiment 5. The means and the 95% confidence intervals of the easy judgment for each grid density for three visual password formats.

Grid Density (columns × rows)	Alphanumeric format (M, 95% CIs)	Pattern format (M, 95% CIs)	Picture format (M, 95% CIs)
2×2	6.30 [5.76,6.83]	6.44 [5.96,6.93]	5.78 [5.19,6.36]
2×3	5.78 [5.11,6.45]	6.22 [5.85,6.59]	5.33 [4.74,5.92]
3×2	6.11 [5.71,6.51]	6.00 [5.40,6.60]	5.15 [4.48,5.81]
3×3	5.81 [5.32,6.31]	5.78 [5.36,6.19]	4.93 [4.37,5.48]
3×4	5.59 [5.04,6.14]	5.33 [4.86,5.81]	4.70 [4.14,5.27]
4×3	5.52 [4.87,6.17]	5.30 [4.89,5.70]	4.96 [4.39,5.54]
4×4	5.26 [4.75,5.77]	5.07 [4.66,5.48]	4.78 [4.17,5.38]
4×5	4.63 [4.09,5.17]	4.44 [3.96,4.93]	4.30 [3.58,5.01]
5×4	4.56 [4.00,5.11]	4.41 [3.87,4.95]	4.33 [3.75,4.91]
5×5	4.67 [4.10,5.24]	4.11 [3.57,4.65]	4.11 [3.44,4.78]
5×6	4.52 [3.94,5.09]	3.44 [2.85,4.04]	4.07 [3.37,4.78]
6×5	4.30 [3.65,4.94]	3.33 [2.60,4.07]	3.85 [3.25,4.45]

6×6	4.56 [3.90,5.21]	3.11 [2.44,3.78]	3.52 [2.79,4.25]
6×7	3.78 [3.06,4.49]	2.89 [2.26,3.52]	3.26 [2.55,3.97]
7×6	3.41 [2.65,4.16]	2.59 [1.96,3.23]	3.30 [2.57,4.02]
7×7	3.52 [2.68,4.36]	2.78 [2.15,3.40]	3.11 [2.36,3.86]

Figure 5.4. shows the participant judgments (n = 27) on whether a grid was potentially safe to use for three visual password formats. As the number of grid cells increased, the participants thought that in an imagined situation using eye tracking, making an alphanumeric visual password would become safer [white circles, ranging from the 2×2 grid (1.26 ± 0.27), to the 7×7 grid (6.44 ± 0.37)]. When the grid became denser, participants also thought the grid increasingly safer to use in an imagined situation in which eye tracking would be used to make the pattern password (gray circles) and the picture password (black circles).
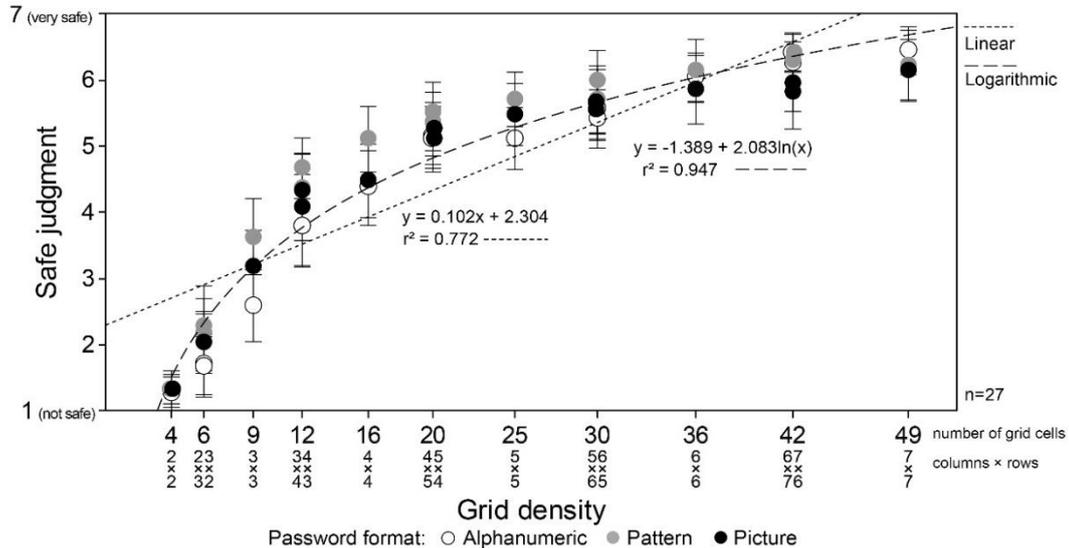
Figure 5.4. Results of Experiment 5. The participant judgments (n = 27) on whether a grid was potentially safe to use for three visual password formats. The circles show the average grid evaluations for alphanumeric passwords (white), pattern passwords (gray), and picture passwords (black). The lines show the linear (short dashes) and logarithmic (long dashes) functions for the relation between the judgment and the grid density for the three visual password formats. Error bars indicate ± 95% confidence intervals around the means.

A regression analysis with a linear (5.1) and a logarithmic (5.2) function was performed to examine the correlation between the "safe" judgment and grid density. The linear ($y = 0.102 + 2.304x$) and logarithmic ($y = -1.389 + 2.083\ln(x)$) regression equations with r^2 values of 0.772 and 0.947, respectively, show that the average "safe" judgment significantly increased ($p < 0.001$) with each increase in the number of grid cells (Appendix S). Here, the fit of the logarithmic function was better than that of the linear function.

Table 5.2 shows the 95% CIs around the means (n = 27) of the "safe" judgment for three visual password formats. A comparison was performed between interval ranges with the midpoint of the rating scale, i.e., "4", to get an indication of which grid density would be useful for each format. For the pattern format, the means and the 95% CIs for the 3×4, 4×4, 4×5, 5×4, 5×5, 5×6, 6×5, 6×6, 6×7, 7×6, and 7×7 grids exceeded the scale midpoint, suggesting that these grid densities were considered potentially safe for this format. For the alphanumeric and picture formats, the means and the 95% CIs for the 4×5, 5×4, 5×5, 5×6, 6×5, 6×6, 6×7, 7×6, and 7×7, grids were higher than the scale midpoint, showing that these grid densities were judged as potentially safe for both these password formats.

Table 5.2. Results of Experiment 5. The means and the 95% confidence intervals of the safe judgment for each grid density for three visual password formats.

Grid Density (columns × rows)	Alphanumeric format (M, 95% CIs)	Pattern format (M, 95% CIs)	Picture format (M, 95% CIs)
2×2	1.26 [1.00,1.52]	1.33 [1.06,1.60]	1.33 [1.11,1.55]
2×3	1.70 [1.24,2.17]	2.19 [1.67,2.70]	2.04 [1.61,2.47]
3×2	1.67 [1.20,2.13]	2.30 [1.71,2.89]	2.04 [1.57,2.51]
3×3	2.59 [2.04,3.14]	3.63 [3.06,4.20]	3.19 [2.65,3.72]
3×4	3.78 [3.20,4.35]	4.67 [4.20,5.13]	4.33 [3.78,4.88]
4×3	3.78 [3.17,4.38]	4.37 [3.84,4.90]	4.07 [3.58,4.56]

4×4	4.37 [3.81,4.93]	5.11 [4.62,5.61]	4.48 [3.93,5.03]
4×5	5.15 [4.73,5.57]	5.37 [4.94,5.81]	5.11 [4.62,5.61]
5×4	5.11 [4.67,5.55]	5.52 [5.07,5.96]	5.26 [4.85,5.66]
5×5	5.11 [4.64,5.58]	5.70 [5.30,6.11]	5.48 [5.01,5.95]
5×6	5.41 [4.97,5.85]	6.00 [5.56,6.44]	5.56 [5.09,6.03]
6×5	5.56 [5.10,6.01]	5.70 [5.20,6.20]	5.67 [5.18,6.16]
6×6	6.04 [5.67,6.41]	6.15 [5.67,6.62]	5.85 [5.34,6.36]
6×7	6.26 [5.94,6.58]	6.41 [6.11,6.70]	5.81 [5.27,6.36]
7×6	6.41 [6.13,6.68]	6.30 [5.89,6.70]	5.96 [5.52,6.41]
7×7	6.44 [6.08,6.81]	6.22 [5.69,6.75]	6.15 [5.69,6.61]

5.3   Discussion

The objective of Experiment 5 was to identify which grid densities potentially are suitable for use with a low-cost eye-tracking device. Sixteen different grid densities, in between 2×2 and 7×7 cells, were evaluated for use with three visual password formats. The participants (n=27) were asked to make an imaginary password on the screen using a mouse, and imagine they were using eye tracking to make the password. Furthermore, the participants were also required to assess the 16 grid densities about whether they are easy to use and potentially safe for making a visual password in an imagined situation using eye tracking.

The results of Experiment 5 generally showed that when grids became denser, participants thought that a grid was more difficult to use but potentially safer for password formation in an imagined situation using eye tracking. As shown in Figure 5.3, both the linear or logarithmic functions were suitable to estimate the "easy" judgment for any number of grid cells between 4 and 49. If the number of grid cells was 24 cells - by using Equations 5.1 and 5.2, respectively - the linear and logarithmic functions predict that, on average, the "easy" judgment is 4.46 and 4.42 points, respectively. The results from both functions were relatively similar. In the same

manner, however, the logarithmic function was more precise than the linear function to estimate the "safe" judgment relative to the number of grid cells (Figure 5.4).

Furthermore, the participants also think that some grid densities are not particularly suitable for visual password formation using eye tracking. As regards safety judgments, the 95% CIs around judgment means of grid densities with either two rows or columns (2×2, 2×3, and 3×2) did not exceed the scale midpoint ("4") in any of the three visual password formats. Grids with two rows or columns, or both, are thus regarded as relatively unsafe for visual password formation with eye tracking. The results for "easy to use" judgments were less clear-cut. However, since the 95% CIs around judgment means of grid densities with either seven rows or columns (6×7, 7×6, and 7×7) did not firmly exceed the scale midpoint ("4") in two of the three visual password formats (the pattern format and the picture format), these grids can be considered as relatively difficult to use. From this, it results that grid densities ranging from 3×3 to 6×6 are considered suitable for visual password formation using actual eye tracking (Experiment 6).

Experiment 5, however, has several limitations. First, with regard to the grid evaluations, there was no information given to the participants on the meaning of "easy to use" or "potential safety" of the grid densities. Second, it is not known how much time was required to create a password, and actual eye tracking was not used. Therefore, more research needs to be done to investigate whether a particular grid formation and password format are easy to use by measuring the time and success rate of visual password authentication using eye tracking. An indication of the subjective

feeling of safety may be the password length. These matters are further investigated in

Chapter 6 (Experiment 6).

Chapter 6. What kind of grid formations and password formats are useful for password authentication with eye-gaze-based input?

6.1   Experiment 6. An investigation of visual password formats and grid formations with a low-cost eye tracker

6.1.1   Purpose

Research on password systems suggests that recognition-based visual password systems are often considered as easier to memorize, and that systems with a denser grid potentially allow more secure password formation. Furthermore, it has been suggested that eye-gaze-based input could be suitable against password theft ("shoulder surfing"), especially in public spaces. (For more details on past research related to visual password authentication with eye-gaze-based input, see Chapter 2.) However, previous studies have shown that the user still had to press a keyboard to trigger gaze input, and only a single password format was used. Next, most password systems have been tested only for users sitting in front of a personal computer. Finally, comparative research about visual password formats has not yet been performed in a systematic way.

Experiment 6, therefore, aimed to investigate what type of password format and grid formation is suitable for password authentication using eye-gaze-based input. Three recognition-based password formats were used (Figure 6.1). The formats were an alphanumeric format, a pattern format, and a picture format, in which the participant was asked to identify and select a sequence of characters, dots, or icons, respectively,

on the screen by using eye-gaze-based input. The formats were used with 16 grids ranging from 3×3 to 6×6 cells (columns × rows; see Figure 6.3 in the Stimuli section).
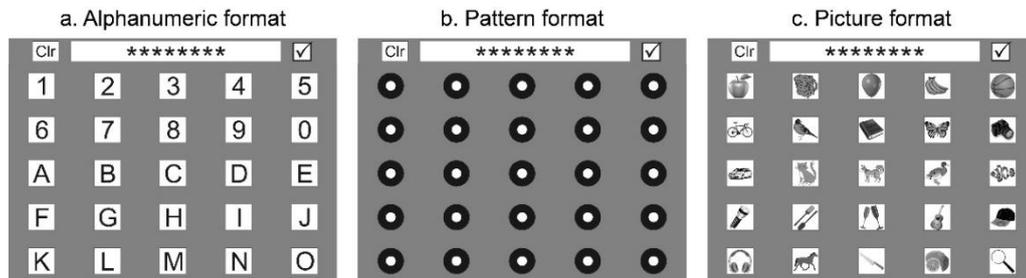
Figure 6.1. Examples of the three recognition-based password formats used in Experiment 6, on a 5×5 grid.

For each format, the participant was asked to perform authentication tasks with a 4-object or 6-object password. Participants performed password registration (Task 1), password confirmation (Task 2), and password login (Task 4). This sequence of tasks is generally performed in any password system and mimics a real situation of password generation. For each of the 16 grid formations, the participant was also asked to judge how easily he/she could perform password input and could recognize a password when authenticating (Task 3, which was performed before password login, Task 4). The participants' task-completion time, their success rate, as well as preference data based on a 7-point rating scale were obtained, in order to test which password format(s) and grid formation(s) would be suitable for eye-gaze-based input.

6.1.2 Method

Participants

Seventeen participants (8 males, 9 females) with normal or corrected-to-normal vision participated. They were 21-44 years of age ($M = 27.1$ years, $SD = \pm 5.9$ years).

The participant's height was in between 157 and 182 cm (M = 167.8 cm, SD = ± 6.7 cm). Fourteen participants were Asian (Japanese, Chinese, and Indonesian), 1 participant was Caucasian, and 2 participants were Latino/Hispanic. The participants were paid for their participation. Data from two participants were not used for statistical analyses. One participant had difficulty employing the eye-tracking system, while the other had difficulty recalling and recognizing the passwords. After each participant had received an explanation and instructions about the experiment, he/she was asked to provide written informed consent (Appendix T) as to his/her participation.

Apparatus

A monitor (Hewlett-Packard LP2065, 20-in, refresh rate 60 Hz) with a resolution of 1600×1200 pixels was used to present the experiment interface (Figure 6.2a). An eye-tracker device (Tobii EyeX©) was mounted on the lower edge of the monitor, at a height of 133 cm from the ground. The angle of the monitor and the eye-tracker were set at two viewing angles of 105 (90+15) and 120 (90+30) degrees. These angles were ideal for participants with a height in between 151-190 cm to register their eye gaze on the eye-tracking system (see Experiment 4, Chapter 4). In order to perform password authentication, the participant was standing in the middle in front of the monitor at a viewing distance of approximately 49 cm, as indicated by a floor mark. This viewing distance is close to the border of the operating distance of the eye-tracker device (for details, see Tobii Eye Tracking Support, 2017), and eye registration was unsuccessful when the participant was standing too close or too far away from the display (see Experiments 1 and 2, Chapter 3). The reason the participant performed

the task while standing was to simulate a situation in which he/she would use eye tracking to register on an ATM-machine with a password.

Figure 6.2. Impression of the experiment set-up of Experiment 6. a. The participant was standing in front of the monitor which showed the experiment interface. b. The experimenter controlled the order of password format, grid, and password length for the participant using another monitor.

Besides the monitor for the experiment interface, another monitor (Lenovo ThinkVision, 20-in, refresh rate 60 Hz) was used by the experimenter to control the order of password format, grid, and password length (Figure 6.2b). Both monitors were mounted on a monitor stand, opposite from each other. All experiment interfaces were programmed in Visual Studio C# (2015), and the data gathered from the participants

were saved in a MySQL database. The experiment was performed under a room lighting condition at an illuminance of 122.35 ± 4.28 lux, as measured using a TOPCON Illuminance Spectro Meter IM-1000 at the location where the participant was standing. The display's luminance was measured using a TOPCON Luminance Meter BM-9. The measurements were performed ten minutes before the start of the experiment.

Stimuli

Sixteen different grid formations were made, ranging from 3×3 to 6×6 cells (columns-by-rows, Figure 6.3). Three password formats depicted in Figure 6.1 were used. The first format was an alphanumeric password format (Figure 6.1a). For this format, alphanumeric characters, i.e., numbers and letters, were presented on a grid. In the case of a grid density of nine cells, i.e., a 3×3 grid, the digits 1, 2, 3, 4, 5, 6, 7, 8, and 9 were presented starting from the top left to the bottom right cell. If the grid density exceeded ten cells, letters (in alphabetical order) were added to the numbers. For example, on a 6×6 grid with 36 cells, the numbers 0 to 9, and the letters A to Z were presented starting from the top left to the bottom right cell. The alphanumeric characters in the grid cells were black with a luminance of 0.14 ± 0.01 cd/m^2, on a white background (2.42 ± 0.09 cd/m^2).

The second format was a pattern format (Figure 6.1b). The pattern format consisted of dots, which could be selected by the participant to create a shape or a pattern as a password. The dots were black (0.14 ± 0.01 cd/m^2) and white (2.40 ± 0.08 cd/m^2). A white dot with a radius of 47 pixels was placed in the middle of a black dot with a radius of 128 pixels, and both dots were presented together as a key. The third

format was a picture format (Figure 6.1c), which consisted of icons in a fixed order on a grid, with the number of icons depending on the grid density (Appendix Q). The icons were in gray-scale with a luminance range of 0.83 to 0.99 cd/m^2, against a white background (2.42 ± 0.11 cd/m^2). Each object key (i.e., alphanumeric character, dot, or icon) of a password was put in the middle of a grid cell with a size of 128×128 pixels, which was 4.16 degrees × 4.47 degrees in visual angle, and every pixel within an object key was 0.028 degrees × 0.030 degrees in visual angle.
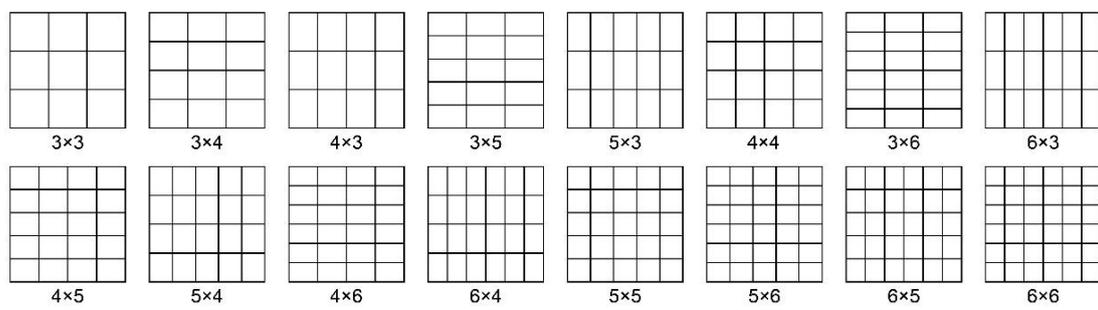
Figure 6.3. A schematic impression of the 16 different grid formations used in Experiment 6. Grid formations of 3×3 to 6×6 cells (columns × rows) were used. Note that the object keys (i.e., alphanumeric characters, dots, or icons) had the same size regardless of the number of grid cells.

The participant had four tasks (see Procedure section below), for which two screen interfaces were made with a size of 1600×1200 pixels set against a gray background (0.90 ± 0.03 cd/m^2) to perform password authentication (Appendix U). In Tasks 1, 2, and 4, the participant used his/her eye gaze to enter a password. For Tasks 1 and 2, the upper part of the screen (1600×125 pixels) consisted of two text boxes, a "Save" key, a "Confirm" key, and a "Clr" key. The "Save" key could be used by the participant to save a password. When gazing at the "Confirm" key, the participant could confirm password input, while the "Clr" was used to clear his/her registered or

confirmed input. The main part of the screen (1600×1075 pixels) displayed the password formats and grids. For Task 4, the upper screen of the task interface (1600×125 pixels) displayed a text box, a "Login" key, and a "Clr" key. The "Login" key could be used by the participant to authenticate his/her password into the system, while the "Clr" was used to clear this. Also in Task 4, the password formats and grids were displayed on the main part of the screen (1600×1075 pixels). When the participant selected an object key on the grid on the main part of the screen, an asterisk would be displayed on the text box at the upper part of the screen, and a chime sound would be played (1538 ms; Appendix R) to indicate that a selection was made. All (object) keys on the upper or main part of the screens could be triggered by eye gaze with a dwell time of 500 ms.

Procedure

The participant was asked to stand in the middle in front of the monitor without crossing a floor mark. While standing, he/she was asked to relax, take a natural viewing position, and make no excessive head movements during the experiment. Following this, the participant was shown a password on the screen, randomly generated for each of the three password formats, consisting either of 4 or 6 objects. He/she was then asked to memorize the password within a minute for a 4-object password and within two minutes for a 6-object password (Appendix V). After memorizing, the participant was instructed to perform the four tasks as described below.

Task 1: Password registration

The participant was instructed to register the memorized password on the screen interface by using his/her eye gaze. The participant could select the appropriate

object keys displayed on a grid that was randomly selected from the 16 different grid formations. The password consisted either of alphanumeric characters (alphanumeric format), dots (pattern format), or icons (picture format). After registration, the participant was instructed to select a "Save" key.

Task 2: Password confirmation

After saving the password, on the same screen, the participant was asked to confirm the password by re-selecting the same object keys on the same grid. Following this, the participant was instructed to select a "Confirm" key. In case the confirmation was incorrect, for example, due to incorrect memorization or incorrect selection of object keys, he/she could retry the confirmation up to five attempts. If the participant failed to confirm the password on the fifth attempt, he/she was instructed to register again (Task 1) using a different password for the same password format and grid.

Task 3: Grid evaluation

After confirming a password, the participant was asked to judge whether he/she considered the grid that was used in Task 1 and 2 as easy to use for password registration and confirmation. This judgment was made on a scale between 1 (not easy) and 7 (very easy). Next, the participant was instructed to evaluate whether the password (4 or 6 objects) was easy to remember. This was also done on a rating scale between 1 (not easy) and 7 (very easy). Since this task did not require eye-gaze-based input, the participant used a mouse to make the rating-scale judgments on the screen. The meaning of "easy to use" was defined as how fast (estimated time needed) and successful (the number of attempts) the participant was in registering and confirming

the password. "Easy to remember" was described as how much effort the participant thought to be necessary to memorize and recall a password.

Task 4: Password login

In this task, the participant was asked to log in into the system with the password that he/she had registered and confirmed before. If the participant noticed an error during login, he/she could retry to enter the password up to five attempts. If the login failed at the fifth attempt, the participant was instructed to register again (Task 1), starting by memorizing a different password for the same password format and grid. After the participant had finished all tasks for each grid formation for the three formats, he/she was asked to fill in a final questionnaire about his/her experience in daily life with passwords in general (Appendix W). The participant was explicitly instructed not to reveal any password or password formation strategy that he/she used in daily life.

The experiment was performed with counterbalance in the order of the three password formats. That is, five participants first performed the tasks in the alphanumeric format, then in the pattern format, and finally in the picture format, for each of the 16 grids. Another five participants started with the pattern format, followed by the picture format, and ended with the alphanumeric format. The remaining five participants started with the picture format and ended with the pattern format. For each format, the order of password length (4 or 6 objects) was varied as well. The time needed and the number of attempts needed by the participant to perform Tasks 1, 2, and 4 were recorded by means of the computer program. During the experiment, the

participant was not informed about this in order to ensure his/her natural attitude towards the tasks.

Before the start of the experiment, each participant needed to register his/her eyes and perform calibration with Tobii EyeX$^©$ software at one of the viewing angles. In order to get familiar with all tasks, a practice program was performed in which the participant practiced Tasks 1, 2, and 4 with a 4-object or a 6-object password, twice for each password format, on a grid randomly chosen from the 16 grids. In between Tasks 2 and 4, Task 3 was practiced as well. The experiment took about 6 hours, divided over 2-hour sessions for 3 days. The procedure was approved by the Ethical Committee of the Faculty of Design, Kyushu University, Japan (131-3).

6.2   Results of Experiment 6

In Tasks 1, 2, and 4, task-completion time and input success rate were obtained. In Task 3, preference data based on a rating scale for the password formats and grid formations were gathered. Data from 4320 trials (15 participants $\times$ 16 grids $\times$ 3 tasks $\times$ 3 password formats $\times$ 2 password lengths) were collected. In Tasks 1, 2 and 4, in which the participant entered the password using his/her eye gaze, 11% (458/4320) of the time measurements were disproportionally slow, i.e., they were outliers in a positive direction. Given the dwell time for eye tracking of 500 ms per object key, disproportionally fast times were not obtained. The Median Absolute Deviation$_n$ method (MAD_n) was used to remove outliers (Leys, Ley, Klein, Bernard, & Licata, 2013). That is, data points that were 2.5 times the MAD_n above the median were removed recursively until no additional outliers were identified.

6.2.1 Task-completion time difference between password formats

The time needed by participants (n = 15) to perform Tasks 1, 2, and 4 (see Procedure section) for 16 grids in three password formats was measured. From here on, this will be called "task-completion time". Since the data were not normally distributed, as confirmed with a Shapiro-Wilk test, non-parametric Friedman tests were performed in order to see whether task-completion time for all 16 grid formations varied with the password format. If significant, pairwise comparisons using Wilcoxon signed-rank tests were performed to see which pair(s) of password formats showed a significant difference with the alpha level was Bonferroni-adjusted to 0.05/3 = 0.017. Figure 6.4 shows the differences in median task-completion time (s) between password formats with 4-object or 6-object passwords for all grids.

The statistical details (Table 6.1, Appendix X) regarding task-completion time are as follows. In Task 1 (password registration), task-completion time over grid density ($df = 2$, n = 16) differed between password formats for both the 4-object ($\chi^2 = 14.00$, $p = 0.001$) and 6-object ($\chi^2 = 16.63$, $p < 0.001$) passwords. Follow-up pairwise comparisons revealed that the time to complete Task 1 with passwords in the alphanumeric format did not differ from task-completion time with passwords in the pattern format (4-object passwords: $Z = -0.63$, $p = 0.532$; 6-object passwords: $Z = -1.60$, $p = 0.109$). Completion time in the picture format, however, took significantly longer than in the alphanumeric format (4-object passwords: $Z = -3.31$, $p = 0.001$; 6-object passwords: $Z = -3.46$, $p = 0.001$) and the pattern format (4-object passwords: $Z = -3.36$, $p = 0.001$; 6-object passwords: $Z = -3.05$, $p = 0.002$).
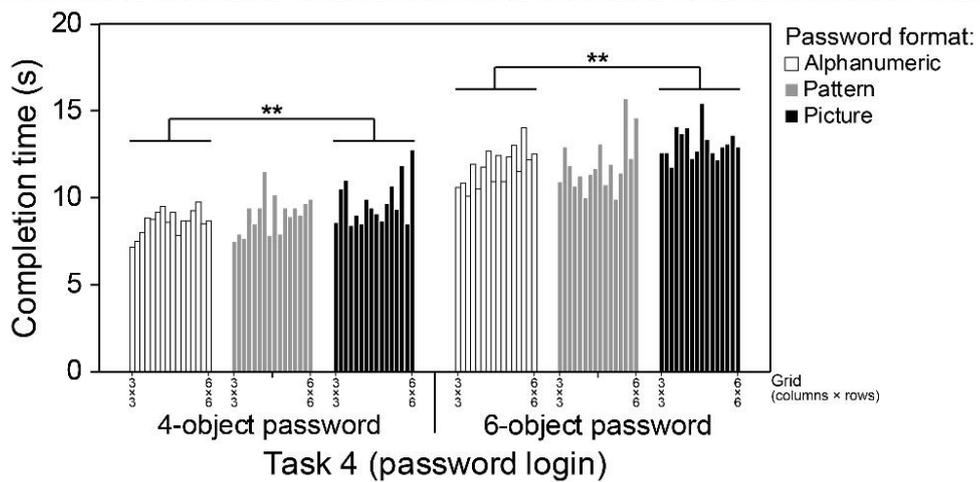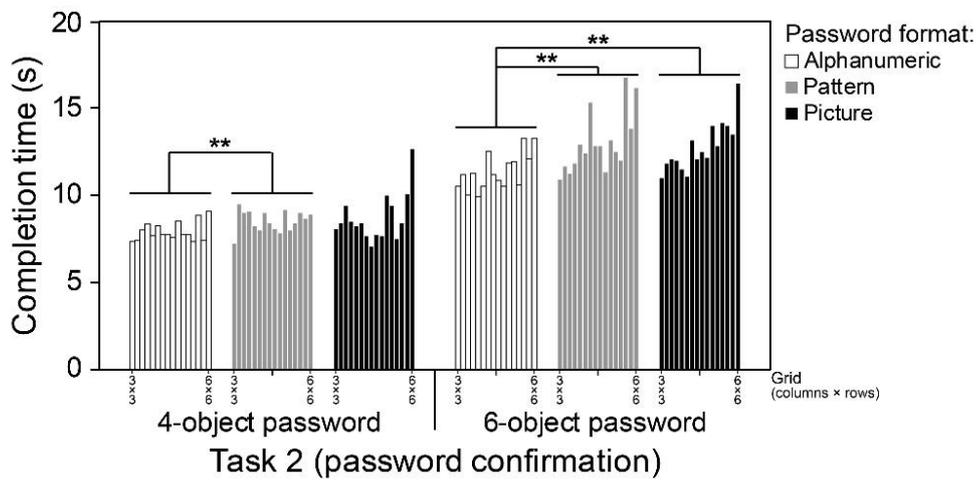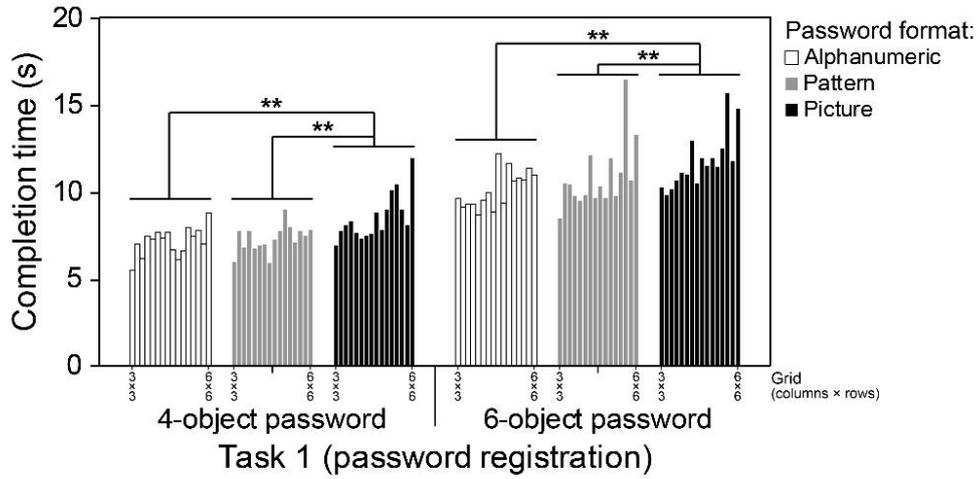
Figure 6.4. Results of Experiment 6. Median task-completion time (s) for alphanumeric, pattern, and picture password formats with 4-object or 6-object keys for 16 grid formations in Task 1 (top), Task 2 (middle), and Task 4 (bottom). An asterisk shows a significant difference in task-completion time between password formats ($p < 0.01$).

For the data obtained in Task 2 (password confirmation), the Friedman test showed significant differences between password formats for both the 4-object ($\chi^2 = 6.13$, $p = 0.047$) and 6-object ($\chi^2 = 24.00$, $p < 0.001$) passwords. Pairwise comparisons revealed that the completion time for passwords in the picture format did not differ from that for the pattern format (4-object passwords: $Z = -0.16$, $p = 0.88$; 6-object passwords: $Z = -0.47$, $p = 0.642$) and for the alphanumeric format (4-object passwords: $Z = -2.22$, $p = 0.026$, which was not significant with Bonferroni correction on the alpha level). However, completion time for alphanumeric passwords was significantly shorter than that for pattern passwords (4-object passwords: $Z = -2.69$, $p = 0.007$; 6-object passwords: $Z = -3.52$, $p < 0.001$) and picture passwords (6-object passwords: $Z = -3.52$, $p < 0.001$).

Completion time of Task 4 (password login) significantly differed between password formats for 4-object ($\chi^2 = 6.50$, $p = 0.039$) and 6-object ($\chi^2 = 7.88$, $p = 0.019$) passwords. The pairwise comparisons showed that completion time for passwords in the pattern format neither differed from that in the alphanumeric format (4-objects passwords: $Z = -1.86$, $p = 0.063$; 6-objects passwords: $Z = -0.52$, $p = 0.605$) nor from completion time in the picture format (4-objects passwords: $Z = -1.50$, $p = 0.134$; 6-objects passwords: $Z = -2.22$, $p = 0.026$, which was also not significant with Bonferroni correction on the alpha level). Completion time for picture passwords, however, was significantly longer than that for alphanumeric passwords (4-objects passwords: $Z = -2.72$, $p = 0.007$, 6-objects passwords: $Z = -2.95$, $p = 0.003$).

Table 6.1. Results of Experiment 6. Pairwise comparisons of task-completion time between password formats for 4-object and 6-object passwords in Task 1, Task 2, and Task 4.

Password length	Pairs of password formats	Task 1	Task 2	Task 4
		Z	Z	Z
4-object passwords	AN > PA	-0.63	**-2.69****	-1.86
	AN > PI	**-3.31****	-2.22	**-2.72****
	PA > PI	**-3.36****	-0.16	-1.50
6-object passwords	AN > PA	-1.60	**-3.52*****	-0.55
	AN > PI	**-3.46****	**-3.52*****	**-2.95****
	PA > PI	**-3.05****	-0.47	-2.22

AN: Alphanumeric format, PA: Pattern format, PI: Picture format
Task 1: password registration, Task 2: password confirmation, Task 4: password login
Z: Wilcoxon signed rank test value
>: faster task-completion time
** $p < 0.01$, *** $p < 0.001$ (after Bonferroni-correction)

6.2.2   The relation between task-completion time and grid density

In Tasks 1, 2 and 4, 16 data points for task-completion time were obtained for each of the three password formats. One data point was acquired for each square grid formation with a grid density of 9 (3×3), 16 (4×4), 25 (5×5), or 36 (6×6) cells. Two data points were obtained for the grids with an equal number of cells, yet each with a horizontal formation (more columns than rows) or a vertical formation (more rows than columns). Two data points were thus obtained for grids with 12 cells (3×4 and 4×3), 15 cells (3×5 and 5×3), 18 cells (3×6 and 6×3), 20 cells (4×5 and 5×4), 24 cells (4×6 and 6×4), and 30 cells (5×6 and 6×5). As shown in Figure 6.4, there was a general trend that participants took more time to input passwords in all three password formats when the grid density became higher. Pearson's correlation analyses were performed over the median of these 16 data points to examine the relation between task-completion time and grid density for 4-object and 6-object passwords in Tasks 1, 2,

and 4 (Appendix Y). The *r*-values are shown in Figure 6.5. First, in general, the participants needed more time to complete 6-object passwords than 4-object passwords for all three formats in all tasks. Second, the results indeed clearly showed that task-completion time increased for denser grids, i.e., grids consisting of more object keys.
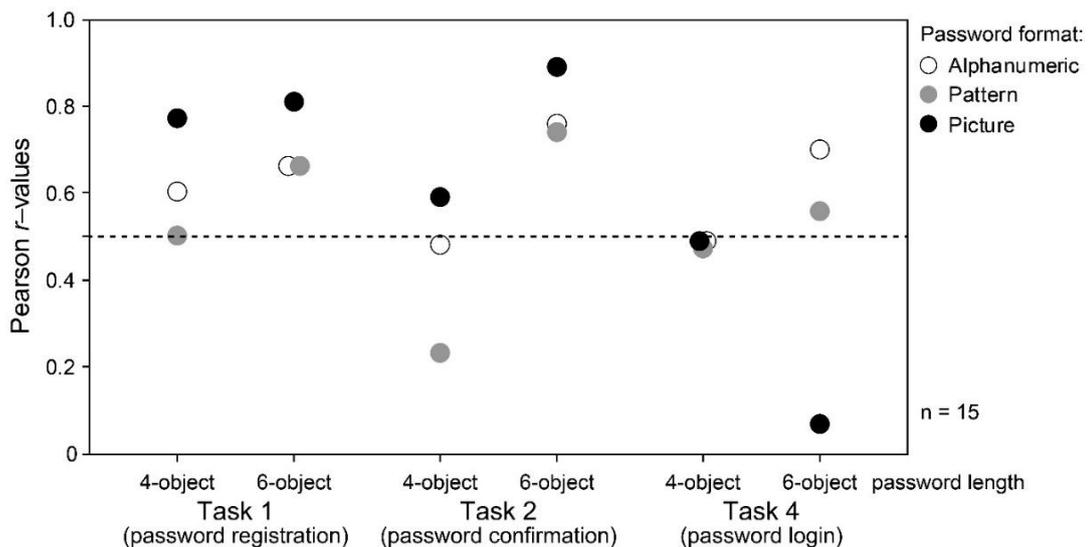
Figure 6.5. Results of Experiment 6. Pearson's correlation values between task-completion time and grid density. Task-completion time was obtained for eye-gaze-based input of 4-object and 6-object passwords in three password formats in Task 1, Task 2, and Task 4. The circles show Pearson *r*-values for the alphanumeric passwords (white), pattern passwords (gray), and picture passwords (black). All Pearson *r*-values higher than 0.50 (dashed line) show a significant positive correlation ($p < 0.05$) between task-completion time and grid density.

The statistical details (Table 6.2a, Appendix Y) are as follows. In Task 1 (password registration), for 4-object passwords, the median completion time ranged in between 5.58-11.96 seconds(s). Pearson's correlation analyses showed that the participants significantly required more completion time with increasing grid density. For the alphanumeric format ($r = 0.60$, n = 15, $p = 0.014$) and the picture format ($r = 0.77$, n = 15, $p < 0.001$) this correlation was significant. For the pattern format, the

correlation bordered on significance ($r = 0.50$, n = 15, $p = 0.050$). As the grid density increased, the participants also significantly needed more time for 6-object passwords, with a median completion time in between 8.55-16.44 s. For the alphanumeric format ($r = 0.66$, n = 15, $p = 0.005$), the pattern format ($r = 0.66$, n = 15, $p = 0.005$), and the picture format ($r = 0.81$, n = 15, $p < 0.001$) this correlation was significant.

Similar to Task 1, the median completion time in Task 2 (password confirmation) increased when the number of grid cells increased. For 4-object passwords, it grew from 7.10 to 12.69 s. The correlation between task-completion time and grid density was significant for the picture format ($r = 0.59$, n = 15, $p = 0.016$), but not for the pattern format ($r = 0.23$, n = 15, $p = 0.395$) and the alphanumeric format ($r = 0.48$, n = 15, $p = 0.061$), although the correlation for the latter bordered on significance. The median task-completion time for 6-object passwords ranged in between 9.97-16.86 s. The correlation between task-completion time and grid density was significant for the alphanumeric format ($r = 0.76$, n = 15, $p < 0.001$), the pattern format ($r = 0.74$, n = 15, $p < 0.001$), and the picture format ($r = 0.89$, n = 15, $p < 0.001$).

Also in Task 4 (password login), there was a general tendency that participants needed more time to enter 4-object passwords when the number of grid cells increased (median completion time from 7.17-12.82 s). However, the correlations between task-completion time and grid density for the alphanumeric format ($r = 0.50$, n = 15, $p = 0.054$), the pattern format ($r = 0.47$, n = 15, $p = 0.064$), and the picture format ($r = 0.49$, n = 15, $p = 0.053$) were not significant, yet all bordered on significance. For 6-object passwords, the median task-completion time ranged from 10.00-15.77 s. There was a statistically significant correlation between task-completion time and grid

density for the alphanumeric format ($r = 0.70$, n = 15, $p = 0.002$) and the pattern format ($r = 0.56$, n = 15, $p = 0.024$), but not for the picture format ($r = 0.07$, n = 15, $p = 0.798$).

6.2.3 The relation between task-success rate and grid density

The password input success rate was measured based on whether the participant could perform Tasks 2 and 4 at the first attempt. For Tasks 2 and 4, data from 2880 trials in total were obtained (15 participants $\times$ 16 grids $\times$ 2 tasks $\times$ 3 password formats $\times$ 2 password lengths). Most of the trials (91%, 2627/2880) were completed at the first attempt with 4-object or 6-object passwords for all grids and password formats. I examined the correlation between task-success rate and grid density for these data. The results showed a negative correlation: when the grid became denser, the number of participants who successfully entered the password with eye-gaze-based input at the first attempt decreased (Appendix Z). Figure 6.6 shows the Pearson r-values for the relation between first-attempt-success rate and grid density in Tasks 2 and 4 for the 4-object and 6-object passwords performed for the three password formats.

The statistical details (Table 6.2b, Appendix Z) are as follows. In Task 2, the first-attempt-success rate with 4-object passwords decreased when the number of grid cells increased for all three formats. Although there was no statistically significant correlation between first-attempt-success rate and grid density for the picture format ($r = -0.40$, n = 15, $p = 0.123$), for the alphanumeric format ($r = -0.67$, n = 15, $p = 0.005$) and the pattern format ($r = -0.64$, n = 15, $p = 0.007$) this negative correlation was significant. For 6-object passwords, the first-attempt-success rate also significantly decreased as grid density increased for the alphanumeric format ($r = -0.69$, n = 15, p

= 0.003), the pattern format (r = -0.76, n = 15, p = 0.001), and the picture format (r = -0.66, n = 15, p = 0.005).
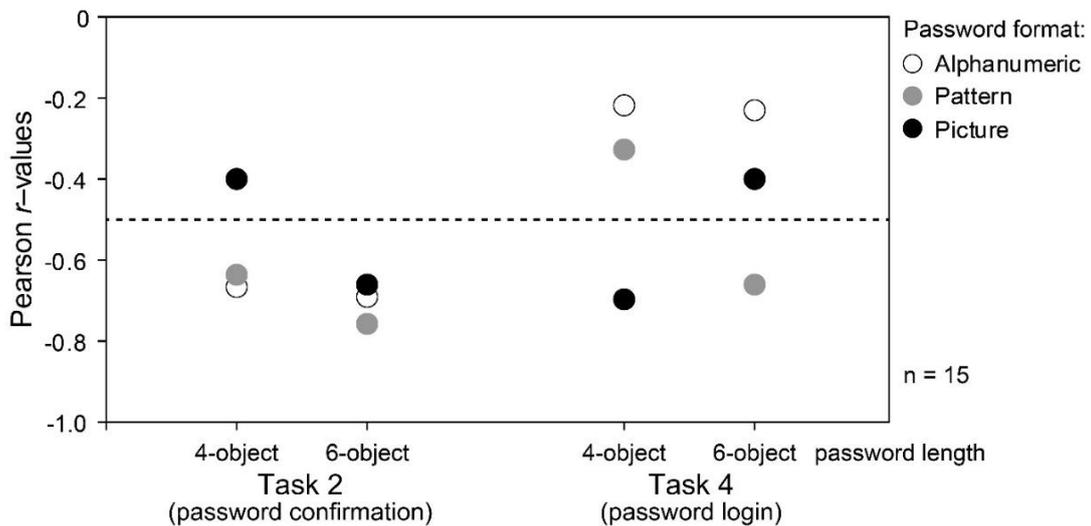
Figure 6.6. Results of Experiment 6. Pearson's correlation values between successful password input at the first attempt and grid density. The first-attempt-success rate was obtained for eye-gaze-based input of 4-object or 6-object passwords in three password formats in Task 2 and Task 4. The circles show Pearson r-values for alphanumeric passwords (white), pattern passwords (gray), and picture passwords (black). All Pearson r-values lower than -0.50 (dashed line) show a significant negative correlation ($p < 0.01$).

In Task 4, the first-attempt-success rate for 4-object passwords also significantly decreased when the number of grid cells increased in the picture format (r = -0.70, n = 15, p = 0.002), but not in the alphanumeric format (r = -0.22, n = 15, p = 0.420) and the pattern format (r = -0.33, n = 15, p = 0.219). For 6-object passwords, first-attempt-success rate showed no significant correlation with grid density for the alphanumeric format (r = -0.23, n = 15, p = 0.392) and the picture format (r = -0.40, n = 15, p = 0.125). There was a statistically significant negative correlation, however,

between first-attempt-success rate and grid density for the pattern format ($r = -0.66$, n
= 15, $p = 0.006$).

Table 6.2. Results of Experiment 6. (a) Correlations (Pearson *r*-values) between task-completion time and grid density in Task 1, Task 2, and Task 4, and (b) correlations (Pearson *r*-values) between the first-attempt-success rate and grid density in Task 2, and Task 4, for 4-object and 6-object passwords in three password formats.

	a. Pearson *r*-values between the task-completion time and grid density					
	4-object passwords			6-object passwords		
Task	Alphanumeric Format	Pattern Format	Picture Format	Alphanumeric Format	Pattern Format	Picture Format
1	**0.60***	0.50	**0.77****	**0.66****	**0.66****	**0.81*****
2	0.48	0.23	**0.59***	**0.76 ****	**0.74****	**0.89*****
4	0.50	0.47	0.49	**0.70****	**0.56***	0.07

	b. Pearson *r*-values between the first-attempt-success rate and grid density					
	4-object passwords			6-object passwords		
Task	Alphanumeric Format	Pattern Format	Picture Format	Alphanumeric Format	Pattern Format	Picture Format
2	**-0.67****	**-0.64****	-0.40	**-0.69****	**-0.76****	**-0.66****
4	-0.22	-0.33	**-0.70****	-0.23	**-0.66****	-0.40

Task 1: password registration, Task 2: password confirmation, Task 4: password login
* p < 0.05, ** p < 0.01, *** p < 0.001

6.2.4 Task-completion time difference between horizontal and vertical grid configurations

In cases of grids with an equal number of cells (i.e., an equal number of object keys), I also checked whether the formation of the grid influenced the task-completion time. Leaving the square grids aside, I directly compared task-completion time for grids with more columns than rows (e.g., columns × rows = 4×3) against grids with more rows than columns (e.g., columns × rows = 3×4). For each task, six pairs of grid

formations were compared using data points with no outliers for each of the three password formats. I called grids with more columns than rows "horizontal" configurations, and grids with more rows than columns "vertical" configurations. Table 6.3 shows the differences in average task-completion time (s) between horizontal and vertical grid configurations for 4-object and 6-object passwords for the three password formats.

For the alphanumeric format in Task 1 (password registration), the average task-completion time for 4-object passwords was nearly similar between horizontal and vertical configurations of 12, 15, 18, 20, 24, and 30 cells (object keys). For the pattern and the picture formats, the time was also nearly similar between horizontal and vertical configurations in between 12 - 30 object keys. No statistically significant difference between horizontal and vertical pairs was found for any of the three password formats. For 6-object passwords, however, pairwise comparisons revealed six significant differences. The first two concerned the alphanumeric format, where vertical configurations of 3×5 and 4×5 (columns × rows) grids required a longer completion time than horizontal configurations of 5×3 and 5×4 grids (Z = -2.20, p = 0.028 and Z = -2.13, p = 0.033, respectively). The next two significant differences were found in the pattern format between vertical 3×6 and 5×6 grids and horizontal 6×3 and 6×5 grids (Z = -2.76, p = 0.006 and Z = -2.73, p = 0.006, respectively). In the picture format, significant differences occurred in the same grid comparisons, i.e., between vertical 3×6 and 5×6 grids and horizontal 6×3 and 6×5 grids (Z = -2.55, p = 0.011 and Z = -2.27, p = 0.023, respectively).

Four significant differences between horizontal and vertical grid configurations with equal grid density were found in the completion time of Task 2 (password confirmation). First, the time to confirm a 4-object alphanumeric password was longer in the vertical 5×6 grid than in the horizontal 6×5 grid (Z = -2.73, p = 0.006). A significant difference also appeared in the task-completion time of the 6-object alphanumeric password (Z = -2.12, p = 0.034) between these grid configurations. In the pattern format, vertical configurations of 3×6 and 5×6 (columns × rows) grids required a longer task-completion time than horizontal configurations of 6×3 and 6×5 (columns × rows) grids (Z = -2.48, p = 0.013 and Z = -2.38, p = 0.017, respectively).

In Task 4 (password login), the eye-gaze-based input also required more time in vertical than in horizontal grid configurations, significantly in four cases. First, the time to complete a 4-object pattern password took longer in the vertical 3×6 grid than in the horizontal 6×3 grid (Z = -2.70, p = 0.007). A significant difference was also found between the vertical 4×5 and the horizontal 5×4 configuration for 4-object pattern passwords (Z = -2.76, p = 0.006). The third significant difference was between the vertical 5×6 and the horizontal 6×5 configuration for 4-object picture passwords (Z = -2.29, p = 0.022). The last significant difference concerned the alphanumeric format, for which completion time of a 6-object password differed between the vertical 3×5 and the horizontal 5×3 configuration (Z = -2.67, p = 0.008).

Table 6.3. Results of Experiment 6. The differences in task-completion time (s) between horizontal and vertical grid configurations for 4-object and 6-object passwords in the three password formats in Task 1, Task 2, and Task 4.

Task	Password format	#Grid cells	Grid density (columns × rows)	4-object password			6-object password		
				N	Mean (SD)	Z	N	Mean (SD)	Z
1	Alphanumeric	12	3×4	12	6.45 (1.01)	-0.47	12	9.02 (1.85)	-0.47
			4×3		6.16 (0.82)			8.69 (1.27)	
		15	3×5	13	7.06 (0.89)	-0.31	12	9.47 (1.20)	**-2.20***
			5×3		7.07 (1.92)			8.47 (1.00)	
		18	3×6	13	7.28 (0.87)	-0.18	12	9.36 (1.02)	-1.33
			6×3		7.44 (1.65)			8.95 (0.87)	
		20	4×5	10	6.35 (1.15)	-0.36	11	10.75 (2.03)	**-2.13***
			5×4		6.60 (1.44)			9.24 (0.44)	
		24	4×6	13	6.90 (0.83)	-1.57	12	10.62 (1.61)	-0.71
			6×4		8.10 (2.01)			10.13 (1.31)	
		30	5×6	12	7.87 (1.26)	-1.26	14	11.11 (2.52)	-0.03
			6×5		7.28 (1.77)			11.39 (3.43)	
	Pattern	12	3×4	13	7.17 (1.36)	-1.22	13	9.58 (2.19)	-0.59
			4×3		6.59 (1.40)			10.48 (2.80)	
		15	3×5	15	7.47 (1.92)	-1.02	12	9.91 (3.17)	-0.47
			5×3		6.91 (1.44)			9.76 (1.85)	
		18	3×6	13	7.61 (2.86)	-1.85	13	12.07 (3.14)	**-2.76****
			6×3		6.16 (0.96)			9.00 (1.02)	
		20	4×5	15	8.06 (2.83)	-0.80	14	10.02 (2.64)	-0.03
			5×4		7.43 (1.74)			9.82 (2.14)	
		24	4×6	14	8.89 (3.25	-1.10	14	11.78 (3.85)	-1.29
			6×4		7.94 (1.82)			10.42 (2.54)	
		30	5×6	14	7.51 (2.45)	-0.47	14	17.80 (5.46)	**-2.73****
			6×5		8.02 (2.92)			11.01 (3.97)	
	Picture	12	3×4	12	6.93 (1.46)	-1.10	11	9.69 (1.18)	-0.27
			4×3		7.85 (1.46)			10.06 (1.21)	
		15	3×5	13	8.63 (1.98)	-0.73	13	11.40 (3.43)	-0.25
			5×3		7.90 (1.94)			11.47 (2.74)	
		18	3×6	11	7.30 (1.65)	-0.53	9	12.72 (2.86)	**-2.55***
			6×3		7.37 (1.16)			10.12 (0.68)	
		20	4×5	13	8.31 (2.07)	-0.59	12	10.58 (1.92)	-1.41
			5×4		7.91 (1.78)			11.98 (4.74)	
		24	4×6	11	8.37 (1.60)	-0.27	12	10.49 (2.12)	-1.57
			6×4		8.58 (2.39)			11.03 (1.90)	
		30	5×6	14	9.16 (2.55)	-0.47	13	14.77 (4.55)	**-2.27***
			6×5		9.03 (2.68)			12.19 (3.65)	

		12	3×4	10	6.99 (1.02)	-0.15	13	11.21 (1.98)	-1.64
			4×3		7.05 (1.43)			10.07 (1.43)	
		15	3×5	14	8.40 (1.55)	-0.85	12	10.95 (1.82)	-1.18
			5×3		7.78 (1.86)			10.26 (2.30)	
		18	3×6	12	7.53 (1.03)	-0.16	9	10.92 (1.82)	-0.77
	Alphanumeric		6×3		7.54 (1.14)			10.02 (1.56)	
		20	4×5	10	7.22 (1.43)	-1.17	11	10.63 (2.35)	-0.27
			5×4		8.07 (1.77)			10.00 (1.52)	
		24	4×6	9	7.35 (1.12)	-1.13	13	11.56 (2.37)	-0.04
			6×4		7.67 (0.74)			11.83 (2.58)	
		30	5×6	15	10.12 (3.91)	**-2.73****	12	13.02 (3.43)	**-2.12***
			6×5		8.21 (1.92)			11.01 (2.76)	

		12	3×4	14	9.24 (2.80)	-0.22	11	11.08 (3.45)	-0.62
			4×3		8.62 (2.52)			10.91 (2.21)	
		15	3×5	13	8.14 (2.07)	-0.87	13	11.47 (3.25)	-1.22
			5×3		8.17 (1.85)			13.10 (3.23)	
		18	3×6	12	10.59 (5.15)	-1.88	14	15.37 (5.38)	**-2.48***
2	Pattern		6×3		7.88 (1.72)			12.31 (3.70)	
		20	4×5	10	7.20 (1.29)	-1.07	10	11.72 (4.10)	-0.26
			5×4		7.62 (1.42)			12.43 (4.15)	
		24	4×6	12	9.29 (2.87)	-0.86	13	12.57 (3.84)	-0.11
			6×4		8.45 (2.58)			12.52 (2.93)	
		30	5×6	10	8.68 (2.97)	-0.76	8	14.22 (3.48)	**-2.38***
			6×5		7.82 (1.86)			10.52 (2.76)	

		12	3×4	13	8.05 (1.25)	-1.29	15	11.98 (3.16)	-0.57
			4×3		9.07 (2.07)			12.39 (3.33)	
		15	3×5	13	8.04 (1.65)	-0.38	11	11.62 (2.72)	-0.89
			5×3		8.46 (2.17)			11.24 (2.28)	
		18	3×6	12	8.08 (2.18)	-1.73	14	14.59 (5.05)	-1.54
	Picture		6×3		6.84 (1.13)			12.46 (3.04)	
		20	4×5	14	8.26 (1.78)	-0.22	12	11.82 (1.79)	-0.86
			5×4		8.48 (2.12)			11.18 (2.58)	
		24	4×6	12	9.78 (3.06)	-0.16	11	13.92 (5.17)	-0.71
			6×4		9.75 (2.41)			11.64 (2.18)	
		30	5×6	13	8.35 (1.80)	-1.22	14	13.63 (3.86)	-0.47
			6×5		9.51 (2.40)			13.43 (3.48)	

		12	3×4	14	7.44 (1.45)	-1.10	11	10.58 (1.78)	-1.07
			4×3		8.27 (1.81)			9.67 (1.59)	
		15	3×5	10	8.18 (1.35)	-0.15	12	12.50 (3.03)	**-2.67****
4	Alphanumeric		5×3		8.91 (3.04)			9.92 (1.59)	
		18	3×6	12	8.74 (1.85)	-0.70	12	12.63 (3.84)	-1.88
			6×3		8.34 (2.09)			10.57 (1.89)	
		20	4×5	10	8.51 (3.24)	-0.87	11	11.80 (3.49)	-0.45

110

		Grid	N	Mean (SD)	Z	N	Mean (SD)	Z
		5×4		7.71 (1.61)			10.60 (1.84)	
	24	4×6	11	8.70 (2.78)	-0.09	13	11.71 (2.41)	-0.8
		6×4		8.44 (1.36)			12.43 (3.70)	
	30	5×6	9	9.72 (2.70)	-1.72	12	12.00 (2.98)	-0.24
		6×5		8.05 (1.51)			11.62 (2.51)	
Pattern	12	3×4	15	8.18 (2.27)	-1.19	13	12.54 (4.92)	-0.87
		4×3		8.90 (3.17)			11.40 (2.98)	
	15	3×5	13	8.75 (2.04)	-0.25	14	11.84 (4.36)	-0.35
		5×3		8.54 (2.15)			11.33 (2.20)	
	18	3×6	10	10.61 (3.26)	**-2.70****	11	12.16 (4.90)	-0.36
		6×3		7.74 (1.66)			10.58 (2.01)	
	20	4×5	11	10.11 (3.44)	**-2.76****	13	13.51 (5.96)	-0.66
		5×4		7.65 (2.01)			11.34 (2.55)	
	24	4×6	10	8.16 (1.55)	-0.56	11	11.82 (4.43)	-1.25
		6×4		8.26 (1.64)			10.14 (2.27)	
	30	5×6	14	9.57 (3.42)	-0.09	10	13.40 (5.19)	-0.56
		6×5		9.76 (3.54)			11.34 (2.63)	
Picture	12	3×4	14	9.37 (2.61)	-0.79	13	12.45 (3.10)	-1.36
		4×3		10.10 (2.45)			11.44 (1.85)	
	15	3×5	13	8.35 (1.16)	-0.66	11	11.91 (3.02)	-1.51
		5×3		8.84 (1.90)			12.74 (2.92)	
	18	3×6	13	10.12 (3.28)	-0.31	12	12.06 (2.06)	-0.55
		6×3		9.95 (3.42)			12.80 (4.04)	
	20	4×5	11	9.51 (2.48)	-0.98	12	15.53 (4.81)	-0.55
		5×4		8.42 (1.78)			15.43 (7.22)	
	24	4×6	13	9.64 (2.90)	-1.36	11	12.40 (3.13)	-1.25
		6×4		10.96 (3.18)			11.10 (2.27)	
	30	5×6	10	11.95 (6.10)	**-2.29***	12	13.15 (4.06)	-0.78
		6×5		8.10 (1.49)			12.83 (2.41)	

N: number of participants, SD: Standard Deviation, Z: Wilcoxon signed rank test value, * = p < 0.05, ** = p < 0.01.

6.2.5 Participant judgments

In Task 3, participant judgments were obtained about the grid densities and formations (Appendix AA). The participants judged the usability of each grid based on how fast (time) and successful (first attempt) they could register and confirm a password with eye-gaze-based input. They were also asked to judge how well they

could recall and recognize a password on each grid density. The participants made judgments on a 7-point rating scale between 1 (not easy) and 7 (very easy). Regarding "easy-to-use" judgments, the participants judged the grid as increasingly less easy to use when the number of object keys increased, either with a 4-object or a 6-object password, for all three password formats. Pearson's correlation analyses showed a statistically significant correlation between the participant judgments and grid density, with r-values ranging from -0.81 to -0.96 (n = 15, $p < 0.001$). Regarding "easy-to-remember" judgments, the participants judged the password as less easy to remember when the number of grid keys increased, for both 4-object and 6-object passwords in all three password formats. Pearson's correlation analyses showed a statistically significant correlation between the participant judgments and grid density, with r-values ranging from -0.61 to -0.96 (n = 15, $p < 0.02$).

6.3   Discussion

In Experiment 6, participants were asked to memorize a 4-object and a 6-object password for three types of password formats and register (Task 1), confirm (Task 2), and log in (Task 4) the password on a grid by using eye-gaze-based input. The three recognition-based password formats were an alphanumeric format, a pattern format, and a picture format (Figure 1). Grid densities and formations were varied in 16 ways in between 3×3 and 6×6 object keys (Figure 2). Task-completion time and success rate data was obtained. Participants also provided preference data about the grid densities and formations (Task 3).

The first purpose of Experiment 6 was to investigate which type of password format is suitable for password authentication using eye-gaze-based input. Experiment 6 showed that for 16 grids, password input with 4-object or 6-object keys required more time in the picture and pattern formats than in the alphanumeric format. In the majority of cases, task-completion time in the alphanumeric format was significantly faster (Figure 6.4). Participants are most likely more familiar with passwords consisting of numbers and letters in daily life, and memorization of alphanumeric passwords by "chunking" (grouping) may have enabled faster recall (Nelson & Vu, 2010). In general, more frequently used items are easier to recall (Kinsbourne & George, 1974) and possibly in the present experiment the participants had not much time to adapt to using icons (picture format) or dots (pattern format). The preference for the alphanumeric format was also reflected in the questionnaire taken after the experiment (Appendix AB), which showed that 12 participants (80%) thought that the alphanumeric password format would be potentially suitable to use with eye-gaze-based input. Only three participants (20%) thought that the picture format could be useful, while none considered the pattern format useful.

The second purpose of Experiment 6 was to investigate what kind of grid formation is useful for password authentication using eye-gaze-based input. The results showed that the participants generally needed more time to complete password registration (Task 1), confirmation (Task 2), and login (Task 4) on denser grids with more object keys, either with a 4-object or a 6-object password, in the three password formats. The majority of the correlations (Pearson r-values) between task-completion time and grid density was significant (Figure 6.5). Previous research on eye

movements already had reported that participants needed less search time for sparse layouts than for dense layouts (Halverson & Hornof, 2004). It is thus likely that the participants needed more time to search the necessary object keys to form the password as the total number of key options increased.

Another possible explanation for the fact that participants needed more time to make the password on a denser grid is the increased chance of incorrect object key selection with eye tracking. Although each object key had the same size regardless of grid density, incorrect key selection might have happened because the distance between object keys narrowed, causing the participant to sometimes unintendedly gaze on an incorrect object key, for example when the screen appeared for the first time. The participant rating scale judgments also showed that they considered a grid as significantly more difficult to use with eye-gaze-based input when the grid became denser.

Another demerit of denser grids found here is that the number of successful password inputs at the first attempt, either for 4-object or 6-object passwords, decreased when the grid became denser. Over half of the correlations (Pearson *r*-values) between first-attempt-success rate and grid density was significant (Figure 6.6). As the number of grid keys increased, the participants thus tended to make more mistakes, i.e., they selected objects incorrectly and needed more attempts. One reason could be that they more often incorrectly gazed at the wrong object key due to the grid density, as described above. Another reason is that with increasing grid density, the passwords became more complex. The passwords used by the participants were randomly generated according to grid density. For example, a password on an

alphanumeric 3×3 grid consisted only of digits, while a password on an alphanumeric 6×6 grid consisted of digits and letters. The combination of the latter might have been more difficult to remember. This result related to the participant judgments, which showed that passwords were judged as significantly less easy to recall and recognize when the grid became denser. Future research is necessary to clarify this issue further.

The last finding related to grid formation is that the time needed to enter a password with eye-gaze-based input was often longer for grids with more rows than columns (vertical configurations) than for grids with more columns than rows (horizontal configurations), under equal grid density. Direct paired comparisons of task-completion time between horizontal and vertical formations with an equal number of grid keys revealed 14 significantly different pairs. In all 14 cases, task-completion time was significantly faster in horizontal grids than in vertical grids (Table 6.3). This strongly suggests that entering a password with eye-gaze-based input is faster on horizontal grids with more columns than rows. Vertical grids with more rows than columns are less efficient for eye-gaze-based input. Studies on the visual search of objects or words have reported similar results. When searching for visual objects on a screen, the direction of the participants' eye movements may occur more frequently horizontally than vertically (Duchowski, 2007). It has further been shown that fixation time in visual search of vertical word lists is longer than fixation time for horizontal word lists (Ojanpää, Näsänen, & Kojo, 2002). A horizontal search model was also preferred for searching a target word in a full-screen search field (Goonetilleke, Lau, & Shih, 2002).

In Experiment 6, as mentioned by some participants, some issues need to be remedied. One limitation is that the "Save", "Clr", and "Confirm" key at the upper part of the interface screen for Tasks 1 and 2 were relatively close together (Appendix U) and that some gaze time needed to be spent on selecting the correct key for these actions. Another issue that requires investigation is dwell time. Here, 500-ms dwell time is used to confirm gaze on a certain object key. It is worthwhile to investigate whether a shorter dwell time can be used since this would speed up the password input process with eye tracking. In Chapter 7 (Experiment 7), these issues are further fixed and investigated.

Chapter 7.   Usability of various dwell times for eye-gaze-based object selection with eye tracking

7.1   Experiment 7. Object selection based on dwell time with a low-cost eye tracker

7.1.1   Purpose

One limitation of Experiment 6 was that only a single dwell time of 500 ms was used. Dwell time is the amount of time a user needs to focus his/her gaze at an object on a display before it is selected. However, the user still needs some practice to select an object using dwell time. For example, on a display with multiple objects, the user must first reliably identify the target object, before performing an action on it. Accordingly, when eye gaze is used to identify a target object, the user may unintentionally and inattentively dwell his/her gaze on the wrong object. As a result, this object may even become selected as the target object – a problem that is known as the Midas-Touch problem (for details on object selection based on dwell time, see Chapter 2). Therefore, interactive interfaces typically use a fixed duration of dwell time for eye-gaze-based object selection. Previous studies have shown that dwell time was fixed at one single value and was not based on user preferences. Furthermore, comparative research on dwell times for eye-gaze-based object selection has not yet been performed.

The purpose of this experiment was to assess the usefulness of various dwell times for selecting three types of objects on a display with eye-gaze-based input. In

short, twelve participants were asked to memorize a sequence of 4 or 6 objects, and to use their eye gaze to enter the sequence onto a user interface with a dwell time of 250, 500, 1000, and 2000 ms per object. The object selection task was performed on four grid formations, consisting of 3×4, 4×3, 4×5, and 5×4 cells (columns-by-rows, see Figure 6.2 in the Stimuli section). Three types of objects were used (Figure 7.1). The first type of objects were alphanumeric characters, similar to those used in the eye-typing task in most previous research with a dwell time (e.g., Majaranta et al., 2003; Bee & André, 2008; Kurauchi et al., 2016). Sequences consisting of patterns of dots and visual icons were also employed. These types of visual objects are commonly used in recognition-based password systems (Biddle et al., 2012), for example to manually unlock smartphones. Besides object selection with eye-gaze-based input, the participants were also asked to evaluate, on a 7-point rating scale, how easily they could perform object selection with each of the four dwell times. The total time necessary to select the correct sequence of objects, the number of object selection corrections, and dwell time evaluations were recorded.
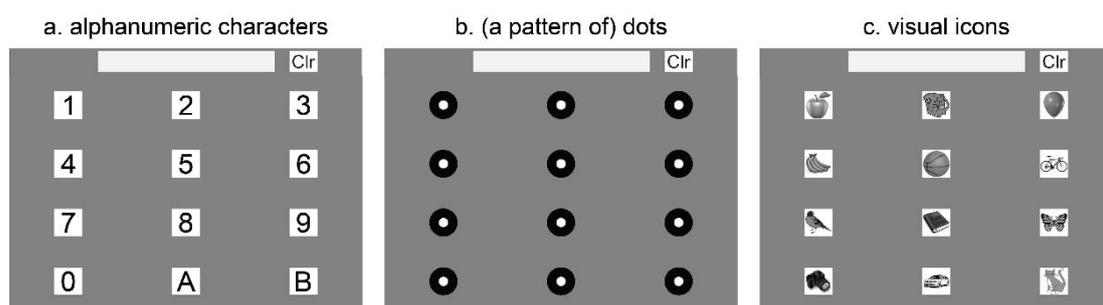
Figure 7.1. Visual (password) objects used in Experiment 7. Examples of the three types of objects that needed to be selected with eye-gaze-based input in this study. A sequence of 4 or 6 objects needed to be selected from four different grids, with four different dwell times.

7.1.2 Method

Participants

The participants were 12 students (5 males and 7 females) of Kyushu University, Japan, who had normal or corrected-to-normal vision. The age of the participants was in between 21 and 45 years ($M = 28.1$ years, $SD = \pm 7.4$ years). Nine participants were Asian (Japanese, Chinese, or Indonesian), 1 participant was Caucasian, and 2 participants were Latino/Hispanic. The average height of the participants was 166.3 cm ($SD = \pm 6.7$ cm). The participants were asked to provide written informed consent as to their participation, after they had received an explanation and instructions about the experiment (Appendix AC). The participants were paid for their participation.

Apparatus

Two monitors (20-in, refresh rate 60 Hz) with a resolution of 1600×1200 pixels were utilized in this experiment. The first monitor was a Hewlett-Packard LP2065, which was used to present the experiment interface. A Tobii Eye Tracker 4C$^{©}$ device was placed on the lower edge of the monitor, at a height of 133 cm from the ground. The monitor and the eye-tracking device were tilted upwards to two viewing angles of 105 and 120 degrees. These angles were suitable for participants in between 151-190 cm in height to register their eye gaze on a very similar eye-tracking system (Experiment 4, Chapter 4). The second monitor (Lenovo ThinkVision) was used by the experimenter as an interface to control the order of dwell time, grid, and the sequence of 4 or 6 objects that needed to be selected. Both monitors were mounted on

a monitor stand, opposite from each other. All experiment interfaces were programmed in Visual Studio C# (2015). Experimental results were saved in a MySQL database. The experiment was performed under room lighting at an illuminance of 124.14 ± 8.23 lux, measured using a TOPCON Illuminance Spectro Meter IM-1000 at the participant's viewing position. Visual object luminance (see below) was measured using a TOPCON Luminance Meter BM-9.

Stimuli

A sequence of 4 or 6 visual objects needed to be selected with eye-gaze-based input, from four different grid formations. These grids consisted of 3×4, 4×3, 4×5, and 5×4 cells (columns-by-rows, Figure 7.2). The number of objects on the display was the same as the grid density, i.e., 12 objects in 3×4 and 4×3 grids, and 20 objects in 4×5 and 5×4 grids. Three types of visual objects were used. The first type were alphanumeric characters consisting of numbers and letters (Figure 7.1a). They were presented in alphabetical order from the top left to the bottom right grid cell. For each cell, the numbers and letters were black with a luminance of 0.17 ± 0.01 cd/m^2 on a white background (3.10 ± 0.17 cd/m^2).
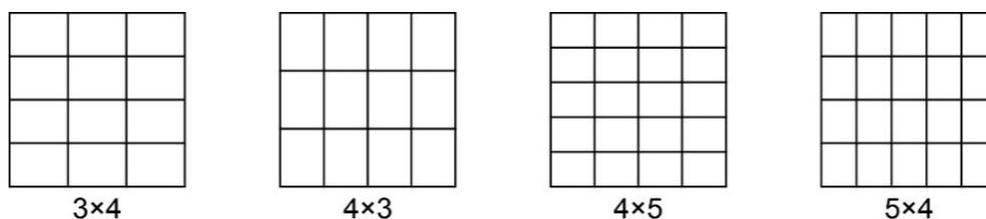
Figure 7.2. The four different grid formations [3×4, 4×3, 4×5, and 5×4 cells (columns × rows)] used in Experiment 7. Note that regardless of the number of grid cells, the visual objects that needed to be selected with eye-gaze-based input (i.e., alphanumeric characters, dots, or icons) had the same size.

The second type of objects that needed to be selected were dots, which could be sequentially selected to create a shape or a pattern (Figure 7.1b). Dot patterns are often used as screen locks on smartphones. A single dot on the display was composed of a white circle (3.12 $\pm$ 0.21 cd/m^2) with a radius of 47 pixels, which was placed in the middle of a black circle (0.18 $\pm$ 0.01 cd/m^2) with a radius of 128 pixels. The third type of objects were visual icons (Figure 7.1c). The visual icons were placed in a fixed order on one of the four grids, in gray-scale with a luminance range of 0.07 to 1.18 cd/m^2, against a white background (3.07 $\pm$ 0.13 cd/m^2). Typical examples of visual icons are depictions of fruits or animals (Appendix Q). For all three types, an object (i.e., alphanumeric character, dot, or icon) was centered in the middle of a grid cell with a size of 128$\times$128 pixels, i.e., 4.16 degrees $\times$ 4.47 degrees in visual angle, and every pixel within an object was 0.028 degrees $\times$ 0.030 degrees in visual angle.

To select a sequence of visual objects (see Procedure section below), a screen interface was made with a size of 1600$\times$1200 pixels, set against a gray background (1.24 $\pm$ 0.05 cd/m^2). In the middle of the upper part of the screen (1600$\times$125 pixels) was a text box (800$\times$100 pixels), and at the top right was a "Clr" key (128$\times$100 pixels). The "Clr" key could be used by the participant to clear incorrectly selected objects one by one. The main part of the screen (1600$\times$1075 pixels) displayed the grids and object types (Appendix AD). When the participant selected an object on the grid on the main part of the screen, an asterisk would appear in the text box at the upper part of the screen, and a chime sound would be played (1538 ms; Appendix R) to indicate that an object was selected. All objects on the upper or main part of the screen could be selected by eye gaze under each of the four dwell time durations.

When selecting objects with eye-gaze-based input, the participant was standing in the middle in front of the screen at a viewing distance of approximately 49 cm, as indicated by a floor mark. Practically, this viewing distance was close to the minimum operating distance of the eye-tracking device (Tobii Eye Tracking Support, 2017), as confirmed in Experiments 1 and 2, Chapter 3. The reason the participant performed the task while standing was to simulate a situation in which he/she would use eye tracking to enter an object sequence, such as a password, on an automated teller machine (ATM).

The object selection task went as follows. First, the participant was shown a 4-object or a 6-object sequence, randomly generated for each of the three object types, on a grid that was randomly selected from the four different grids. Thus, the objects that were presented on the screen consisted of a sequence of alphanumeric characters, dots, or visual icons (Appendix AE). The participant was then asked to memorize a sequence of 4 objects within a minute and a sequence of 6 objects within two minutes. The appropriate grid was displayed on the screen to assist the participant in memorizing the objects' positions within the grid.

After memorizing, the participant was instructed to enter the memorized sequence onto the screen interface by selecting the appropriate objects either with a dwell time of 250, 500, 1000, or 2000 ms. The participant was instructed to use a "Clr" key if he/she had selected an incorrect object. In case the object selection was incorrect, for example, due to the Midas-Touch problem or to selecting objects in the wrong order, the sequence could be attempted up to five times. If the participant failed to

enter the correct objects on the fifth attempt, he/she was instructed to try again using a different sequence for the same dwell time, object type and grid.

After entering the correct objects, the participant was asked to evaluate whether he/she considered the dwell time as easy to use for object selection with eye-gaze-based input. The evaluation was made on a scale between 1 (not easy) and 7 (very easy). The participant used a mouse to make the rating-scale judgments on the screen. The meaning of "easy to use" was defined as how fast (estimated time needed) and with how few corrections the participant thought he/she had entered the objects.

Next to this subjective dwell time evaluation by the participant, objective measurements of the total object selection time (for 4 objects or 6 objects) were obtained by means of the computer program. The number of object selection corrections for each combination of object type and dwell time was calculated from 240 overall attempts (12 participants $\times$ 4 grids $\times$ 5 attempts). In order to ensure a natural attitude towards the task, the participant was not informed about these time measurements before the experiment.

The experiment was performed with counterbalance in the order of the four dwell time durations. That is, three participants first selected objects with the dwell time of 250 ms, then with the dwell time of 500, 1000 ms, and finally with the dwell time of 2000 ms, for each of the four grids. Next, three participants started with the dwell time of 500 ms, followed by the dwell time of 1000, 2000 ms, and ended with the dwell time of 250 ms. Another three participants started with 1000 ms, followed by 2000, 250 ms, and ended with the dwell time of 500 ms. The remaining three participants started with the dwell time of 2000 ms and ended with the dwell time of

1000 ms. In the same way, the order of object type was also counterbalanced within every four participants. The number of objects in a sequence (4 or 6) was varied as well, for each object type and dwell time.

Before the start of the experiment, the participant needed to register his/her eyes and perform calibration with Tobii Eye Tracker 4C$^{©}$ software. In order to get used to the task, a practice program was prepared in which the participant practiced object selection for a given object type and sequence, on a grid randomly chosen from the four grids. The evaluation task was practiced as well. The experiment took about 1 hour and 30 minutes. The procedure was approved by the Ethical Committee of the Faculty of Design, Kyushu University, Japan (131-3).

7.2   Results of Experiment 7

The total time necessary to select the correct sequence of 4 or 6 objects will be called 'object selection time' from hereon. For every participant, object selection time, the number of object selection corrections, and the dwell-time evaluation data obtained with the rating scale were recorded.

7.2.1   Object selection time

Object selection time for 4- or 6-object sequences was obtained in 576 trials in total (12 participants $\times$ 4 dwell time durations $\times$ 4 grids $\times$ 3 object types). The Median Absolute Deviation$_n$ method (MAD$_n$) was used to detect and remove outliers in the data. A removal criterion of 2.5 times the MAD$_n$ above the median was used recursively until no additional outliers were identified (Leys et al., 2013). By using this method, 9 out of 576 trials (2%) were removed for 4-object sequences. No trials

were removed for 6-object sequences. Shapiro-Wilk tests showed that the object selection time data obtained under each of the four dwell time durations were not normally distributed ($p < 0.05$), either with 4-object or 6-object sequences. To compare object selection time between object types and between grid formations, non-parametric Kruskal-Wallis tests were therefore performed for 4-object sequences, since the data were unpaired after outlier removal (9 trials). Data for 6-object sequences were subjected to non-parametric Friedman tests (Appendix AF).

The results (Figure 7.3) showed that for both 4-object and 6-object sequences, object selection time – in case of correct object selection – did not significantly differ between object types under each of the four dwell time durations. For 4-object sequences, object selection time also did not differ significantly between grid formations under each dwell time duration. For 6-object sequences, however, object selection time significantly differed between grid formations for a dwell time of 500 ms for alphanumeric characters ($df = 3$, $\chi^2 = 13.40$, $p < 0.004$), and for a dwell time of 2000 ms for visual icons ($df = 3$, $\chi^2 = 9.90$, $p < 0.019$). For example, when using a 500-ms dwell time, participants needed a little more time to select alphanumeric objects on a 5×4 grid than on 4×3 and 3×4 grids, respectively ($p = 0.015$, $p = 0.034$). However, after Holm-Bonferroni correction for multiple comparisons based on ranks, these differences did not pass the significance level. Also for selecting icon objects with a dwell time of 2000 ms, differences were not significant after Holm-Bonferroni correction. Although robust significant differences have been found in a relatively large range of 3×3 until 6×6 grids (see Experiment 6, Chapter 6), object selection time in this experiment thus did not vary widely among the four grid formations. Given this,

object selection time was averaged over grid formations for each of the three object types for further statistical analysis.
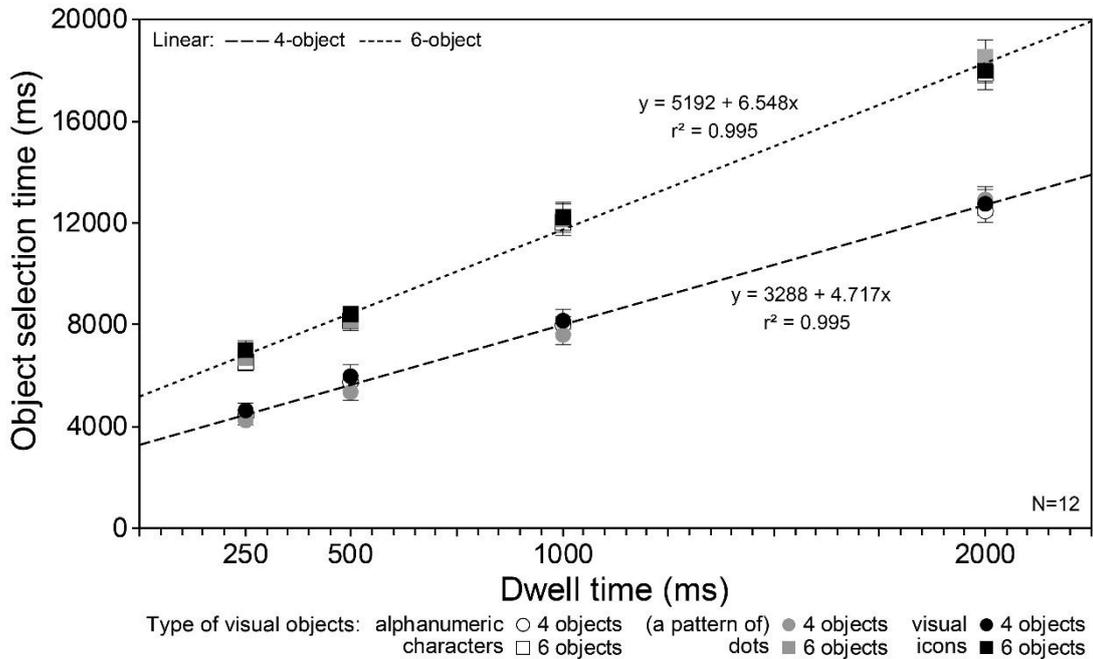
Figure 7.3. Results of Experiment 7. The relation between dwell time per object and the total selection time for 4- or 6-object sequences in milliseconds (ms) without object correction. Eye-gaze-based selection time for 4-object (circles) and 6-object (squares) sequences of three types of objects was obtained with dwell time durations of 250, 500, 1000, and 2000 ms per object. The continuous lines show a linear function for 4-object and 6-object sequences through the four dwell time durations in the three object types. Error bars indicate ± 95% confidence intervals around the means.

The results also showed that object selection time for 6-object sequences was systematically longer than for 4-object sequences without object correction, as indicated by the 95% confidence intervals. Furthermore, as expected, for all three types of visual objects, the resulting grand averages showed that 4-object and 6-object selection time linearly increased with dwell time, as shown in Figure 7.3. A regression analysis with a linear function (similar to Equation 5.1 as in Chapter 5) was performed

in order to examine the correlation between object selection time (4 and 6 objects) and dwell time duration. The reason why the linear function used here was to estimate object selection time for any duration among the four dwell times. In this case, x is the dwell time duration (from 250 to 2000 ms), and y is the object selection time. The regression equations for 4-object sequences ($y = 3288 + 4.717x$) and 6-object sequences ($y = 5192 + 6.548x$) show that the average object selection time respectively increased ($r^2 = 0.995$; $p < 0.001$) with each increase in dwell time duration (Appendix AG).

The time needed to search an object in the display under each dwell time duration can be expressed with the following equation:

$$ST = \frac{OT - (DT \times NPO)}{NPO} \tag{7.1}$$

where ST is object search time (ms), OT represents object selection time (ms), DT is dwell time (250, 500, 1000, 2000 ms), and NPO indicates the number of objects in a sequence (4 or 6).

Using Equation 7.1, the data show that the participants needed about 1000 ms to search a single target object on the display (see Table 7.1), regardless of object type and sequence length (4 or 6 objects). For 4-object alphanumeric, dot, and icon sequences, the average search time per object (ST) was 976, 952, and 1042 ms, respectively. For 6-object alphanumeric, dot, and icon sequences, the average object search time (ST) was 923, 965, and 965 ms, respectively.

Table 7.1. Results of Experiment 7. Object selection time (OT) and object search time (ST) for 4-object and 6-object sequences with eye-gaze-based-input. OT and ST were obtained for each of three visual object types, under dwell time durations of 250, 500, 1000, and 2000 ms per object.

Number of objects in a sequence	Object type	Dwell Time (ms)								Average object search time	SD
		250		500		1000		2000			
		OT	ST	OT	ST	OT	ST	OT	ST		
4 objects	Alphanumeric characters	4398	849	5749	937	7985	996	12481	1120	976	114
	(a pattern of) Dots	4295	824	5386	846	7605	901	12946	1236	952	192
	Visual icons	4649	912	6012	1003	8217	1054	12795	1199	1042	120
6 objects	Alphanumeric characters	6557	843	8180	863	12025	1004	17892	982	923	82
	(a pattern of) Dots	6751	875	8198	866	12150	1024	18573	1095	965	113
	Visual icons	7036	923	8414	903	12212	1035	17990	998	965	63

OT: object selection time (ms); ST: object search time (ms); SD: standard deviation (ms).

7.2.2 Number of object selection corrections

The number of object selection corrections for each dwell time duration and object type was obtained from 240 overall attempts (12 participants × 4 grids × 5 attempts). Data showed that all participants had accomplished the object selection tasks before the fifth and final attempt. Table 7.2 shows that with a dwell time of 250 ms, the number of object selection corrections was relatively high, especially for dot and icon objects. However, in all conditions the number of object selection corrections strongly decreased – often to zero – as dwell time duration increased. Because of the relatively low number of object selection corrections overall, further statistical analyses were not performed.

Table 7.2. Results of Experiment 7. The number and percentages of object selection corrections for 4-object and 6-object sequences in three object types, under dwell time durations of 250, 500, 1000, and 2000 ms.

Number of objects in a sequence	4 objects				6 objects			
Dwell Time (ms)	250	500	1000	2000	250	500	1000	2000
Alphanumeric characters	3 (1%)	0 (0%)	1 (0%)	0 (0%)	10 (4%)	0 (0%)	0 (0%)	0 (0%)
(a pattern of) Dots	16 (7%)	0 (0%)	0 (0%)	1 (0%)	43 (18%)	9 (4%)	6 (3%)	4 (2%)
Visual icons	14 (6%)	2 (1%)	0 (0%)	0 (0%)	25 (10%)	5 (2%)	0 (0%)	0 (0%)

Note: Percentages (%) are based on 240 overall attempts (n=12 × 4 grids × 5 attempts).

7.2.3 Dwell time evaluations

The participants assessed the usability of each dwell time on a rating scale. Participants were instructed to base their evaluations on their subjective impression of how fast (estimated time needed) and with how few corrections they thought they had entered the object sequence. Evaluations were made on a 7-point rating scale between 1 (not easy) and 7 (very easy). A dwell time of 250 ms was judged as less easy to use for dot and visual icon objects. As the dwell time duration increased, participants considered 500-ms dwell time as easier to use for all three object types. However, when the dwell time duration equal to or higher than 1000 ms, participants thought a dwell time generally not easy to use for object selection of two or three object types.

In order to explore the obtained data further, some further analyses were performed. First, a regression analysis with a quadratic function (7.2) was performed to examine the relation between the "easy" ratings and dwell time duration and also to figure out the optimum dwell time for each object type.

$$y = ax^2 + bx + c \tag{7.2}$$

where x is the dwell time duration (from 250 to 2000 ms), y is the rating for dwell time evaluation (from 1 to 7), a, and b are the coefficients of the equation with $a \neq 0$, while c is a constant number.

$$y = -0.000001182x^2 + 0.001x + 5.338 \tag{7.3}$$

$$y = -0.000001745x^2 + 0.003x + 3.428 \tag{7.4}$$

$$y = -0.000004529x^2 + 0.009x + 1.698 \tag{7.5}$$

The regression equations 7.3 to 7.5, respectively, for alphanumeric, dot and icon objects, with r^2 values of 0.828, 0.801, 0.862 ($p > 0.05$), respectively, suggest that the continuous curves did not fit very well to the "easy" rating data in all cases (Appendix AH). Furthermore, these equations also could not estimate the maximum points of the usability of the dwell time for each object type, especially around the 500-ms dwell time, as shown in Figure 7.4. The reason could be that only four dwell times were investigated and the distance between dwell time durations was wide, particularly between 1000 and 2000 ms. Moreover, the data also showed that participants did not consider any duration is suitable for selecting various object types after 1000-ms dwell time.
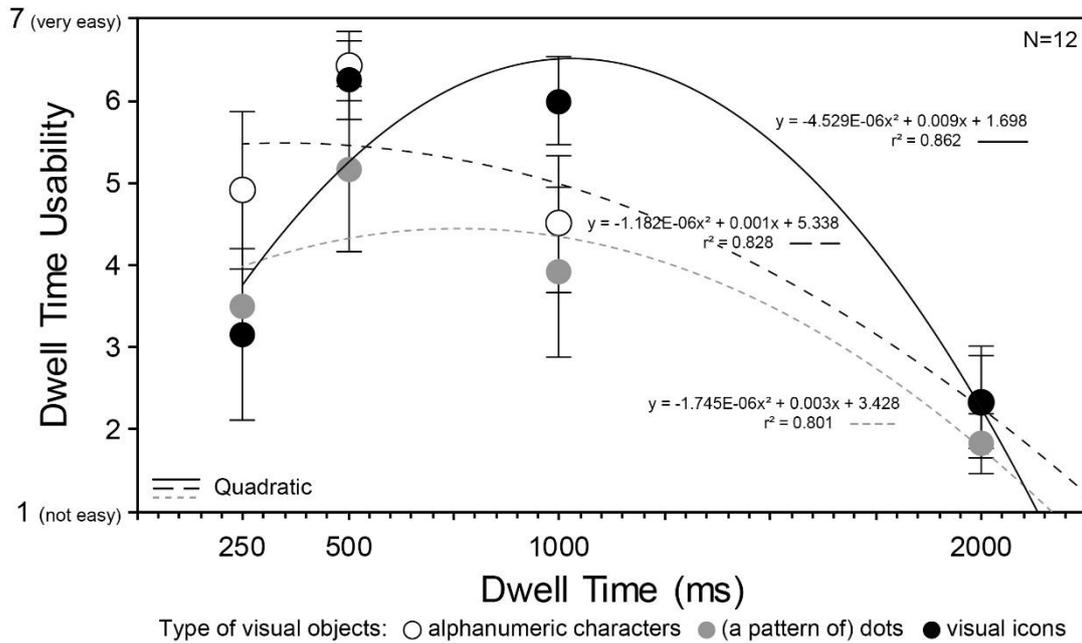
Figure 7.4. Results of Experiment 7. Average evaluations of the usability of each dwell time duration (ms) for eye-gaze-based selection of three object types. The continuous curves show a quadratic function for each object type through the four dwell time durations – Equations 7.3 to 7.5. Error bars indicate ± 95% confidence intervals around the means.

Hence, the quadratic function (Equation 7.2) was used again in order to determine the maximum points for each object type by means of the first three dwell time durations, i.e., 250, 500, and 1000 ms. Based on the quadratic function, the maximum points for each object type through the first three dwell time durations were obtained by using the following equations:

$$y = 0.000013x^2 + 0.01584x + 1.78 \tag{7.6}$$

$$y = -0.00001224x^2 + 0.01586x + 0.3 \tag{7.7}$$

$$y = -0.000017093x^2 + 0.02514x + 2.0466 \tag{7.8}$$

The regression equations 7.6 to 7.8, respectively, for alphanumeric, dot, and icon objects with $r^2 = 1.00$, respectively, were a better fit to the data than equations 7.3 to 7.5, respectively, when one does not regard the 2000-ms dwell time data (Appendix AH). According to these equations, the optimum estimated dwell time for alphanumeric objects is 600 ms (Eq. 7.6), for dot objects it is 650 ms (Eq. 7.7), and for visual icons it is 725 ms (Eq. 7.8).
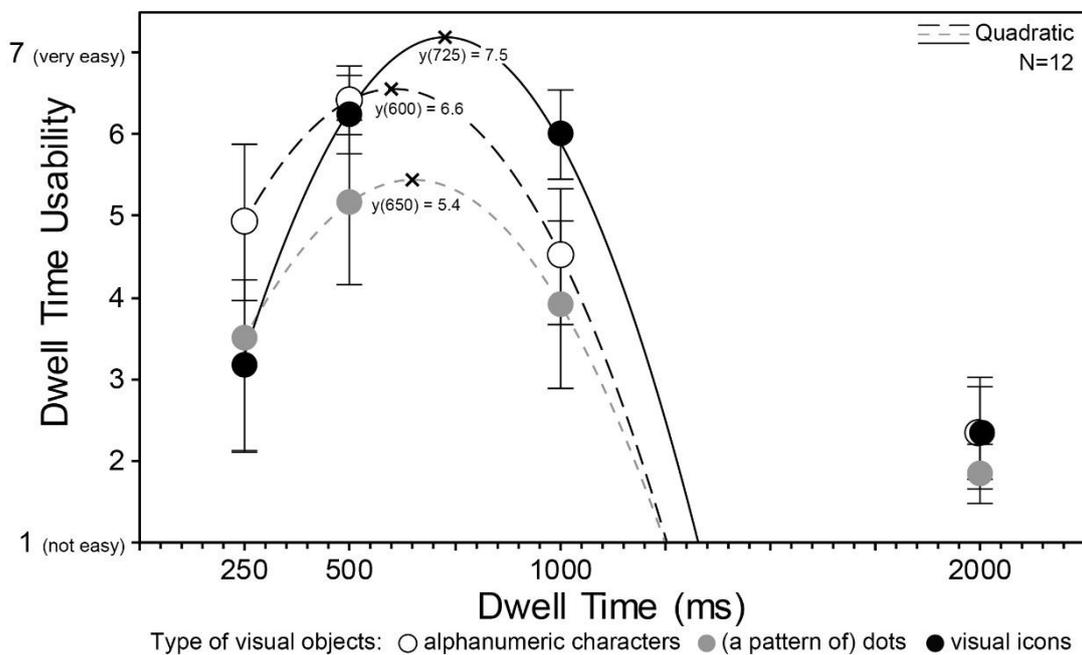
Figure 7.5. Results of Experiment 7. The continuous curves show a quadratic function for each maximum point – Equations 7.6 to 7.8. The crosses show the estimated maximum points of dwell time usability ratings for each object type through the first three dwell time durations of 250, 500, and 1000 ms. Error bars indicate ± 95% confidence intervals around the means.

However, the quadratic function and the equations as shown in Figure 7.5 still were not the best-fitting to estimate the data trend between the "easy" ratings and dwell time duration for each object type, because the 2000-ms dwell time was not included. In order to fit the data better, it appears that a right-skewed parabolic function would

be a better candidate. Therefore, a more complex function likely can be used to show a continuous trend for right-skewed data points, e.g., the Linex loss function (Zieliński, 2005). This function is usually used to draw asymmetric curves through data points in order to make predictions from the data – see Appendix AI, for brief information about this function. However, it is obvious from the data that the optimum dwell times estimated with this loss function will be rather similar to those shown in Figure 7.5.

With regard to the statistical analysis, since the data were not normally distributed, as confirmed by Shapiro-Wilk tests ($p < 0.05$, for four dwell time durations in three object types), non-parametric Friedman tests were performed over the data. If significant, pairwise comparisons using Wilcoxon signed-rank tests with Holm-Bonferroni-correction on the alpha-level based on ranks [0.05/(m-k+1), where m = 6 indicates the number of pairs and k = 1, 2, …, m, indicates the level of ranks] were performed to see which pair(s) of dwell time evaluations were significantly different. Figure 7.4 shows the dwell time evaluations for each object type.

The statistical details regarding the object dwell time evaluations (also see Table 7.3, Appendix AJ) are as follows. For alphanumeric characters, the Friedman test ($df = 3$, n = 12) showed significantly different evaluations between dwell time durations ($\chi^2 = 25.23$, $p < 0.001$). Follow-up pairwise comparisons revealed five significantly different pairs. First, object selection with a dwell time of 500 ms was assessed as significantly easier than with a dwell time of 250, 1000, and 2000 ms ($Z = -2.31$, $p = 0.021$, $Z = -3.10$, $p = 0.002$, $Z = -3.13$, $p = 0.002$, respectively). Furthermore, object selection with dwell times of 250 and 1000 ms was considered significantly

easier than with a dwell time of 2000 ms ($Z = $ -2.66, $p = $ 0.008, $Z = $ -3.09, $p = $ 0.002, respectively).

For dots, dwell time durations were also evaluated differently ($\chi^2 = $ 18.47, $p < $ 0.001). Pairwise comparisons revealed three significant differences. First, object selection with a dwell time of 500 ms was considered significantly easier than with a dwell time of 1000 ms ($Z = $ -2.54, $p = $ 0.011), and 2000 ms ($Z = $ -2.97, $p = $ 0.003). Second, object selection with a dwell time of 1000 ms was assessed as significantly easier than with a dwell time of 2000 ms ($Z = $ -2.82, $p = $ 0.005). No statistically significant differences were found between the evaluations of a 250-ms and a 1000-ms dwell time ($Z = $ -0.67, $p = $ 0.503), a 500-ms dwell time ($Z = $ -2.22, $p = $ 0.027), and a 2000-ms dwell time ($Z = $ -2.09, $p = $ 0.036), although the latter two comparisons bordered on significance after Holm-Bonferroni correction on the alpha level ($p = $ 0.017 and $p = $ 0.025, respectively).

Also for selecting visual icons, the participant evaluations ($df = $ 3, n $ = $ 12) significantly differed ($\chi^2 = $ 26.84, $p < $ 0.001), with four significant pairwise comparisons. Object selection with a dwell time of 500 ms was considered significantly easier than with a dwell time of 250 ms ($Z = $ -2.82, $p = $ 0.005), and 2000 ms ($Z = $ -3.09, $p = $ 0.002). A dwell time of 1000 ms was judged as significantly easier to use than a dwell time of 250 ms ($Z = $ -2.95, $p = $ 0.003), or 2000 ms ($Z = $ -3.09, $p = $ 0.002).

Table 7.3. Results of Experiment 7. Pairwise comparisons of object dwell time usability for three object types.

Alphanumeric characters		Dots		Visual icons	
Dwell time (ms) comparison	Z-score	Dwell time (ms) comparison	Z-score	Dwell time (ms) comparison	Z-score
250 - 1000	-0.63	250 - 500	**-2.22**[†]	250 - 2000	-1.27
250 > 2000	**-2.66**[**]	250 - 1000	-0.67	500 > 250	**-2.82**[**]
500 > 250	**-2.31**[*]	250 - 2000	**-2.09**[†]	500 - 1000	-0.75
500 > 1000	**-3.10**[**]	500 > 1000	**-2.54**[*]	500 > 2000	**-3.09**[**]
500 > 2000	**-3.10**[**]	500 > 2000	**-2.97**[**]	1000 > 250	**-2.95**[**]
1000 > 2000	**-3.09**[**]	1000 > 2000	**-2.82**[**]	1000 > 2000	**-3.09**[**]

Z-score: Wilcoxon signed rank test value; >: significantly easier to use;
* p < 0.05, ** p < 0.01 (after Holm-Bonferroni-correction); † p < 0.10 (bordered on significance).

Besides differences in the usability of dwell time durations within a certain object type, differences in dwell time usability between object types were also examined with Friedman tests (Table 7.4, Appendix AK). For a dwell time of 250 ms, the participant evaluations ($df = 2$, n = 12) significantly differed ($\chi^2 = 13.32$, $p = 0.001$). After Holm-Bonferroni correction on the alpha level based on ranks [0.05/(m-k+1), where m = 3 indicates the number of pairs and k = 1, 2, …, m, indicates the level of ranks], pairwise comparisons with Wilcoxon signed-rank tests showed that alphanumeric object selection with a dwell time of 250 ms was considered significantly easier than visual icon selection ($Z = -2.97$, $p = 0.003$). For a 1000-ms dwell time the participant evaluations ($df = 2$, n = 12) also significantly differed ($\chi^2 = 13.28$, $p = 0.001$). Visual icon selection with a 1000-ms dwell time was considered significantly easier than selecting an alphanumeric character ($Z = -2.45$, $p = 0.014$) or a dot ($Z = -2.97$, $p = 0.003$).

Table 7.4. Results of Experiment 7. Differences in usability evaluations of four object dwell time durations between object types.

Dwell Time (ms)					
250		500	1000		2000
Object type comparison	Z-score		Object type comparison	Z-score	
Alphanumeric character > Visual icon	**-2.97****	Usability evaluations for this dwell time did not significantly differ between object types.	Visual icon > Dot	**-2.97****	Usability evaluations for this dwell time did not significantly differ between object types.
Alphanumeric character – Dot	**-**1.85		Visual icon > Alphanumeric character	**-2.45***	
Dot – Visual icon	**-**0.50		Dot – Alphanumeric character	**-**0.92	

Z-score: Wilcoxon signed rank test value;
>: significantly easier to use;
* p < 0.05, ** p < 0.01 (after Holm-Bonferroni-correction).

7.3 Discussion

The objective of Experiment 7 was to investigate the usability of various dwell times for selecting a sequence of 4 or 6 objects on four different grids. In this experiment, object selection from a display with eye-gaze-based input was investigated with object dwell times of 250, 500, 1000, and 2000 ms. Three different object types were used. Twelve participants were asked to memorize a 4-object and 6-object sequence, and to use their eye gaze to enter the sequence of objects onto a user interface. The selection time for each sequence and the number of object selection corrections were obtained and analyzed. Besides entering the sequence of objects onto the user interface, the participants were also required to evaluate the usability of the four dwell time durations.

The results of Experiment 7 showed that, first, the number of object selection corrections decreased with every increase in dwell time duration (Table 7.2). Most

object selection corrections were made with the dwell time of 250 ms for each object type and sequence. This dwell time is close to the minimal border of the fixation range, which is 200-600 ms (Sibert & Jacob, 2000; Cantoni, Galdi, Nappi, Porta, & Riccio, 2015). The fixation range is the time necessary to stabilize the eyes to fixate on something in the visual field. It is likely that the Midas-Touch problem occurred with the 250-ms dwell time (Jacob, 1991; Velichkovsky, Rumyantsev, & Morozov, 2014). With a relatively short dwell time, participants may have unintentionally selected objects while still scanning the display to identify potential target objects to form the correct sequence. These unwanted selections needed to be corrected, and hence the number of selection corrections was relatively high under a 250-ms dwell time per object.

The data suggest, however, that the number of object selection corrections under the 250-ms dwell time varied strongly with object type. For alphanumeric objects, a dwell time of 250 ms was considered relatively easy to use (Table 7.4). Under this dwell time the participants made fewer corrections either with 4-object sequences (1%) or 6-object sequences (4%) as compared with the other two object types (Table 7.2). It is most probable that experience with alphanumeric passwords in daily life, e.g., to access systems through manual input, improved input performance as well as usability evaluations. Indeed, the number of object selection corrections and dwell time evaluations for dots and visual icons were different. For example, a dwell time of 1000 ms was considered relatively easy to use for selecting visual icon objects. Perhaps because the participants were not familiar with using visual icons, they needed

relatively more time to search them on the display – something which they could not do easily with the 250-ms dwell time due to the Midas-Touch problem.

Second, object selection time, i.e., the total time needed by the participants to select a sequence of objects with eye-gaze-based input, varied depending on the number and type of objects to be selected. As one would expect, selecting 6 objects to enter a 6-object sequence overall took significantly more time than selecting 4 objects to enter a 4-object sequence. Furthermore, object selection time in case of correct object selections significantly increased with every increase in dwell time duration (Figure 7.3). This is in line with a previous study on object selection time using eye tracking (Penkar, Lutteroth, & Weber, 2012). However, although the object selection time increased when dwell time increased, the search time for various types of visual objects on a grid was stable at about 1000 ms per object (Table 7.1). As dwell time increased, the average object search time in a 4-object sequence ranged in between 952 and 1042 ms per object. Similarly, the average object search time in a 6-object sequence ranged from 923 to 965 ms per object. It is important to note that this search time was established for every object type, thus regardless of whether participants were experienced with the type of objects or not. If indeed the search time for object selection on a display is approximately steady at about 1000 ms per object, the total eye-gaze-based input time for similar types of search tasks on a grid-based display will be rather easy to compute.

The most salient finding in the participant evaluations was that a dwell time of 500 ms was easiest to use for eye-gaze-based selection of all three visual object types (Figure 7.4). Although very few object selection corrections were necessary with a

dwell time of 1000 and 2000 ms, these longer dwell times may have caused fatigue when fixating. As noted by Majaranta, Aula, and Räihä, (2004) in their eye-typing research, indeed object selection with a dwell time of 2000 ms could cause discomfort to users. Some participants even mentioned after this experiment that they thought there was something wrong with the interface or the eye-tracking system when using the 2000-ms dwell time.

Given the participant evaluations, the relatively low number of object selection corrections – in particular for alphanumeric sequences – and the relatively low object selection time for a complete sequence of objects, a dwell time of 500 ms is recommendable for eye-gaze-based object selection. However, fitting some mathematical functions through the data suggested that other dwell time durations between 250 and 1000 ms can be also possible to use for selecting various object types. Therefore, in the future, more research needs to be done to confirm this issue with a wider dwell time duration ranging from 200 ms to around 1000 ms.

Chapter 8. General discussion and conclusions

This dissertation investigated the physical aspects of using low-cost eye-tracking devices for registration users' eyes into a certain system that requires interactive interfaces. Furthermore, with the same low-cost eye trackers, this dissertation also investigated the cognitive aspects of users' abilities for selecting a sequence of visual objects on a grid-based interface screen with eye-gaze-based input. To achieve this, a series of experiments has been conducted in this dissertation.

The findings obtained related to the physical aspects of using low-cost eye-tracking devices (the Tobii EyeX$^©$ and Eye Tribe$^©$) in Chapters 3 and 4 of this dissertation were the following. In Chapter 3, Experiments 1 and 2 were described to obtain maximum and minimum viewing distances and the highest and lowest viewing angle under three lighting conditions at which the participants' eyes could be registered into low-cost eye-tracker interfaces. The participants were asked to stand in front of a display and the participants were instructed to register their eyes into eye-tracker interfaces. Both the average viewing distances and viewing angles under different lighting conditions were obtained. Although no performance comparisons were made between eye trackers, on average, the low-cost Eye Tribe$^©$ eye-tracking device in Experiment 2 could register participants' eyes at somewhat larger and closer viewing distances and at higher and lower viewing angles, than the low-cost Tobii EyeX$^©$ eye-tracking device used in Experiment 1.

Furthermore, an important result from Experiment 1 is that the minimum and maximum viewing limit of the Tobii EyeX$^©$ did not vary significantly with different lighting conditions, and neither did the viewing angle range. Experiment 2, however,

showed that the maximum viewing distance of the Eye Tribe$^©$ was significantly different between two conditions of artificial lighting, although the average difference was just 1 cm (room light: 76.8 cm; full light: 77.8 cm). The lowest viewing angle that allowed registration into the Eye Tribe$^©$ system also differed significantly between the two artificial lighting conditions, the average angle difference was small too − 0.6 degrees (room light: 81.3 degrees; full light: 81.9 degrees). Generally, the results of Experiments 1 and 2 indicated that both low-cost eye-tracking devices had relatively stable measuring results. From the results in Chapter 3, it can be concluded that

(i)   the performance of the two low-cost eye trackers tested in this dissertation was robust enough under different conditions of illuminance and luminance for participants' eyes registration at various viewing distances and viewing angles.

However, the calibration quality of Eye Tribe$^©$ was poor especially for participants who wore glasses. In Experiment 2, from 10 participants who wore glasses, relatively poor calibration was obtained in the natural light for 6 participants, in the room light for 5 participants, and in the full light for 7 participants. This result was in line with the previous finding that the calibration quality of eye-tracking devices was indeed poorer for participants with glasses than those without glasses (Funke et al., 2016).

In Chapter 4 (Experiment 3), the aim was to investigate the effect of wearing glasses on the calibration process into the eye-tracking device under the same lighting conditions as used in Experiments 1 and 2. The same participants with uncorrected vision carried out the calibration without glasses and with non-prescription, clear

replica glasses, while the same participants with prescription glasses did the calibration both with and without prescription glasses, if they were able to do so. The calibration quality and calibration time data for the same participants in both with and without glasses conditions were obtained. Experiment 3 in Chapter 4 showed that

(ii)   with the low-cost eye-tracking devices used in the studies described in this dissertation, the calibration quality was poorer, and the calibration time was longer for participants who wore glasses as compared to the same participants without glasses under three different lighting conditions.

The obtained result of Experiment 3 was not due to poor eyesight in the group with corrected vision; the participants with prescription glasses should have had difficulty performing the calibration task without their glasses, while the participants without glasses (or with contact lenses) should have had no increased difficulty performing the calibration with replica glasses. A possible explanation for this could be the presence of light reflection and glare when using glasses. Furthermore, when tracking an object on the screen at a viewing angle of 90 degrees, participants' eyes were sometimes obstructed by the thick frame of the glasses. It is imaginable that if eye tracking is going to be used in public settings for future applications, users will perform registration under different viewing angles. Users may be standing in front of a device (e.g., ticketing machines or automated teller machines) and register from viewing angles that differ because of differences in their height. Users may be sitting behind a device as well, because they are in a wheelchair or use eye tracking while sitting behind a steering wheel in a vehicle (Kandil, Rotter, & Lappe, 2010).

In Chapter 4, therefore, the first goal of Experiment 4 was to investigate the ideal angle of the display and the Eye Tribe© device for participants with different heights for the registration into the eye-tracker interface. The participants were asked to stand at a natural viewing position in front of the eye tracking device – as if they were using an ATM machine –, while some participants were also asked to sit in front of the eye tracker. The angle of the display and the eye tracker were systematically varied. Experiment 4 with the Eye Tribe© eye-tracking device showed a significant correlation between the height of the participants and the angles of the eye tracker and the display under which they could register themselves. In summary, if the participant was tall or viewing from a high position looking down on the display, the display should be angled upwards in a more horizontal position for registration to occur. By contrast, people sitting down on chairs or in wheelchairs would benefit from a vertical display position or a downward angle.

The second goal of Experiment 4 was to perform calibration for participants with and without (replica) glasses at the different display angles under two different lighting conditions. The same participants with uncorrected vision carried out the calibration without glasses and with non-prescription, clear replica glasses, while the same participants with prescription glasses did the calibration both with and without glasses if they were able to perform. Since the reflection of the light spot appeared on the display at certain angles in the full light condition, the natural and room light conditions were only used in Experiment 4. The calibration quality and calibration time data were obtained. Although no statistics were performed due to differences in group sizes, the calibration quality and calibration time for participants with glasses in

Experiment 4 seemed better than in Experiment 3 under two lighting conditions. The results of Experiment 4 in Chapter 4 strongly suggest that

(iii)   in order to achieve good and fast calibration with the low-cost eye-tracking devices used in the studies described in this dissertation, the eye-tracking device and the display need to be set at a certain angle, depending on the viewing height of participants. If viewing angles can be adjusted, participants with glasses should be able to perform the calibration without problems.

The findings obtained with regard to the cognitive aspects of users' abilities for selecting a sequence of visual objects on a grid-based interface screen with eye-gaze-based input, as described in Chapters 5 to 7 of this dissertation, were the following. The similar low-cost eye-tracker devices (the Tobii EyeX© and Tobii Eye Tracker 4C©) were utilized as an eye-gaze-based input device to select a sequence of visual objects on various grid-based screen interfaces. The visual objects that were used consisted of alphanumeric characters, dots, and visual icons, from which a single visual password with a different number of objects needed to be selected. An interactive interface with multiple objects shown on the screen typically employs a grid to organize objects based on sequenced columns and rows (for details on visual password systems with higher grid densities, see Chapter 2).

The goal in Chapter 5 (Experiment 5) was to identify which grid densities potentially are suitable to use with low-cost eye trackers. In Experiment 5, sixteen different grid formations were used in between 2×2 to 7×7 cells. When sitting in front of a computer, twenty-seven participants were asked to create an imaginary password

by selecting objects on the display using manual input with a mouse, thus without eye tracking. The imaginary password consisted of four to eight objects (alphanumeric characters, dots, or visual icons). Next, the participants were also asked to judge the 16 grid densities about whether the grids are easy to use and potentially safe for making a visual password in an imagined situation using eye tracking. The user judgments of 16 different grid densities for three visual password formats were obtained. Experiment 5 showed that a grid generally was considered more difficult to use, but potentially safer to make any password in three formats, when the number of grid cells increased. Furthermore, for each visual password format, some grid densities were thought to be relatively difficult to use (e.g., a 7×7 grid) or potentially unsafe (e.g., a 2×2 grid). The results thus suggest that grid densities from 3×3 to 6×6 cells are suitable to use with low-cost eye tracking devices.

In Chapter 6, the goal of Experiment 6 was to investigate which type of password format and grid formation is suitable for password authentication using eye-gaze-based input. The Tobii EyeX© device was used. Participants were asked to memorize a 4-object and a 6-object password for three types of password formats and register (Task 1), confirm (Task 2), and log in (Task 4) the password on a grid by using eye-gaze-based input. The three recognition-based password formats were an alphanumeric format, a pattern format, and a picture format. Grid densities and formations were varied in 16 ways in between 3×3 and 6×6 object keys, following the results of Experiment 5. Task-completion time and task-success rate data were obtained. Participants also provided preference data based on a rating scale about the grid densities and formations (Task 3). Experiment 6 showed that, first, task-

completion time was mostly shorter for the alphanumeric password format than for the pattern or picture format. Second, task-completion time of 4-object or 6-object passwords generally increased as the grid density increased, while the task-success rate at the first attempt decreased when the grid density increased. Finally, task-completion time often was faster for grids with more columns than rows (horizontal formations, e.g., 4×3 cells) than for grids with more rows than columns (vertical formations, e.g., 3×4 cells). From the results of Experiment 6 in Chapter 6, it can be concluded that

(iv)   the chance of performing quick and successful password authentication by eye-gaze-based input improves with horizontal grids (e.g., with more columns than rows, as in 4×3, 5×3, 6×3, 5×4, 6×4, or 6×5 grids) with relatively low grid densities.

It can be said that horizontal grids are more efficient for password formation with eye-gaze-based input. This corroborates studies on the visual search of objects or words, which have reported that the direction of the users' eye movements often occurs more horizontally than vertically (Duchowski, 2007; Ojanpää et al., 2002; Goonetilleke et al., 2002). Participants also needed more time to make the password on a denser grid in which the chance of unwanted object selection with eye tracking increases. Furthermore, with increasing grid density, the passwords became more complex. The results of Experiment 6 in Chapter 6 also strongly suggests that

(v)   the alphanumeric password format is the easiest to use for object selection with eye-gaze-based input, in that password input consisting of alphanumeric characters required relatively less time and relatively few mistakes were made.

This result could be because participants most likely are more familiar with passwords consisting of numbers and letters in which they may memorize passwords by "chunking", which allows them to recall passwords fast (Nelson & Vu, 2010). It was known that often-used objects are easier to remember (Kinsbourne & George, 1974). In the case of using icons (picture format) or dots (pattern format) in Experiment 6, the participants had not much time to adapt to these objects.

A limitation of Experiment 6 was that a dwell time duration of 500 ms was used to select a sequence of visual objects in passwords with eye-gaze-based input. This duration may have been too short for participants or for some visual objects. In Chapter 7, the goal of Experiment 7 was to investigate the usability of various dwell times for selecting a sequence of 4 or 6 objects on four different grids. In this experiment, object selection from a display with eye-gaze-based input was investigated with object dwell times of 250, 500, 1000, and 2000 ms. Three different object types were used. A Tobii Eye Tracker 4C$^{©}$ device was used as an eye-gaze-based input device. Twelve participants were asked to memorize a 4-object and 6-object sequence, and to use their eye gaze to enter the sequence of objects onto a user interface. Besides entering the sequence of objects onto the user interface, the participants were also required to evaluate the usability of the four dwell time

durations. The selection time for each sequence and the number of object selection corrections, and dwell time evaluations were obtained.

Experiment 7 showed that regardless of the number and type of objects that had to be selected, the participants needed about 1000 ms to search a single target object on the display. Under this steady search time, as expected, the total time necessary to select 4 or 6 objects (object selection time) increased when dwell time increased, but with fewer object selection corrections. Experiment 7 also showed that a dwell time of 500 ms per object was evaluated as easier to use for eye-gaze-based selection of all three types of visual objects. The results of Experiment 7 in Chapter 7 strongly suggests that

(vi) a dwell time of 500 ms was recommendable for object selection using eye-gaze-based input. This is based on the relatively low number of object selection corrections (particularly for alphanumeric sequences), the relatively low object selection time for a complete sequence of objects, and the participant evaluations of dwell time durations.

Compared to this dwell time, with relatively short dwell time (e.g., 250-ms), participants may have unintentionally selected objects while still scanning the display to identify potential target objects to form the correct sequence. In reading tasks, regardless of whether words are familiar or not, lexical activation and recognition require fixation durations on average in between 200 and 300 ms (Rayner & Pollatsek, 1989). In scene perception and visual search tasks, the fixation duration is approximately in between 180 and 330 ms (Rayner, 2009). Thus, with a short dwell

time of 250 ms, as used here, the visual object recognition process is still in progress. In Experiment 7, therefore, the participant sometimes selected undesired objects by chance with a dwell time of 250 ms. Furthermore, it has been reported that object selection with more than 1000-ms dwell time could cause discomfort to users (Majaranta et al., 2004). Goldberg and Kotval (1999) also indicated that longer dwell time were inconvenient in retrieving general information from a display.

Overall, this dissertation describes research conducted to support the varying needs of the user in general or in certain situations to perform interactive interface tasks such as password authentication using eye-gaze-based input. It is expected that this dissertation will be a useful starting point and resource for researchers. Furthermore, it is also hoped that the dissertation will give helpful guidelines for developers of gaze-based authentication systems.

A limitation of the research is that since low-cost eye tracking devices were used in the experiments in this dissertation, it was not possible to obtain some important data related to the eye movement metrics, e.g., calibration time, fixation duration, first fixation, saccades, etc. In the future, more research needs to be done to confirm the present conclusions with a sophisticated eye tracking device. Some other issues will need further investigation. First, the efficiency of the low-cost eye-tracker device with regard to pupil size and occlusion still needs to be investigated. Mostly, eye-tracking techniques rely on the visibility of the pupil and most registration errors occur due to (partial) pupil occlusion. Partial pupil occlusion mainly occurs when wearing glasses and it varies with ethnicity (Blignaut & Wium, 2014). Furthermore, if

the size of grid cells that used to place objects is too small, the target objects probably difficult to be reliably selected by using eye-gaze-based input. It is known that visual acuity in eye tracking depends on the size of target objects in the visual angle (see Chapter 2). Further research is thus necessary to investigate how big or small the size of grids can be used for eye-gaze-based object selection.

Another issue is the effect of practice as regards object selection time and the number of object selection corrections. For example, it is known that participants may get used to searching visual objects on screen after a few days of practice (Baluch & Itti, 2010; Clark, Appelbaum, van den Berg, Mitroff, & Woldorff, 2015). As a result, visual search time may decrease and visual search accuracy may increase or at least become consistent. It is important to investigate whether object selection time and the number of selection corrections would decrease when the participants have more time to practice, especially with dot and visual icon objects. There is a possibility that the Midas-Touch problem will also become less with increasing user experience, which would suggest that an object dwell time of even less than 500 ms may become feasible in certain interactive interface systems. Finally, one issue open to investigation is whether older users would have the same dwell time usability evaluations as younger users. It is known that the visual area in which information can be obtained within one eye fixation reduces in size as a function of age (Ball et al., 1988). Furthermore, visual search accuracy is generally also affected by age (Madden, Gottlob, & Allen, 1999; Lee, Kim, & Ji, 2019). More research is thus necessary.

References

Alam, S. (2016). SUIS: An online graphical Signature-Based User Identification System. In *Proceedings of the 2016 Sixth International Conference on Digital Information and Communication Technology and its Applications (DICTAP)* (pp. 85–89). IEEE. https://doi.org/10.1109/DICTAP.2016.7544006.

Aviv, A. J., Budzitowski, D., & Kuber, R. (2015). Is Bigger Better? Comparing User-Generated Passwords on 3x3 vs. 4x4 Grid Sizes for Android's Pattern Unlock. In *Proceedings of the 31st Annual Computer Security Applications Conference on - ACSAC 2015* (pp. 301–310). New York, New York, USA: ACM Press. https://doi.org/10.1145/2818000.2818014.

Ball, K. K., Beard, B. L., Roenker, D. L., Miller, R. L., & Griggs, D. S. (1988). Age and *visual* search: expanding the useful field of view. *Journal of the Optical Society of America A*, *5*(12), 2210–2219. https://doi.org/10.1364/JOSAA.5.002210.

Baluch, F., & Itti, L. (2010). Training top-down attention improves performance on a triple-conjunction search task. *PLoS ONE*, *5*(2). https://doi.org/10.1371/journal.pone.0009127.

Bee, N., & André, E. (2008). Writing with Your Eye: A Dwell Time Free Writing System Adapted to the Nature of Human Eye Gaze. In *Perception in Multimodal Dialogue Systems* (Vol. 5078 LNCS, pp. 111–122). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-69369-7_13.

Bee, N., Wagner, J., André, E., Charles, F., Pizzi, D., & Cavazza, M. (2010). Interacting with a gaze-aware virtual character. In *Proceedings of the 2010 workshop on Eye gaze in intelligent human machine interaction - EGIHMI '10* (pp. 71–77). New York, New York, USA: ACM Press. https://doi.org/10.1145/2002333.2002345.

Biddle, R., Chiasson, S., & van Oorschot, P. C. (2012). Graphical Passwords: Learning from the First Twelve Years. *ACM Computing Surveys*, *44*(4), 1–41. https://doi.org/10.1145/2333112.2333114.

Bishop, M. (2005). Psychological Acceptability Revisited. In L. Cranor and S. Garfinkel (Eds.), *Security and Usability: Designing Secure Systems That People*

Can Use (pp. 1-12), O'Reilly Media, Sebastopol, C.A.

Blignaut, P., & Wium, D. (2014). Eye-tracking data quality as affected by ethnicity and experimental design. Behavior Research Methods, 46(1), 67–80. https://doi.org/10.3758/s13428-013-0343-0.

Boraston, Z., & Blakemore, S.-J. (2007). The application of eye-tracking technology in the study of autism. *The Journal of Physiology*, *581*(3), 893–898. https://doi.org/10.1113/jphysiol.2007.133587.

Brigham, F., Zaimi, E., Jo Matkins, J., Shields, J., McDonnouugh, J., & Jakubecy, J. (2001). The eyes may have it: Reconsidering eye-movement research in human cognition. In Scruggs, T. and Mastropieri, M. (Eds.), *Technological Applications* (pp. 39-59), Emerald Group Publishing Limited, Bingley, https://doi.org/10.1016/S0735-004X(01)80005-7.

Cantoni, V., Galdi, C., Nappi, M., Porta, M., & Riccio, D. (2015). GANT: Gaze analysis technique for human identification. *Pattern Recognition*, *48*(4), 1027–1038. https://doi.org/10.1016/j.patcog.2014.02.017.

Chun, M. M., & Wolfe, J. M. (2005). Visual Attention. In *Blackwell Handbook of Sensation and Perception* (pp. 272–310). Malden, MA, USA: Blackwell Publishing Ltd. https://doi.org/10.1002/9780470753477.ch9.

Clark, K., Appelbaum, L. G., van den Berg, B., Mitroff, S. R., & Woldorff, M. G. (2015). Improvement in Visual Search with Practice: Mapping Learning-Related Changes in Neurocognitive Stages of Processing. *Journal of Neuroscience*, *35*(13), 5351–5359. https://doi.org/10.1523/JNEUROSCI.1152-14.2015.

De Angeli, A., Coventry, L., Johnson, G., & Renaud, K. (2005). Is a picture really worth a thousand words? Exploring the feasibility of graphical authentication systems. *International Journal of Human-Computer Studies*, *63*(1–2), 128–152. https://doi.org/10.1016/j.ijhcs.2005.04.020.

De Luca, A., Denzel, M., & Hussmann, H. (2009). Look into my Eyes! Can you guess my Password? In *Proceedings of the 5th Symposium on Usable Privacy and Security - SOUPS '09*. New York, New York, USA: ACM Press. https://doi.org/10.1145/1572532.1572542.

De Luca, A., Weiss, R., & Drewes, H. (2007). Evaluation of eye-gaze interaction methods for security enhanced PIN-entry. In *Proceedings of the 19th*

Australasian conference on Computer-Human Interaction: Entertaining User Interfaces - OZCHI '07 (pp. 199–202). New York, New York, USA: ACM Press. https://doi.org/10.1145/1324892.1324932.

Dhamija, R., & Perrig, A. (2000). Deja vu: A User Study Using Images for Authentication. In *Proceedings of the 9th conference on USENIX Security Symposium - SSYM'00*. Retrieved from http://portal.acm.org/citation.cfm?id=1251306.1251310.

Di Giorgio, E., Turati, C., Altoè, G., & Simion, F. (2012). Face detection in complex visual displays: An eye-tracking study with 3- and 6-month-old infants and adults. *Journal of Experimental Child Psychology*, *113*(1), 66–77. https://doi.org/10.1016/j.jecp.2012.04.012.

Duchowski, A. T. (2007). Industrial Engineering and Human Factors. In: Springer. *Eye Tracking Methodology: Theory and Practice Second Edition* (pp. 242–260). Verlag London.

Duchowski, A. T. (2018). Gaze-based interaction: A 30 year retrospective. *Computers & Graphics*, *73*, 59–69. https://doi.org/10.1016/j.cag.2018.04.002.

Dunphy, P., Fitch, A., & Olivier, P. (2008). Gaze-contingent passwords at the ATM. In *Proceedings of The 4th Conference on Communication by Gaze Interaction – COGAIN* (pp. 59–62). COGAIN NoE. Retrieved from http://homepages.cs.ncl.ac.uk/p.m.dunphy/COGAIN2008.pdf.

Ellis, E. M., Borovsky, A., Elman, J. L., & Evans, J. L. (2015). Novel word learning: An eye-tracking study. Are 18-month-old late talkers really different from their typical peers? *Journal of Communication Disorders*, *58*, 143–157. https://doi.org/10.1016/j.jcomdis.2015.06.011.

Ellis, S.L., Candrea, R., Misner, J.W., Craig, C.J., Lankford, C., & Hutchinson, T.E. (1998) Windows to the soul? What eye movements tell us about software usability. In: *Proceedings of 7th annual conference of the usability professionals' association*, Washington, pp 151–178.

Fitts, P. M., Jones, R.E. & Milton, J.L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review*, *9*(2), 24-29.

Forget, A., Chiasson, S., & Biddle, R. (2010). Shoulder-surfing resistance with eye-gaze entry in cued-recall graphical passwords. In *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems - CHI '10 (pp. 1107–1110). New York, New York, USA: ACM Press. https://doi.org/10.1145/1753326.1753491.

Funke, G., Greenlee, E., Carter, M., Dukes, A., Brown, R., & Menke, L. (2016). Which Eye Tracker Is Right for Your Research? Performance Evaluation of Several Cost Variant Eye Trackers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 1240–1244. https://doi.org/10.1177/1541931213601289.

Goonetilleke, R. S., Lau, W. C., & Shih, H. M. (2002). Visual search strategies and eye movements when searching Chinese character screens. *International Journal of Human-Computer Studies*, *57*(6), 447–468. https://doi.org/10.1006/ijhc.2002.1027.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and construct. *International Journal of Industrial Ergonomics*, *24*(6), 631–645. https://doi.org/10.1016/S0169-8141(98)00068-7.

Graham, D. J., & Jeffery, R. W. (2012). Predictors of nutrition label viewing during food purchase decision making: an eye tracking investigation. *Public Health Nutrition*, *15*(2), 189–197. https://doi.org/10.1017/S1368980011001303.

Haber, R.N. & Hershenson, M. (1973) The Psychology of Visual Perception. London: Holt, Rinehart and Winston.

Halverson, T., & Hornof, A. J. (2004). Local Density Guides Visual Search: Sparse Groups are First and Faster. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(16), 1860–1864. https://doi.org/10.1177/154193120404801615.

Hansen, J. P., Johansen, A. S., Hansen, D. W., Itoh, K., & Mashino, S. (2003). Command Without a Click: Novice User Studies of Dwell Time Typing by Mouse and Eye-Gaze Selections. *Human-Computer Interaction -- INTERACT'03*, 121–128.

Hutchinson, T. E., White, K. P., Martin, W. N., Reichert, K. C., & Frey, L. A. (1989). Human-computer interaction using eye-gaze input. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(6), 1527–1534. https://doi.org/10.1109/21.44068.

Hyrskykari, A., Majaranta, P., Aaltonen, A., & Räihä, K.-J. (2000). Design Issues of iDict: A Gaze-Assisted Translation Aid. In *Proceedings of the symposium on Eye tracking research & applications - ETRA '00* (pp. 9–14). New York, New York, USA: ACM Press. https://doi.org/10.1145/355017.355019.

Jacob, R. J. K. (1991). The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems*, *9*(2), 152–169. https://doi.org/10.1145/123078.128728.

Jacob, R. J. K. (1995). Eye Tracking in Advanced Interface Design. In *Barfield W, Furness TA (eds) Virtual environments and advanced interface design* (pp. 258–288). Oxford University Press, New York.

Jacob, R. J. K., & Karn, K. S. (2003). Commentary on Section 4 - Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. In *The Mind's Eye* (pp. 573–605). Elsevier. https://doi.org/10.1016/B978-044451020-4/50031-1.

Janthanasub, V., & Meesad, P. (2015). Evaluation of a Low-cost Eye Tracking System for Computer Input. *KMUTNB International Journal of Applied Science and Technology*, *8*(3), 1–12. https://doi.org/10.14416/j.ijast.2015.07.001.

Jermyn, I., Mayer, A., Monrose, F., Reiter, M. K., & Rubin, A. D. (1999). The Design and Analysis of Graphical Passwords. In *Proceedings of the 8th USENIX Security Symposium* (pp. 1–14). USENIX. Retrieved from https://www.usenix.org/legacy/publications/library/proceedings/sec99/full_pape rs/jermyn/jermyn.pdf?CFID=1379024&CFTOKEN=9569d9d8626e3dbd-50C19994-0569-6569-867B982A21E34346.

Kandil, F. I. (2010). Car drivers attend to different gaze targets when negotiating closed vs. open bends. *Journal of Vision*, *10*(4), 1–11. https://doi.org/10.1167/10.4.24.

Kinsbourne, M., & George, J. (1974). The mechanism of the word-frequency effect on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(1), 63–69. https://doi.org/10.1016/S0022-5371(74)80031-9.

Kumar, M., Garfinkel, T., Boneh, D., & Winograd, T. (2007). Reducing Shoulder-surfing by Using Gaze-based Password Entry. In *SOUPS '07 Proceedings of the 3rd symposium on Usable privacy and security* (pp. 13–19). Retrieved from https://cups.cs.cmu.edu/soups/2007/proceedings/p13_kumar.pdf.

Kurauchi, A., Feng, W., Joshi, A., Morimoto, C., & Betke, M. (2016). EyeSwipe: Dwell-free Text Entry Using Gaze Paths. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (pp. 1952–1956). New York, New York, USA: ACM Press. https://doi.org/10.1145/2858036.2858335.

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1358), 1231–1239. https://doi.org/10.1098/rstb.1997.0105.

Lee, S. C., Kim, Y. W., & Ji, Y. G. (2019). Effects of visual complexity of in-vehicle information display: Age-related differences in visual search task in the driving context. *Applied Ergonomics*, *81*(May), 102888. https://doi.org/10.1016/j.apergo.2019.102888.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013.

Madden, D. J., Gottlob, L. R., & Allen, P. A. (1999). Adult age differences in visual search accuracy: Attentional guidance and target detectability. *Psychology and Aging*, *14*(4), 683–694. https://doi.org/10.1037/0882-7974.14.4.683.

Maeder, A., Fookes, C., & Sridharan, S. (2005). Gaze based user authentication for personal computer applications. In *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004.* (pp. 727–730). IEEE. https://doi.org/10.1109/ISIMP.2004.1434167.

Majaranta, P. (2009). *Text Entry by Eye Gaze. Dissertations in Interactive Technology, number 11*.

Majaranta, P., Ahola, U., & Špakov, O. (2009). Fast gaze typing with an adjustable dwell time. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09* (p. 357). New York, New York, USA: ACM Press. https://doi.org/10.1145/1518701.1518758.

Majaranta, P., Aula, A., & Räihä, K.-J. (2004). Effects of feedback on eye typing with a short dwell time. In *Proceedings of the Eye tracking research & applications symposium on Eye tracking research & applications - ETRA'2004* (Vol. 1, pp. 139–146). New York, New York, USA: ACM Press.

https://doi.org/10.1145/968363.968390.

Majaranta, P., & Bulling, A. (2014). Eye Tracking and Eye-Based Human–Computer Interaction. In *Advances in Physiological Computing* (pp. 39–65). https://doi.org/10.1007/978-1-4471-6392-3_3.

Majaranta, P., MacKenzie, I. S., Aula, A., & Räihä, K.-J. (2003). Auditory and visual feedback during eye typing. In *CHI '03 extended abstracts on Human factors in computer systems - CHI '03* (p. 766). New York, New York, USA: ACM Press. https://doi.org/10.1145/765978.765979.

Majaranta, P., & Räihä, K.-J. (2002). Twenty Years of Eye Typing: Systems and Design Issues. In *Proceedings of the symposium on Eye tracking research & applications - ETRA '02* (p. 15). New York, New York, USA: ACM Press. https://doi.org/10.1145/507072.507076.

Mihajlov, M., & Jerman-Blazic, B. (2018). Eye Tracking Graphical Passwords. In D. Nicholson (Ed.), *Advances in Intelligent Systems and Computing* (Vol. 593, pp. 37–44). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-60585-2_4.

Mihajlov, M., Trpkova, M., & Arsenovski, S. (2013). Eye tracking recognition-based graphical authentication. In *2013 7th International Conference on Application of Information and Communication Technologies* (pp. 1–5). IEEE. https://doi.org/10.1109/ICAICT.2013.6722632.

Miniotas, D., Spakov, O., & Evreinov, G. (2003). Symbol Creator : An Alternative Eye-based Text Entry Technique with Low Demand for Screen Space. *Proceedings of Human-Computer Interaction - INTERACT'03*, (c), 137–143.

Navab, A., Gillespie-Lynch, K., Johnson, S. P., Sigman, M., & Hutman, T. (2012). Eye-Tracking as a Measure of Responsiveness to Joint Attention in Infants at Risk for Autism. *Infancy*, *17*(4), 416–431. https://doi.org/10.1111/j.1532-7078.2011.00082.x.

Nelson, D., & Vu, K.-P. L. (2010). Effectiveness of image-based mnemonic techniques for enhancing the memorability and security of user-generated passwords. *Computers in Human Behavior*, *26*(4), 705–715. https://doi.org/10.1016/j.chb.2010.01.007.

Nyström, M., Andersson, R., Holmqvist, K., & van de Weijer, J. (2013). The influence

of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, *45*(1), 272–288. https://doi.org/10.3758/s13428-012-0247-4.

Ojanpää, H., Näsänen, R., & Kojo, I. (2002). Eye movements in the visual search of word lists. *Vision Research*, *42*, 1499–1512.

Ooms, K., Lapon, L., Dupont, L., & Popelka, S. (2015). Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental set-ups. *Journal of Eye Movement Research*, *8*(1), 1–24. https://doi.org/10.16910/jemr.8.1.55.

Penkar, A. M., Lutteroth, C., & Weber, G. (2012). Designing for the Eye – Design Parameters for Dwell in Gaze Interaction. In *Proceedings of the 24th Australian Computer-Human Interaction Conference on - OzCHI '12* (pp. 479–488). New York, New York, USA: ACM Press. https://doi.org/10.1145/2414536.2414609.

Rayner, K. (2009). *Eye movements and attention in reading, scene perception, and visual search. Quarterly Journal of Experimental Psychology* (Vol. 62). https://doi.org/10.1080/17470210902816461

Rayner, K., & Pollatsek, A. (1989). *The Psychology of Reading*. Englewood Cliffs, NJ: Prentice Hall.

Renaud, K. (2005). Evaluating Authentication Mechanisms. In L. Cranor and S. Garfinkel (Eds.), *Security and Usability: Designing Secure Systems That People Can Use* (pp. 103-128), O'Reilly Media, Sebastopol, C.A.

Shechner, T., Jarcho, J. M., Britton, J. C., Leibenluft, E., Pine, D. S., & Nelson, E. E. (2013). Attention bias of anxious youth during extended exposure of emotional face Pairs: An eye-tracking study. *Depression and Anxiety*, *30*(1), 14–21. https://doi.org/10.1002/da.21986.

Sibert, L. E., & Jacob, R. J. K. (2000). Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00* (Vol. 11, pp. 281–288). New York, New York, USA: ACM Press. https://doi.org/10.1145/332040.332445.

Spakov, O., & Miniotas, D. (2004). On-line adjustment of dwell time for target selection by gaze. In *Proceedings of the third Nordic conference on Human-computer interaction - NordiCHI '04* (pp. 203–206). New York, New York, USA: ACM Press. https://doi.org/10.1145/1028014.1028045.

Stampe, D. M., & Reingold, E. M. (1995). Selection By Looking: A Novel Computer Interface And Its Application To Psychological Research. In *Journal of Geotechnical and Geoenvironmental Engineering ASCE* (Vol. 120, pp. 467–478). https://doi.org/10.1016/S0926-907X(05)80039-X.

Stawicki, P., Gembler, F., Rezeika, A., & Volosyak, I. (2017). A Novel Hybrid Mental Spelling Application Based on Eye Tracking and SSVEP-Based BCI. *Brain Sciences*, *7*(12), 35. https://doi.org/10.3390/brainsci7040035.

Tao, H., & Adams, C. (2008). Pass-Go : A Proposal to Improve the Usability of Graphical Passwords. *International Journal of Network Security*, *7*(2), 273–292. https://doi.org/10.6633/IJNS.

Thorpe, J., & van Oorschot, P. C. (2004). Towards Secure Design Choices for Implementing Graphical Passwords. In *20th Annual Computer Security Applications Conference* (pp. 50–60). IEEE. https://doi.org/10.1109/CSAC.2004.44.

Tobii (2006) User Manual: Tobii Eye Tracker and ClearView analysis software. Tobii Technology AB.

Tobii Eye Tracking Support (2017). What's the difference between Tobii Eye Tracker 4C and Tobii EyeX?. Available from: https://help.tobii.com/hc/en-us/articles/212814329-What-s-the-difference-between-Tobii-Eye-Tracker-4C-and-Tobii-EyeX-.

Velichkovsky, B. B., Rumyantsev, M. A., & Morozov, M. A. (2014). New Solution to the Midas Touch Problem: Identification of Visual Commands Via Extraction of Focal Fixations. *Procedia Computer Science*, *39*(C), 75–82. https://doi.org/10.1016/j.procs.2014.11.012.

Ware, C., & Mikaelian, H. H. (1987). An evaluation of an eye tracker as a device for computer input2. *ACM SIGCHI Bulletin*, *17*(SI), 183–188. https://doi.org/10.1145/30851.275627.

Zieliński, R. (2005). Estimating quantiles with Linex loss function. Applications to VaR estimation. Applicationes Mathematicae, 32(4), 367–373. https://doi.org/10.4064/am32-4-1.

Appendix A.   Instruction and informed consent of Experiment 1 and 2

Instruction and Informed Consent of Eye-tracking experiment for participants on 8th floor building 3 (Ohashi Campus), at Kyushu University, Department of Human Science, Japan.

Dear participant,

Thank you for joining our eye tracking experiment.
We would like to measure
1. The maximum and minimum viewing distances at which we can use the eye tracker,
2. The up-down angles at which the eye tracker is effective.

We expect individual differences based on eye-size and other characteristics. If you agree, we would also like to take a picture of your eyes. There are no (health) risks involved in joining the experiment and bear in mind that you can opt-out of the experiment any time - participation is on voluntary basis. The experiment will take about 20 minutes.

To process the data accurately, we would like to ask you for some information. We will use the information to analyze our data and, possibly, for data publication of group means. However, we guarantee your privacy: your data will be numbered and we will not disclose data of single individuals.

Here are our questions:
1. Do you wear glasses or contact lenses? [yes / no]
2. How tall are you?   ...... cm
3. What is your age?   ......   years old
4. Can we take a picture of your eyes? [yes / no]

Experiment instructions:
1. We will first calibrate the eye tracker when it is at a distance of 40 cm from your eyes. Please follow the instructions of the experimenter.
2. We then will measure at what maximum and minimum distance the eye tracker works. Please follow the instructions of the experimenter.
3. We will then measure at what maximum upper and lower angle the eye tracker works, from a viewing distance of 40 cm from the eye tracker. Please follow the instructions of the experimenter.

Thank you for your participation!
Yesaya Tommy Paulus, Gerard B. Remijn

Eye tracking experiment - Written informed consent
Experimenter: Yesaya Tommy Paulus

Your signature on this form indicates that you understand to your satisfaction the information provided to you about your participation in this experiment, and agree to participate as a research participant.

You are free to withdraw from this experiment at any time. You should feel free to ask for clarification or new information throughout your participation.

Participant's Name:

Participant's Signature:

Date:

Appendix B.   Experiment set-up (Experiment 1 - Tobii EyeX$^©$ eye tracker)
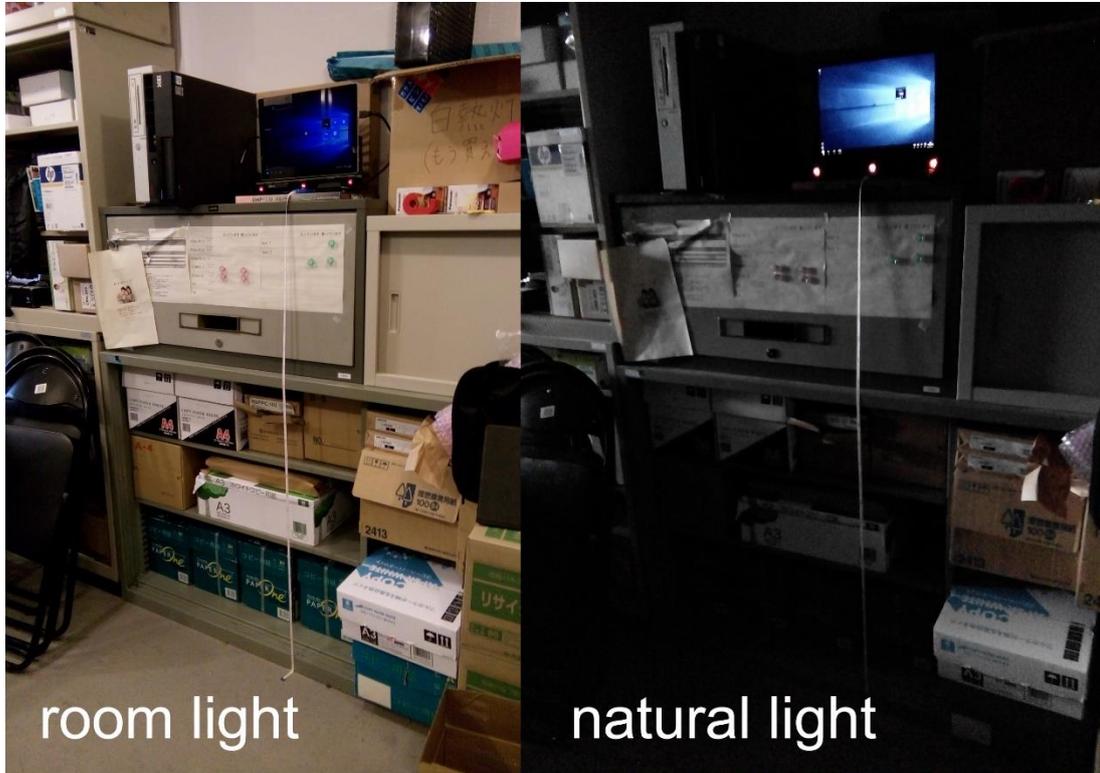
Appendix C.　Participants' face photos of Experiment 1

Appendix D.   Statistical analysis of Experiment 1 data

1. Normality check of maximum viewing distance data

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.155	25	.123	.941	25	.155
Room_L	.125	25	.200[*]	.947	25	.214
Full_L	.107	25	.200[*]	.945	25	.196

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

2. Normality check of minimum viewing distance data

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.205	25	.008	.867	25	.004
Room_L	.197	25	.014	.856	25	.002
Full_L	.156	25	.118	.901	25	.019

a. Lilliefors Significance Correction

3. Friedman Test between lighting conditions for maximum viewing distance,

Test Statistics[a]

N	25
Chi-Square	2.362
df	2
Asymp. Sig.	.307

a. Friedman Test

and minimum viewing distance

Test Statistics[a]

N	25
Chi-Square	.775
df	2
Asymp. Sig.	.679

a. Friedman Test

4. Normality check of highest viewing angle data

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.173	25	.052	.944	25	.186
Room_L	.135	25	.200[*]	.902	25	.020
Full_L	.119	25	.200[*]	.969	25	.630

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

5. Normality check of lowest viewing angle data

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.070	25	.200[*]	.985	25	.961
Room_L	.106	25	.200[*]	.969	25	.620
Full_L	.188	25	.023	.807	25	.000

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

6. Friedman Test between lighting conditions for highest viewing angle,

Test Statistics[a]

N	25
Chi-Square	3.083
df	2
Asymp. Sig.	.214

a. Friedman Test

and lowest viewing angle

Test Statistics[a]

N	25
Chi-Square	.061
df	2
Asymp. Sig.	.970

a. Friedman Test

Appendix E.   Experiment set-up (Experiment 2 - Eye Tribe© eye tracker)

Participants' face photos of Experiment 2

Appendix F.   Statistical analysis of Experiment 2 data

1. Normality check of maximum viewing distance data

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.083	28	.200[*]	.980	28	.852
Room_L	.178	28	.023	.927	28	.053
Full_L	.118	28	.200[*]	.976	28	.757

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

2. Normality check of minimum viewing distance data

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.171	28	.035	.927	28	.052
Room_L	.206	28	.004	.942	28	.126
Full_L	.138	28	.187	.915	28	.026

a. Lilliefors Significance Correction

3. Friedman Test between lighting conditions for maximum viewing distance.

Test Statistics[a]

N	28
Chi-Square	9.406
df	2
Asymp. Sig.	.009

a. Friedman Test

Pairwise comparisons with Wilcoxon signed-rank tests.

Test Statistics[a]

	Room_L - Natural_L	Full_L - Natural_L	Full_L - Room_L
Z	-1.917[b]	-.378[b]	-2.820[c]
Asymp. Sig. (2-tailed)	.055	.706	.005

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.
c. Based on negative ranks.

4. For minimum viewing distance

Test Statistics[a]

N	28
Chi-Square	2.583
df	2
Asymp. Sig.	.275

a. Friedman Test

5. Normality check of highest viewing angle data

Tests of Normality

	Kolmogorov-Smirnov[a]	Shapiro-Wilk

	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.192	28	.010	.914	28	.025
Room_L	.137	28	.192	.932	28	.069
Full_L	.160	28	.065	.949	28	.186

a. Lilliefors Significance Correction

6. Normality check of lowest viewing angle data

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.161	28	.062	.942	28	.126
Room_L	.118	28	.200*	.972	28	.643
Full_L	.078	28	.200*	.989	28	.985

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

7. Friedman Test between lighting conditions for highest viewing angle,

Test Statistics[a]

N	28
Chi-Square	.491
df	2
Asymp. Sig.	.782

a. Friedman Test

and lowest viewing angle

Test Statistics[a]

N	28
Chi-Square	9.135
df	2
Asymp. Sig.	.010

a. Friedman Test

Pairwise comparisons with Wilcoxon signed-rank tests.

Test Statistics[a]

	Room_L - Natural_L	Full_L - Natural_L	Full_L - Room_L
Z	-1.663[b]	-.793[c]	-2.815[c]
Asymp. Sig. (2-tailed)	.096	.428	.005

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.
c. Based on negative ranks.

Appendix G.   Instruction and informed consent in Experiment 3

Instruction and Informed Consent of Eye-tracking experiment for participants on 8th floor building 3 (Ohashi Campus), at Kyushu University, Department of Human Science, Japan.

Dear participant,

Thank you for joining our eye tracking experiment. We would like to measure
1. The calibration quality of the eye-tracker for users with glasses and without glasses,
2. The time to get the results of the calibration.

We expect individual differences based on eye-size and other characteristics. If you agree, we would also like to take a picture of your eyes. There are no (health) risks involved in joining the experiment and bear in mind that you can opt-out of the experiment any time - participation is on voluntary basis. The experiment will take about 40 minutes.

To process the data accurately, we would like to ask you for some information. We will use the information to analyze our data and, possibly, for data publication of group means. However, we guarantee your privacy: your data will be numbered and we will not disclose data of single individuals.

Here are our questions:
1. Do you wear glasses? [yes / no]
2. How thick your glasses? ...... mm
3. Do you wear contact lenses? [yes / no]
4. How tall are you?   ...... cm
5. What is your age?   ......   years old
6. Can we take a picture of your eyes? [yes / no]

Experiment instructions:
1. We will measure the calibration quality for users who wear glasses at a viewing distance of 40 cm. If users can see without glasses, we will measure the calibration quality for users with or without glasses. If users cannot see without glasses, we will measure the calibration quality for users only with glasses, and at the same time we will measure the time it takes for this measurement. Please follow the instructions of the experimenter.
2. We then will measure the calibration quality for users who do not wear glasses at a viewing distance of 40 cm always with or without (replica) glasses, and also at the same time we will measure the time it takes for this measurement. Replica glasses are non-prescription clear glasses. Please follow the instructions of the experimenter.

Thank you for your participation!
Yesaya Tommy Paulus, Gerard B. Remijn

Eye tracking experiment - Written informed consent
Experimenter: Yesaya Tommy Paulus

Your signature on this form indicates that you understand to your satisfaction the information provided to you about your participation in this experiment, and agree to participate as a research participant.

You are free to withdraw from this experiment at any time. You should feel free to ask for clarification or new information throughout your participation.

Participant's Name:

Participant's Signature:

Date:

Appendix H.   Replica glasses used in Experiment 3 and 4

Participants' face photos of Experiment 3 (n = 16)
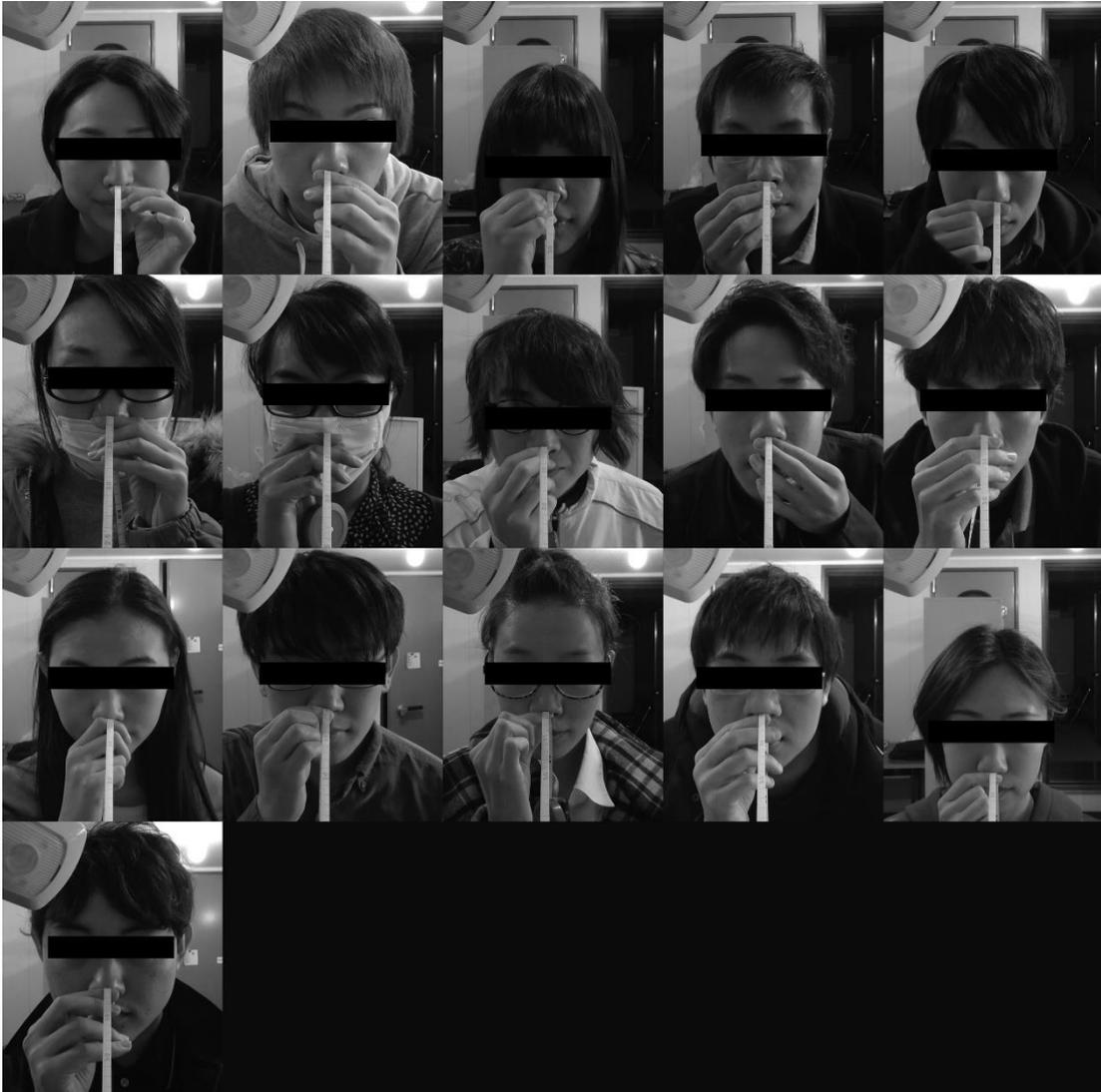
Appendix I.   Example movements of a circle (animation)

appendix\9 point.swf

or
appendix\9 point.exe

Appendix J.   Statistical analysis of Experiment 3 data

1. Normality check of calibration quality with glasses

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.216	16	.044	.846	16	.012
Room_L	.354	16	.000	.649	16	.000
Full_L	.372	16	.000	.698	16	.000

a. Lilliefors Significance Correction

2. Normality check of calibration quality without glasses

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.231	16	.022	.825	16	.006
Room_L	.223	16	.033	.837	16	.009
Full_L	.231	16	.022	.825	16	.006

a. Lilliefors Significance Correction

3. Friedman Test between lighting conditions for calibration quality with glasses,

Test Statistics[a]

N	16
Chi-Square	5.826
df	2
Asymp. Sig.	.054

a. Friedman Test

and calibration quality without glasses

Test Statistics[a]

N	16
Chi-Square	.905
df	2
Asymp. Sig.	.636

a. Friedman Test

4. Normality check of calibration time with glasses

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.263	16	.004	.859	16	.019
Room_L	.185	16	.145	.861	16	.020
Full_L	.178	16	.190	.844	16	.011

a. Lilliefors Significance Correction

5. Normality check of calibration time without glasses

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.

Natural_L	.138	16	.200*	.937	16	.315
Room_L	.322	16	.000	.736	16	.000
Full_L	.221	16	.035	.863	16	.021

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

6. Friedman Test between lighting conditions for calibration time with glasses,

Test Statistics[a]

N	16
Chi-Square	5.375
df	2
Asymp. Sig.	.068

a. Friedman Test

and calibration time without glasses

Test Statistics[a]

N	16
Chi-Square	2.419
df	2
Asymp. Sig.	.298

a. Friedman Test

Pairwise comparisons for calibration quality between glasses conditions under each lighting

Test Statistics[a]

natural	WithoutG - WithG
Z	-1.940[b]
Asymp. Sig. (2-tailed)	.052

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Test Statistics[a]

room	WithoutG - WithG
Z	-2.170[b]
Asymp. Sig. (2-tailed)	.030

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Test Statistics[a]

full	WithoutG - WithG
Z	-2.472[b]
Asymp. Sig. (2-tailed)	.013

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Pairwise comparisons for calibration time between glasses conditions under each lighting

Test Statistics[a]

natural	WithoutG - WithG
Z	-1.704[b]
Asymp. Sig. (2-tailed)	.088

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

Test Statistics[a]

room	WithoutG - WithG
Z	-2.999[b]
Asymp. Sig. (2-tailed)	.003

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

Test Statistics[a]

full	WithoutG - WithG
Z	-3.258[b]
Asymp. Sig. (2-tailed)	.001

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

Appendix K.   Instruction and informed consent in Experiment 4

Instruction and Informed Consent of Eye-tracking experiment for participants on 8th floor building 3 (Ohashi Campus), at Kyushu University, Department of Human Science, Japan.

Dear participant,

Thank you for joining our eye tracking experiment. We would like to measure
1. The angle of the eye-tracker and the display screen for calibration of participants with/without glasses,
2. The time to get the results of the calibration under various angle conditions.

We expect individual differences based on eye-size and other characteristics. If you agree, we would also like to take a picture of your eyes. There are no (health) risks involved in joining the experiment and bear in mind that you can opt-out of the experiment any time. We provide a payment for participation of JPY 1000. The experiment will take about 60 minutes.

To process the data accurately, we would like to ask you for some information. We will use the information to analyze our data and, possibly, for data publication of group means. However, we guarantee your privacy: your data will be numbered and we will not disclose data of single individuals.

Here are our questions:
1. Do you wear glasses? [yes / no]
2. How thick are your glasses? ...... mm
3. Do you wear contact lenses? [yes / no]
4. How tall are you? ...... cm
5. What is your age? ...... years old
6. Can we take a picture of your eyes? [yes / no]

Experiment instructions:
1. Please stand in the middle on front of the screen and don't cross the marking on the floor, relax and take a natural viewing position, and don't move your head even when the angle of the display is changed.
2. We will measure at different angles of the eye-tracker and the display screen <u>for participants with glasses</u>. If you can see without glasses, we will measure the calibration quality for participants with and without glasses. If you cannot see without glasses, we will measure the calibration quality for participants only with glasses. At the same time, we will measure the time necessary for the calibration, but only when your eyes appear on the calibration display. Please follow the instructions of the experimenter.
3. We will measure at different angles of the eye-tracker and the display screen <u>for</u>

participants without glasses with and without (replica) glasses. At the same time, we will measure the time necessary for the calibration, but only when your eyes appear on the calibration display. Replica glasses are non-prescription, clear glasses. Please follow the instructions of the experimenter.

Thank you for your participation!
Yesaya Tommy Paulus, Gerard B. Remijn

Eye tracking experiment - Written informed consent
Experimenter: Yesaya Tommy Paulus

Your signature on this form indicates that you understand to your satisfaction the information provided to you about your participation in this experiment, and agree to participate as a research participant.

You are free to withdraw from this experiment at any time. You should feel free to ask for clarification or new information throughout your participation.

Participant's Name:

Participant's Signature:

Date:

Appendix L.   Experiment set-up (Experiment 4 - Eye Tribe$^©$ eye tracker)

Appendix M. Participants' face photos of Experiment 4 (n = 30)

Appendix N.   Statistical analysis of Experiment 4 data

A Pearson's correlation results between variabels

Correlations

		Height	FirstAngle	SecondAngle
Height	Pearson Correlation	1	.940**	.972**
	Sig. (2-tailed)		.000	.000
	N	30	30	16
FirstAngle	Pearson Correlation	.940**	1	1.000**
	Sig. (2-tailed)	.000		.000
	N	30	30	16
SecondAngle	Pearson Correlation	.972**	1.000**	1
	Sig. (2-tailed)	.000	.000	
	N	16	16	16

**. Correlation is significant at the 0.01 level (2-tailed).

Data Analysis for the first angle data

1. Normality check of calibration quality with glasses

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.313	30	.000	.731	30	.000
Room_L	.302	30	.000	.749	30	.000

a. Lilliefors Significance Correction

2. Normality check of calibration quality without glasses (three participants with prescription glasses were not able to perform the calibration without glasses at their first angle)

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.382	27	.000	.698	27	.000
Room_L	.346	27	.000	.740	27	.000

a. Lilliefors Significance Correction

3. Friedman Test between lighting conditions for calibration quality with glasses,

Test Statistics[a]

N	30
Chi-Square	.077
df	1
Asymp. Sig.	.782

a. Friedman Test

and calibration quality without glasses

Test Statistics[a]

N	27
Chi-Square	.091
df	1
Asymp. Sig.	.763

a. Friedman Test

4. Normality check of calibration time with glasses

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.270	30	.000	.782	30	.000
Room_L	.331	30	.000	.673	30	.000

a. Lilliefors Significance Correction

5. Normality check of calibration time without glasses (three participants with prescription glasses were not able to perform the calibration without glasses at their first angle)

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.171	27	.041	.867	27	.002
Room_L	.204	27	.005	.799	27	.000

a. Lilliefors Significance Correction

6. Friedman Test between lighting conditions for calibration time with glasses,

Test Statistics[a]

N	30
Chi-Square	.034
df	1
Asymp. Sig.	.853

a. Friedman Test

and calibration time without glasses

Test Statistics[a]

N	27
Chi-Square	.615
df	1
Asymp. Sig.	.433

a. Friedman Test

7. Pairwise comparisons for calibration quality between glasses conditions under natural and room lighting (n = 27)

Test Statistics[a]

natural	WithoutG - WithG
Z	-1.915[b]
Asymp. Sig. (2-tailed)	.056

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

Test Statistics[a]

room	WithoutG - WithG
Z	-.032[b]
Asymp. Sig. (2-tailed)	.975

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

8. Pairwise comparisons for calibration time between glasses conditions under natural and room lighting (n = 27)

Test Statistics^a

natural	WithoutG - WithG
Z	-2.423^b
Asymp. Sig. (2-tailed)	.015

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

Test Statistics^a

room	WithoutG - WithG
Z	-1.828^b
Asymp. Sig. (2-tailed)	.068

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

Data Analysis for the second angle data

1. Normality check of calibration quality with glasses

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.238	16	.016	.819	16	.005
Room_L	.350	16	.000	.692	16	.000

a. Lilliefors Significance Correction

2. Normality check of calibration quality without glasses

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.296	16	.001	.796	16	.002
Room_L	.395	16	.000	.601	16	.000

a. Lilliefors Significance Correction

3. Friedman Test between lighting conditions for calibration quality with glasses,

Test Statistics[a]

N	16
Chi-Square	.667
df	1
Asymp. Sig.	.414

a. Friedman Test

and calibration quality without glasses

Test Statistics[a]

N	16
Chi-Square	1.000
df	1
Asymp. Sig.	.317

a. Friedman Test

4. Normality check of calibration time with glasses

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.273	16	.002	.806	16	.003
Room_L	.217	16	.043	.770	16	.001

a. Lilliefors Significance Correction

5. Normality check of calibration time without glasses

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Natural_L	.344	16	.000	.723	16	.000
Room_L	.248	16	.010	.872	16	.029

a. Lilliefors Significance Correction

6. Friedman Test between lighting conditions for calibration time with glasses,

Test Statisticsᵃ

N	16
Chi-Square	2.571
df	1
Asymp. Sig.	.109

a. Friedman Test

and calibration time without glasses

Test Statisticsᵃ

N	16
Chi-Square	.286
df	1
Asymp. Sig.	.593

a. Friedman Test

7. Pairwise comparisons for calibration quality between glasses conditions under natural and room lighting (n = 16)

Test Statisticsᵃ

natural	WithoutG - WithG
Z	-2.124[b]
Asymp. Sig. (2-tailed)	.034

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

Test Statisticsᵃ

room	WithoutG - WithG
Z	-1.709[b]
Asymp. Sig. (2-tailed)	.088

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

8. Pairwise comparisons for calibration time between glasses conditions under natural and room lighting (n = 16)

Test Statisticsᵃ

natural	WithoutG - WithG
Z	-2.435[b]
Asymp. Sig. (2-tailed)	.015

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

Test Statisticsᵃ

room	WithoutG - WithG
Z	-.683[b]
Asymp. Sig. (2-tailed)	.495

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

Appendix O.   Instruction and informed consent in Experiment 5

Instruction and Informed Consent of Grid experiment for participants on 8th floor building 3 (Ohashi Campus), at Kyushu University, Department of Human Science, Japan.

Dear participant,

Thank you for agreeing to participate in today's experiment. We are investigating visual passwords. The goal of the experiment is to assess and evaluate which is the best grid for each visual password format. We would like to obtain your preferences by means of a rating scale.

Note that we would like you to make some "ideal" passwords during the experiment. Please NEVER use a password that you use in daily life.

There are no (health) risks involved in joining the experiment and bear in mind that you can opt-out of the experiment any time - participation is on a voluntary basis. The experiment will take about 60 minutes. If you have a question or problem at any point in our experiment, please do not hesitate to ask the instructor.

To process the data accurately, we would like to ask you for some information. We will use the information to analyze our data and, possibly, for data publication of group means. However, we guarantee your privacy: your data will be numbered, and we will not disclose data of single individuals.

Here are our questions:
1. Do you wear glasses? [yes / no]
2. Do you wear contact lenses? [yes / no]
3. How tall are you?   ...... cm
4. What is your age?   ......   years old
5. Can you join again for the next experiment? [yes / no]

General experiment instructions:
1. The program will be conducted with counterbalance between visual password formats for each participant automatically based on the participant's ID.
2. We will ask you to make (draw or select) some "ideal" passwords on the monitor screen for each grid of the three visual password formats. Please follow the instructions of the experimenter.
3. We will ask you to answer some questions about the password you will make for each grid and ask you to evaluate some assessments for the use of the grids on three visual password formats. Please follow the instructions of the experimenter.

Thank you for your participation!
Yesaya Tommy Paulus, Gerard B. Remijn

Grids Experiment, Experimenter: Yesaya Tommy Paulus - Written informed consent

Your signature on this form indicates that you understand to your satisfaction the information provided to you about your participation in this experiment, and agree to participate as a research participant.

You are free to withdraw from this experiment at any time. You should feel free to ask for clarification or new information throughout your participation.

Participant's Name: Date:

Participant's Signature:

Appendix P.   Experiment set-up (Experiment 5)

Appendix Q.   Visual icons used in Experiments 5 to 7

Appendix R.　The frequency analysis of a sound used in Experiments 5 to 7
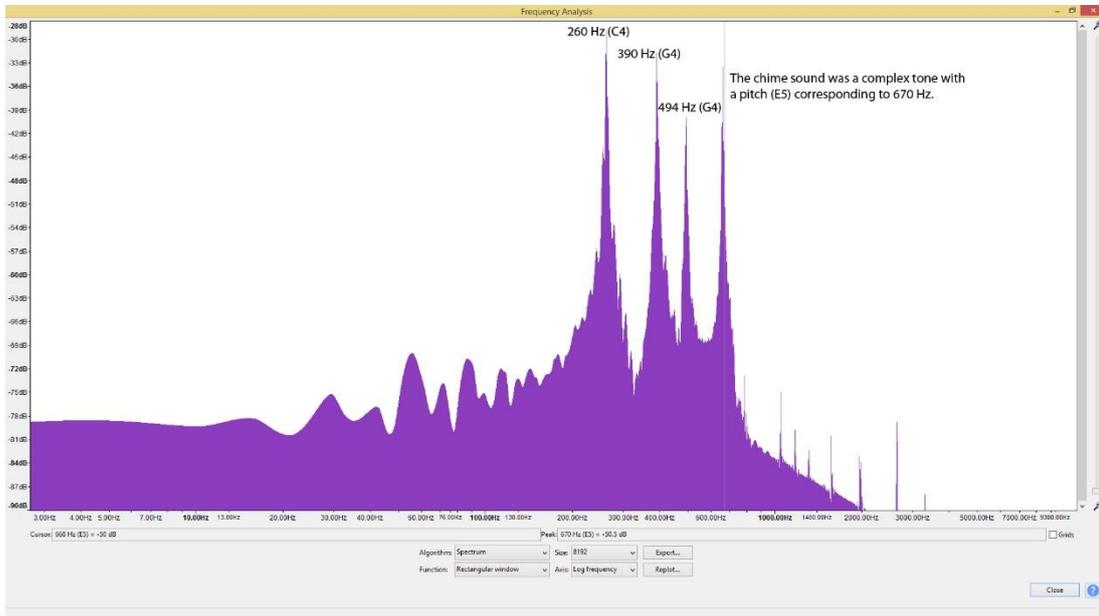
The sound file is in appendix\Windows Background.wav

Appendix S.   Regression analysis of Experiment 5

Easy judgments: regression analysis with a linear function ($y = ax + b$) for all visual password formats

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.922	.851	.847	.405

The independent variable is GC.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	43.017	1	43.017	261.814	.000
Residual	7.558	46	.164		
Total	50.575	47			

The independent variable is GC.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
GC	-.068	.004	-.922	-16.181	.000
(Constant)	6.087	.111		54.900	.000

Easy judgments: regression analysis with a logarithmic function [$y = \ln(x)$] for all visual password formats

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.922	.850	.847	.406

The independent variable is GC.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	42.983	1	42.983	260.456	.000
Residual	7.591	46	.165		
Total	50.575	47			

The independent variable is GC.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
GC	-1.257	.078	-.922	-16.139	.000
(Constant)	8.168	.231		35.369	.000

Safe judgments: regression analysis with a logarithmic function [$y = \ln(x)$] for all visual password formats

Safe judgment

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.973	.947	.946	.378

The independent variable is GC.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	117.924	1	117.924	825.208	.000
Residual	6.574	46	.143		
Total	124.498	47			

The independent variable is GC.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
GC	2.083	.073	.973	28.726	.000
(Constant)	-1.389	.215		-6.464	.000

Safe judgments: regression analysis with a linear function ($y = ax + b$) for all visual password formats

Safe judgment

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.879	.772	.767	.786

The independent variable is GC.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	96.093	1	96.093	155.614	.000
Residual	28.405	46	.618		
Total	124.498	47			

The independent variable is GC.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
GC	.102	.008	.879	12.475	.000
(Constant)	2.304	.215		10.719	.000

Appendix T.   Instruction and informed consent in Experiment 6

Instruction and Informed Consent of grids experiment for participants on 8th floor building 3 (Ohashi Campus), at Kyushu University, Department of Human Science, Japan.

Dear participant,

Thank you for agreeing to participate in today's experiment. We are investigating the use of different grid densities for visual password formats using eye-tracking. The goal of the experiment is to identify whether a particular grid and visual password format is easy to use when authenticating a password with actual eye tracking.

Note that we would like you to register, confirm, and enter (log in) a short or long visual password during the experiment. A password was generated randomly with a minimum length of four (short) and a maximum length of six (long) visual objects for each grid density on three visual password formats. The passwords are set differently to each grid density for each format.

There are no (health) risks involved in joining the experiment and bear in mind that you can opt-out of the experiment any time. We provide a payment for the participation of JPY 4000 upon finishing the experiment, and also some snacks and drinks are provided during the break time. The experiment will take about 360 minutes, divided over 3 sessions. If you have a question or problem at any point in our experiment, please do not hesitate to ask the experimenter. Please follow the instructions of the experimenter.

Pre-experiment instructions:
1. We will ask you to register and calibrate your eyes on Tobii EyeX software at one of the viewing angles.
2. We will ask you to do practice for registering, confirming, entering (log in) a short or long visual password on each grid density for all formats using eye tracking - twice for each format. Note that before entering a password, we will ask you to answer two questions (make an evaluation for each grid density)

Overall, the general instructions are divided into three tasks as shown in the following:
Task 1 (Registration / Confirmation)
1. Please stand in the middle on the front of the screen, don't cross the marking on the floor, relax and take a natural viewing position, and don't move your head during the experiment.
2. We will show and provide a short or long visual password to you for each grid density of three visual password formats.
3. Please memorize the short password in one minute or the long password in two minutes.

4. We will ask you to register the memorized visual password (short or long) by selecting visual objects (alphanumeric characters, dots, visual icons) on the screen for each grid density on three visual password formats using eye tracking.
5. On the same screen, we will ask you to confirm the current password (short or long password) by re-selecting the same visual objects correctly for each grid density on three visual password formats using eye tracking.
   - If the password does not match with your current password, i.e., the one that you memorized, you could retry the confirmation until five times. If you cannot correctly confirm the password five times, you should return to step 4 with a different password (short or long).
   - If your password matches with your current password (correct confirm), you can go to step 6.

Task 2 (Rating-scale judgment)
6. We will ask you to answer on a two 7-point rating scale questions on the screen about:
   - whether you think that a particular grid density will be easy to use for registering and confirming a short or long visual password using eye tracking.
   - whether you think that a short or long visual password will be easy to memorize and recall on each grid density of three formats with eye tracking.

Task 3 (Login)
7. We will ask you to entry (login) to the system by selecting a sequence of visual objects on the screen that matches your current password.
   - If your password does not match with your current password, you could retry the login until five times. If you cannot correctly login within five times, you should return to step 4 with a different password (short or long).
   - If your password matches with your current password (correct login), you can return to step 4 with the next password length or next grid.

8. After finishing all tasks for all grids and formats, please fill in the questionnaire regarding your experience.

Note that you will perform this experiment with counterbalance in order of three visual password formats. Also, the order of password lengths will be counterbalanced among participants.

Thank you for your participation!
Yesaya Tommy Paulus, Gerard B. Remijn

Grids Experiment, Experimenter: Yesaya Tommy Paulus - Written informed consent

Your signature on this form indicates that you understand to your satisfaction the information provided to you about your participation in this experiment, and agree to participate as a research participant.

You are free to withdraw from this experiment at any time. You should feel free to ask for clarification or new information throughout your participation.

Participant's Name: Date:

Participant's Signature:

Appendix U.   Screen interfaces of Experiment 6

Screen interface for Tasks 1 (registration) and 2 (confirmation) with a 5×5 grid.
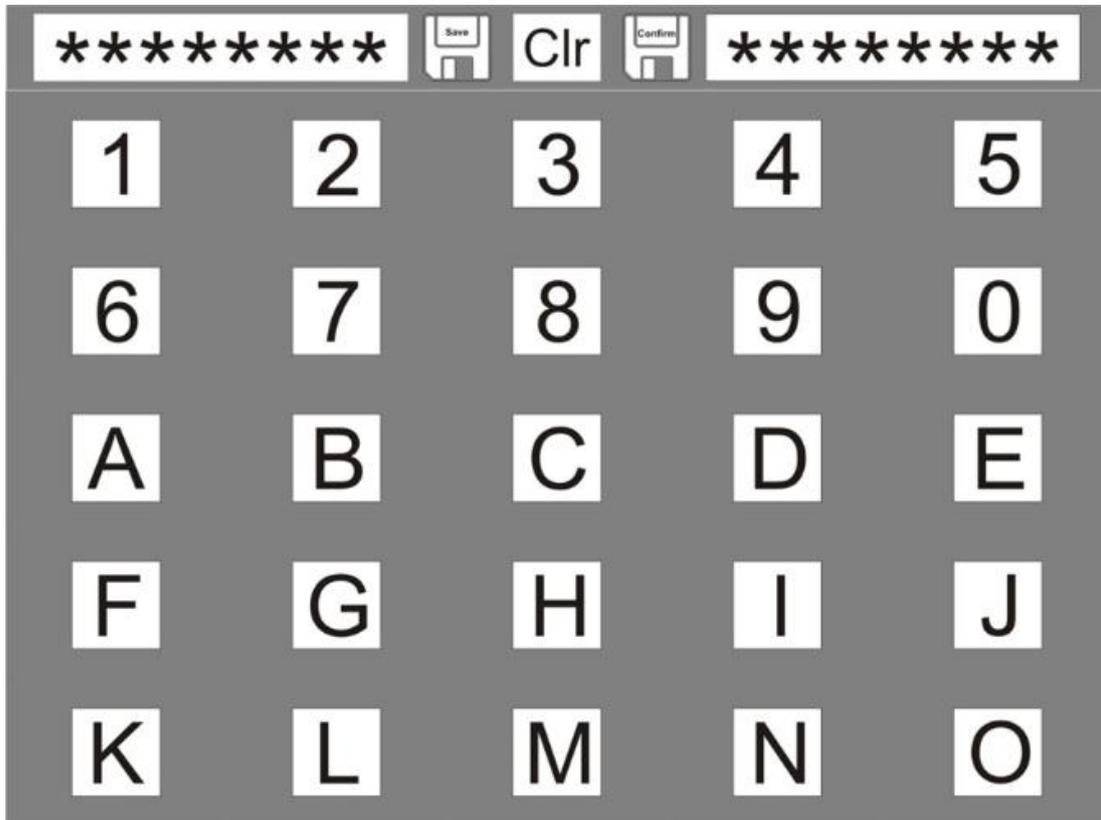
Screen interface for Task 4 (login) with a 5×5 grid.

Screen interface for Task 3 (grid evaluation)

Appendix V.   Sequence memorized visual passwords in Experiment 6

4-object   6-object

Alphanumeric characters

A pattern of dots
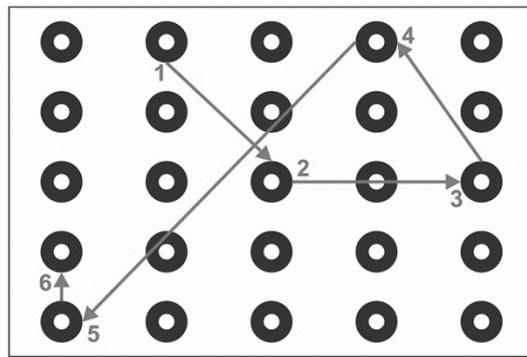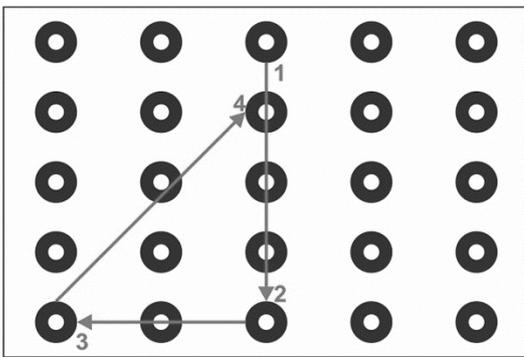
Visual Icons
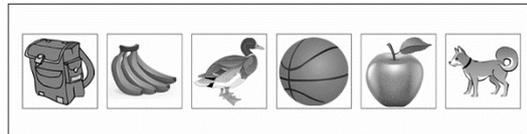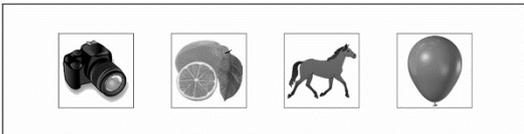
Questionnaire

To process the data accurately, we would like to ask you for some information. We will use the information to analyze our data and, possibly, for data publication of group means.

There are no right and no wrong answers – important to us is that you tell your personal opinion. Probably some possible answers might seem not applicable or appropriate. In this case, please choose the answer that applies the most. Please do not skip any answers and do not think much about each question. Please try to answer spontaneously and as accurately as possible.

All data will be handled in absolute confidentiality, anonymity, and we will not disclose data of single individuals. Here are our questions:

1.   What is your age? _____ years old

2.   How tall are you? _____ cm

3.   What is your gender? [male / female]

4.   Do you wear glasses? [yes / no]

5.   Do you wear contact lenses? [yes / no]

6.   What is your ethnicity? Check (√) only one

    ☐ Asian   ☐ Latino/Hispanic

    ☐ Caucasian   ☐ Other: _____

7.   What is the highest level of education you have completed?

    ☐ High school   ☐ Doctoral degree

    ☐ Bachelor's degree   ☐ Other: _____

    ☐ Master's degree

8.   Are you currently a student or a company employee? [not a student or a company employee]

    If a student, what is your major?

    _____

    If a company employee, what is your occupation?

    _____

9.   Do you use a personal computer (PC) in daily or for work? [yes / no]

10. Please read this statement: "Most problems I find when using a PC, I can fix on my own."

    How does this apply to you (encircle a number): 1 = Always, 2 = Very Often, 3 = Sometimes, 4 = Rarely, 5 = Never

11. Do you have a degree in or are you currently studying toward a degree in an IT-related field (e.g., information technology, computer science, electrical engineering, computer security etc.)? [yes / no]

12. Have you ever (select all that apply):

    ☐ Configured a firewall on a computer

    ☐ Created a database

    ☐ Installed a computer program

    ☐ Written a computer program

    ☐ None of the above

13. Do you have any experience using a text-based password (e.g., a PIN with numbers or letters) in public settings (e.g., on ATMs, on Laptops / PCs, or Smartphones)? [yes / no]

14. Do you have any experience using a visual password in public settings (e.g., selecting visual icons or patterns on ATMs, on Laptops / PCs, or Smartphones)? [yes / no]

    If yes, what experience are those?

15. Do you have any experience using a visual password with eye tracking in public settings (e.g., on ATMs, on Laptops / PCs, or Smartphones)? [yes / no]

    If yes, what experience are those?

16. Do you have any mobile device (e.g., a tab/pad, or a smartphone)? [yes / no]

    If yes, please specify the brand and model of your mobile device

17. If you have a mobile device, which of the following options best describes your screen lock habits?

    ☐ I use a numeric passcode (PIN: Personal Identification Number) to lock/unlock screen on my device

☐ I use a pattern lock screen on my device

☐ I have no lock screen setting on my device

☐ I don't have a mobile device

18. Please give a description of your first impressions regarding the login process during the experiment.

```



```

19. Do you have any strategy to memorize the password on each format? [yes / no]

If yes, please describe your strategy

```
Alphanumeric:


Pattern:


Picture:


```

20. Do you need more time to memorize? [yes / no]

If yes, which password length [short / long]

21. Which visual password format do you think is suitable (best) to use in a real situation with eye tracking? [Alphanumeric / Pattern / Picture]

Please describe your reason

```



```

22. After finishing this experiment, would you use the visual password system with grid and eye tracking to:

☐ login into a computer or a laptop

☐ access an email

☐ login into an online banking system

☐ unlock a screen of the mobile device (instead of entering the PIN)

☐ Other: _____

☐ None of the above

Appendix X.   Friedman Tests for task-completion time

Task 1: 4-object password

Test Statistics[a]

N	16
Chi-Square	14.000
df	2
Asymp. Sig.	.001

a. Friedman Test

Test Statistics[a]

4-object password	Pattern - Alphanumeric	Picture - Alphanumeric	Picture - Pattern
Z	-.625[b]	-3.309[b]	-3.361[b]
Asymp. Sig. (2-tailed)	.532	.001	.001

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

Task 1: 6-object password

Test Statistics[a]

N	16
Chi-Square	16.625
df	2
Asymp. Sig.	.000

a. Friedman Test

Test Statistics[a]

	Pattern - Alphanumeric	Picture - Alphanumeric	Picture - Pattern
Z	-1.603[b]	-3.464[b]	-3.051[b]
Asymp. Sig. (2-tailed)	.109	.001	.002

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

Task 2: 4-object password

Test Statistics[a]

N	16
Chi-Square	6.125
df	2
Asymp. Sig.	.047

a. Friedman Test

Test Statistics[a]

	Pattern - Alphanumeric	Picture - Alphanumeric	Picture - Pattern
Z	-2.689[b]	-2.223[b]	-.155[b]
Asymp. Sig. (2-tailed)	.007	.026	.877

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

Task 2: 6-object password

Test Statistics[a]

N	16
Chi-Square	24.125
df	2
Asymp. Sig.	.000

a. Friedman Test

Test Statistics[a]

	Pattern - Alphanumeric	Picture - Alphanumeric	Picture - Pattern
Z	-3.516[b]	-3.516[b]	-.465[c]
Asymp. Sig. (2-tailed)	.000	.000	.642

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.
c. Based on positive ranks.

Task 4: 4-object password

Test Statistics[a]

N	16
Chi-Square	6.500
df	2
Asymp. Sig.	.039

a. Friedman Test

Test Statistics[a]

	Pattern - Alphanumeric	Picture - Alphanumeric	Picture - Pattern
Z	-1.862[b]	-2.715[b]	-1.500[b]
Asymp. Sig. (2-tailed)	.063	.007	.134

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

Task 4: 6-object password

Test Statistics[a]

N	16
Chi-Square	7.875
df	2
Asymp. Sig.	.019

a. Friedman Test

Test Statistics[a]

	Pattern - Alphanumeric	Picture - Alphanumeric	Picture - Pattern
Z	-.517[b]	-2.948[b]	-2.223[b]
Asymp. Sig. (2-tailed)	.605	.003	.026

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

Appendix Y.   Pearson's correlation analyses for task-completion time

Task 1: 4-object password

Correlations

		Grid_Cells	AN_4object	PA_4object	PI_4object
Grid_Cells	Pearson Correlation	1	.600*	.497	.768**
	Sig. (2-tailed)		.014	.050	.001

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).
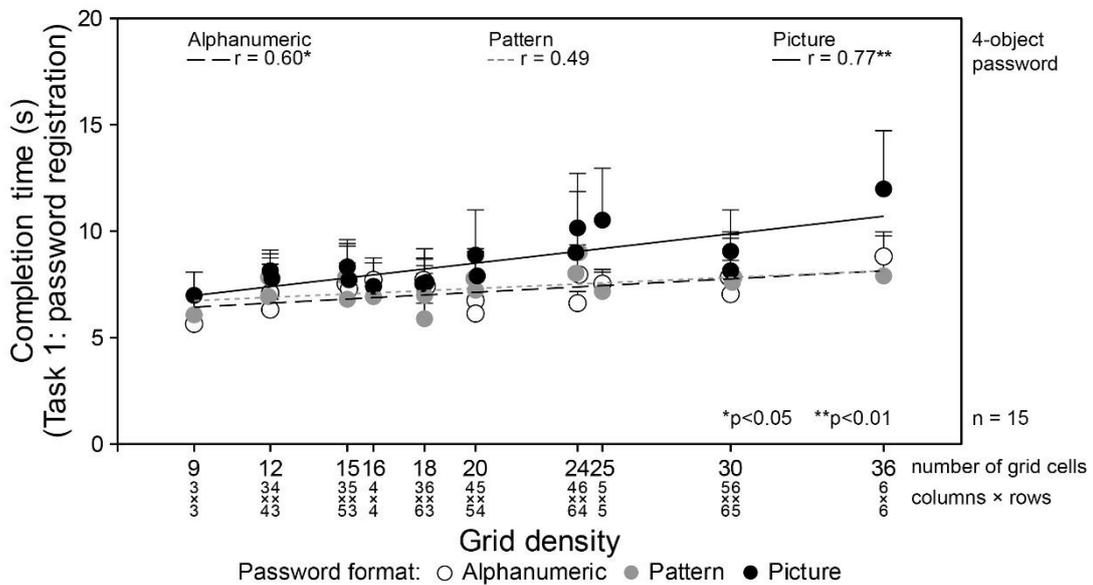
Task 1: 6-object password

Correlations

		Grid_Cells	AN_6object	PA_6object	PI_6object
Grid_Cells	Pearson Correlation	1	.661**	.664**	.814**
	Sig. (2-tailed)		.005	.005	.000

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
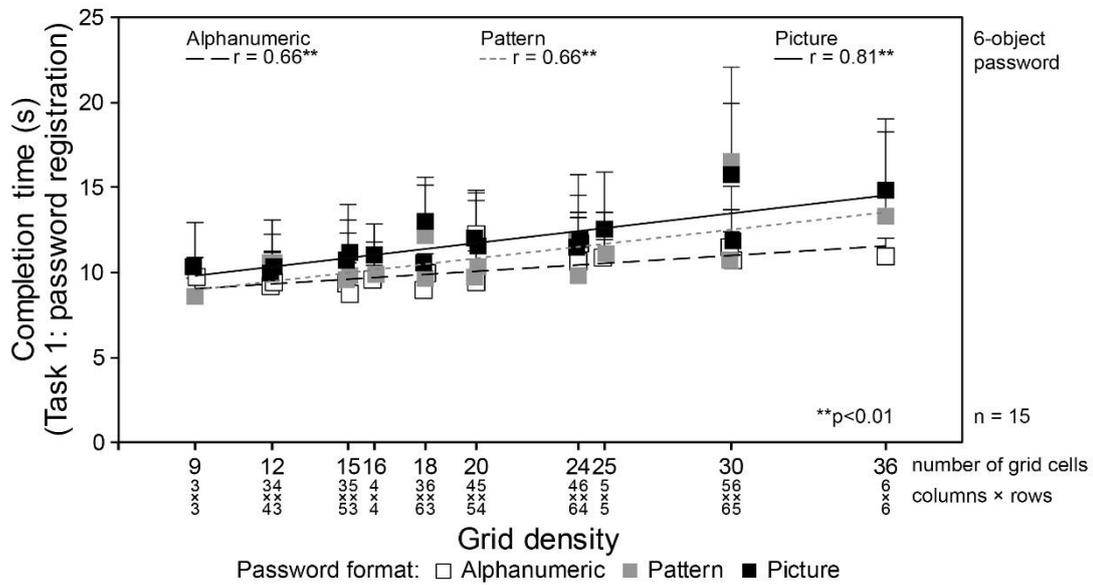
Task 2: 4-object password

Correlations

		Grid_Cells	AN_4object	PA_4object	PI_4object
Grid_Cells	Pearson Correlation	1	.478	.228	.590*
	Sig. (2-tailed)		.061	.395	.016

*. Correlation is significant at the 0.05 level (2-tailed).
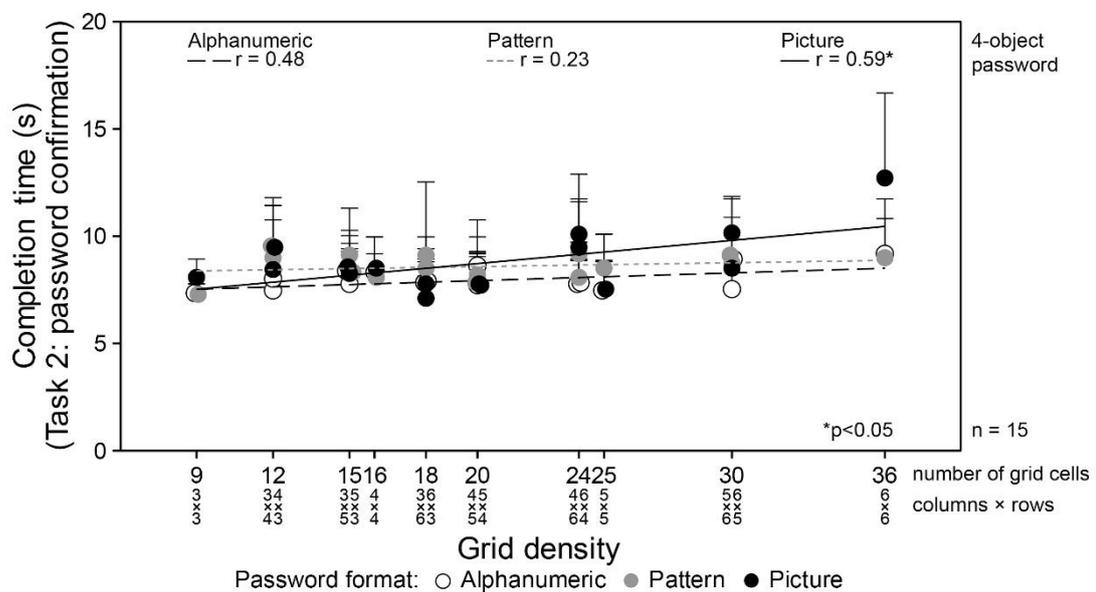
Task 2: 6-object password

Correlations

		Grid_Cells	AN_6object	PA_6object	PI_6object
Grid_Cells	Pearson Correlation	1	.761**	.736**	.895**
	Sig. (2-tailed)		.001	.001	.000

**. Correlation is significant at the 0.01 level (2-tailed).
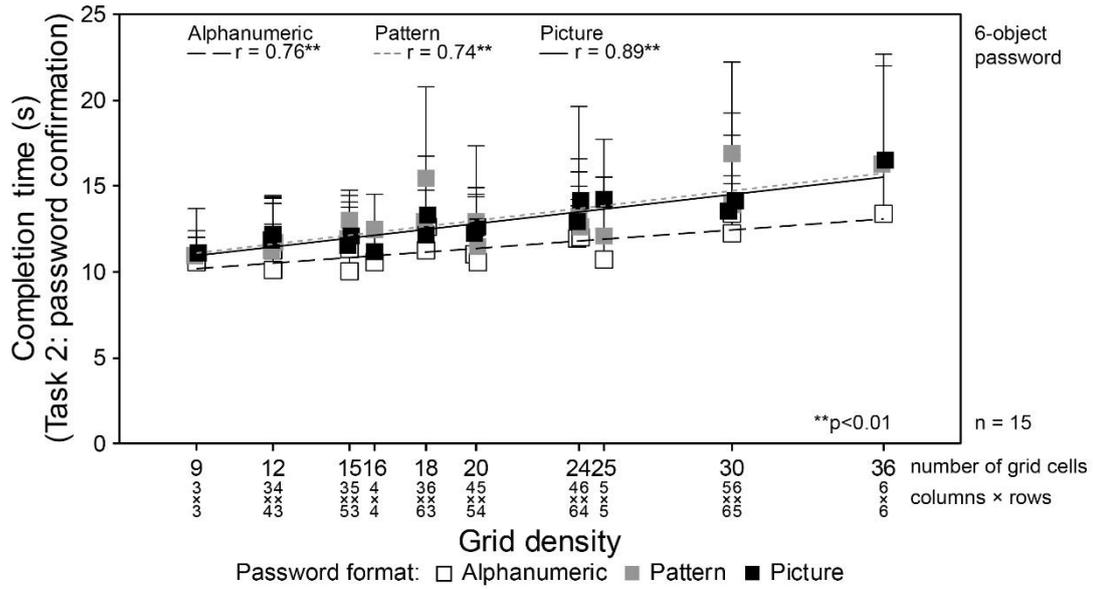
Task 4: 4-object password

Correlations

		Grid_Cells	AN_4object	PA_4object	PI_4object
Grid_Cells	Pearson Correlation	1	.489	.473	.492
	Sig. (2-tailed)		.055	.064	.053

**. Correlation is significant at the 0.01 level (2-tailed).
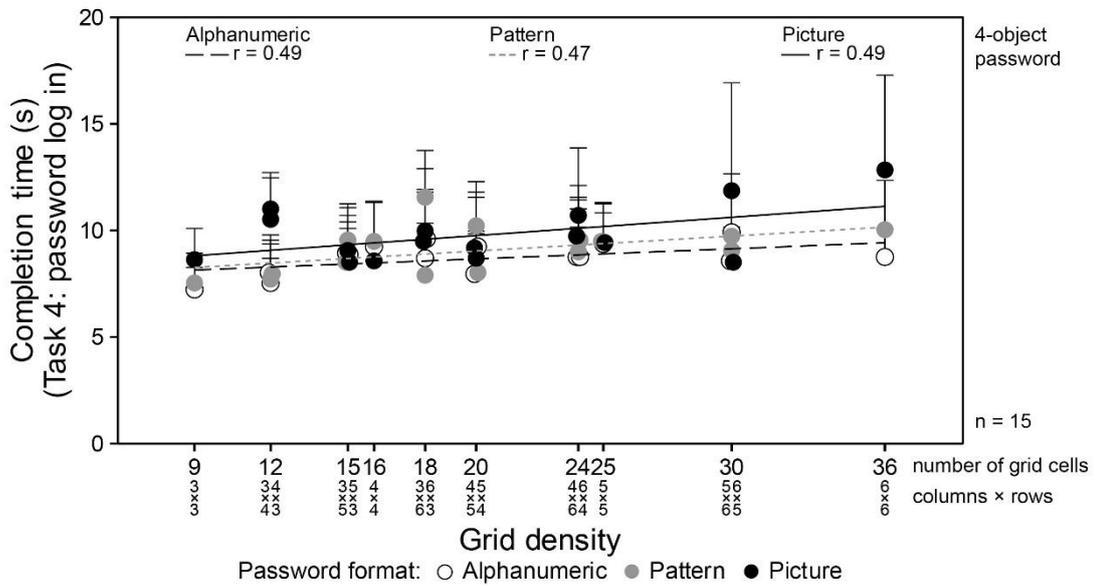
Task 4: 6-object password

Correlations

		Grid_Cells	AN_6object	PA_6object	PI_6object
Grid_Cells	Pearson Correlation	1	.704**	.558*	.070
	Sig. (2-tailed)		.002	.025	.798

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
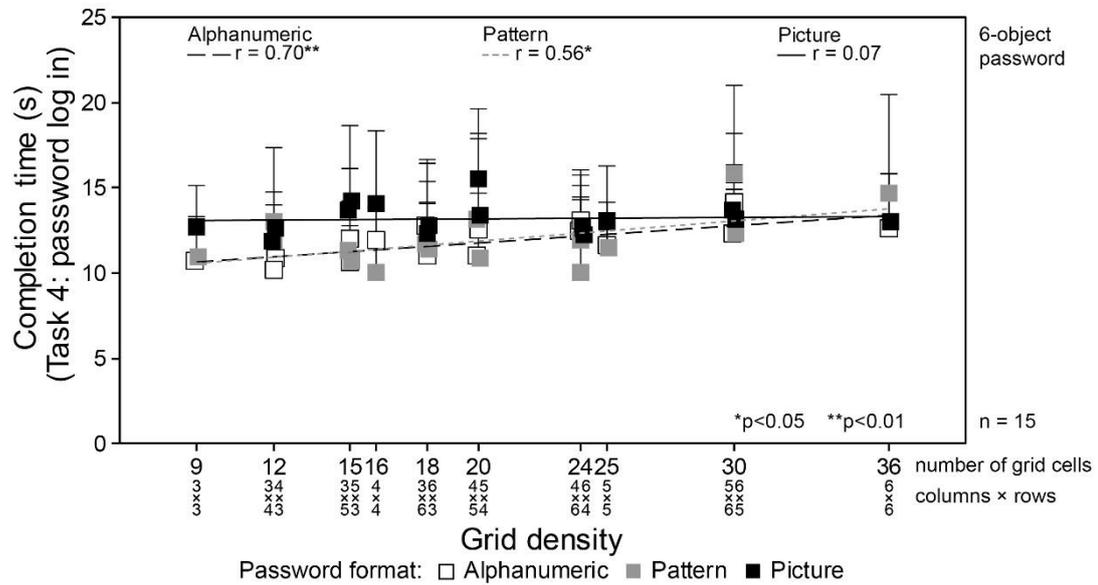
Appendix Z.   Pearson's correlation analyses for task-success rate

Task 2: 4-object password

Correlations

		Grid_Cells	AN_4object	PA_4object	PI_4object
Grid_Cells	Pearson Correlation	1	-.665**	-.644**	-.402
	Sig. (2-tailed)		.005	.007	.123

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
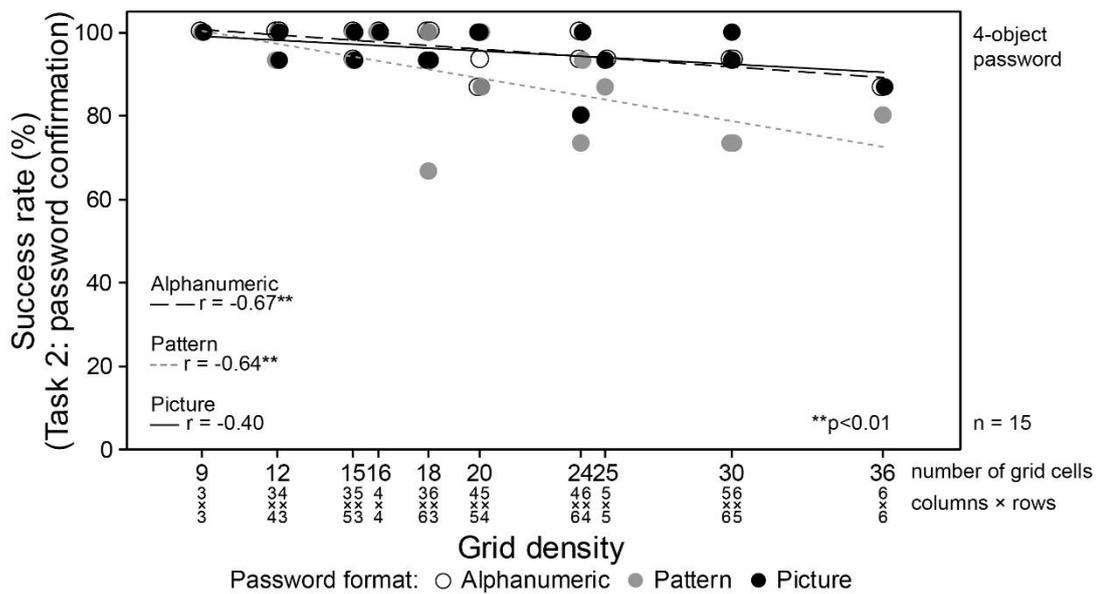
Task 2: 6-object password

Correlations

		Grid_Cells	AN_6object	PA_6object	PI_6object
Grid_Cells	Pearson Correlation	1	-.687**	-.764**	-.662**
	Sig. (2-tailed)		.003	.001	.005

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
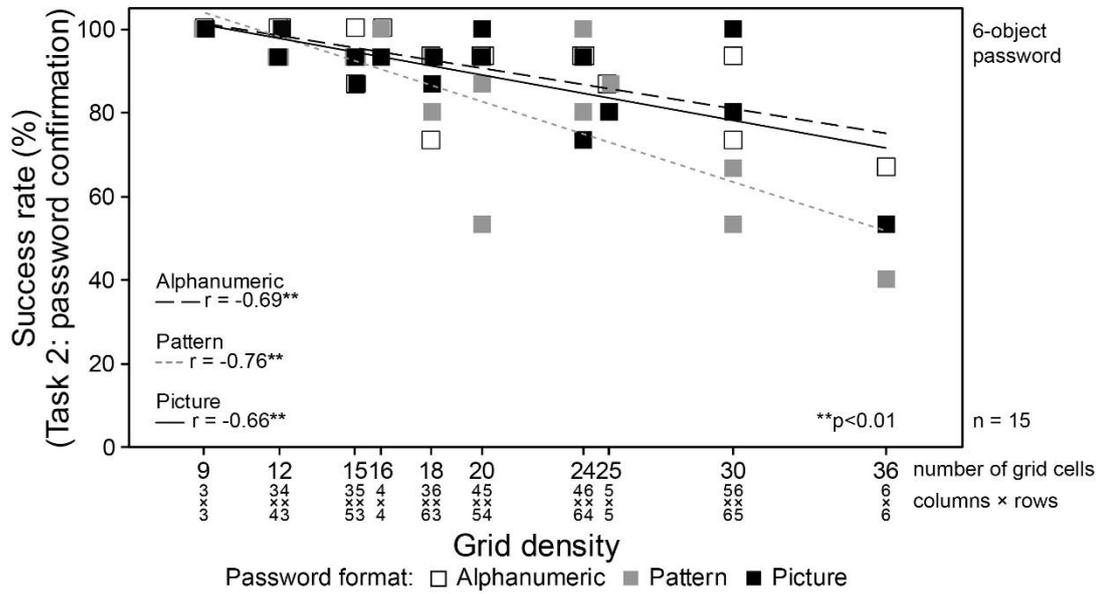
Task 4: 4-object password

Correlations

		Grid_Cells	AN_4object	PA_4object	PI_4object
Grid_Cells	Pearson Correlation	1	-.217	-.326	-.702**
	Sig. (2-tailed)		.420	.219	.002

**. Correlation is significant at the 0.01 level (2-tailed).
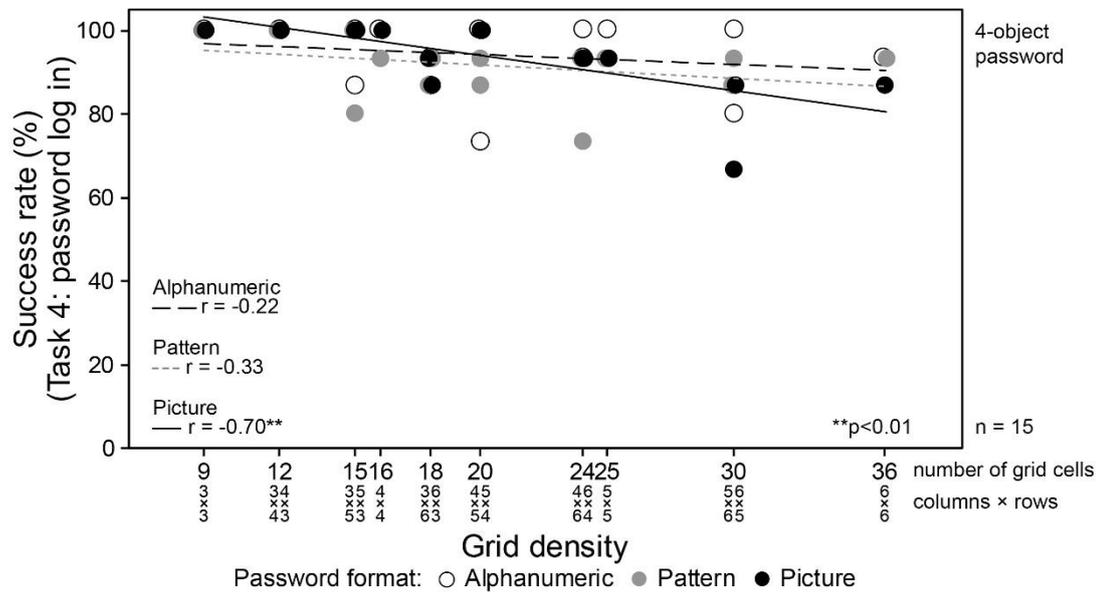
Task 4: 6-object password

Correlations

		Grid_Cells	AN_6object	PA_6object	PI_6object
Grid_Cells	Pearson Correlation	1	-.230	-.659**	-.400
	Sig. (2-tailed)		.392	.006	.125

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
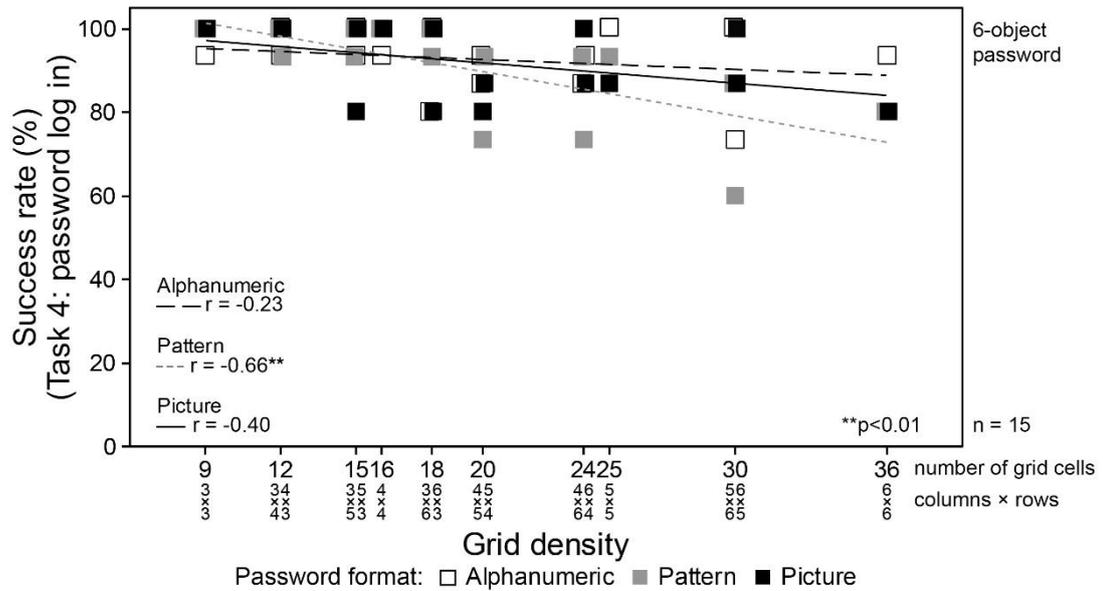
Appendix AA. Pearson's correlation analyses for grid evaluation

Easy-to-use judgment: 4-object password and 6-object password.

Correlations

4-object password		Grid_Cells	AN_4object	PA_4object	PI_4object
Grid_Cells	Pearson Correlation	1	-.814**	-.958**	-.936**
	Sig. (2-tailed)		.000	.000	.000

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

6-object password		Grid_Cells	AN_6object	PA_6object	PI_6object
Grid_Cells	Pearson Correlation	1	-.869**	-.944**	-.903**
	Sig. (2-tailed)		.000	.000	.000

**. Correlation is significant at the 0.01 level (2-tailed).

Easy-to-remember judgment: 4-object password and 6-object password.

Correlations

4-object password		Grid_Cells	AN_4object	PA_4object	PI_4object
Grid_Cells	Pearson Correlation	1	-.614*	-.956**	-.638**
	Sig. (2-tailed)		.011	.000	.008

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

6-object password		Grid_Cells	AN_6object	PA_6object	PI_6object
Grid_Cells	Pearson Correlation	1	-.821**	-.857**	-.787**
	Sig. (2-tailed)		.000	.000	.000

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix AB. User preferences from a questionnaire

	The question number 21 in the questionnaire
ID	Which visual password format do you think is suitable (best) to use in a real situation with eye tracking?
1	Picture, it seems memorable because has a story
2	Alphanumeric, it's easy to remember it and someone can set the password by their memory
3	Picture, it seems sader and easy to use
4	Alphanumeric, we familiar with this format
5	Alphanumeric, The reason is: the pattern is difficult to remember when it's not created by the user self, picture can be defined in many ways so it can be mistaken, and only alphanumeric can be read or remember straightly.
6	Alphanumeric: it's the easiest one to remember, Picture: it's interesting and hard to forget.
7	Alphanumeric, easy/simple/ fast in use but maybe the number of passwords can be added using 7 or 8 digits.
8	Alphanumeric, the pattern and the picture are very interesting but seem to be easy. So I will choose the one not easy to remember.
9	Alphanumeric, because Pattern is difficult to memorize.
10	Alphanumeric is used in daily life so it's easy to use, Picture is easy to memorize (and funny).
11	Alphanumeric, it is easy to memorize for me
12	Alphanumeric, it is easy to remember. Picture, there may be cultural differences
13	Alphanumeric, it is better to change digit's place every time, Pattern is hard to remember by other
14	Alphanumeric is very common in my daily life
15	Picture, it was the easiest one for me to remember. Even after finishing the experiment I could still remember some of the stories.

The alphanumeric format: 12 out of 15.
The picture format: 3 out of 15, and the pattern format was none.

Instruction and Informed Consent of dwell time experiment for participants on 8th floor building 3 (Ohashi Campus), at Kyushu University, Department of Human Science, Japan.

Dear participant,

Thank you for agreeing to participate in today's experiment. The goal of the experiment is to investigate the usability of various dwell times for selecting a sequence of 4 or 6 objects on four different grids with eye-gaze-based input. Dwell time is the gaze time needed to select specific objects on a display, for example by using eye tracking.

Note that we would like you to input a 4-object or 6-object password by using your eye-gaze with a dwelling time of 250 ms, 500 ms, 1000 ms, or 2000 ms, respectively, during the experiment. A password was generated randomly with a minimum length of 4 and a maximum length of 6 visual objects. The passwords are set differently to each grid and each password format. Please NEVER use a password formation strategy that you use in daily life.

There are no (health) risks involved in joining the experiment and bear in mind that you can opt-out of the experiment at any time. We provide a payment for the participation of JPY 1000 upon finishing the experiment, and also some candies are provided during the break time. The experiment takes about 1 hour and 30 minutes. If you have a question or problem at any point in our experiment, please do not hesitate to ask the experimenter. Please follow the instructions of the experimenter.

To process the data accurately, we would like to ask you for some information. We will use the information to analyze our data and, possibly, for data publication of group means. However, we guarantee your privacy: your data will be numbered, and we will not disclose data of single individuals.

Here are our questions:
1. Do you wear glasses? [yes / no]
2. Do you wear contact lenses? [yes / no]
3. How tall are you?   ...... cm
4. What is your age?   ......   years old
5. Did you participate in my previous experiment? [yes / no]

Pre-experiment instructions:
1. We will ask you to register and calibrate your eyes on Tobii EyeX software.
2. We will ask you to do practice for entering a password using eye tracking with a dwell time on a grid and a password format. Note that after entering a password,

we will ask you to answer a question about making an evaluation for each dwell time.

The main instructions:
1. Please stand in the middle on the front of the screen, don't cross the marking on the floor, relax and take a natural viewing position, and don't move your head during the experiment.
2. We will show a 4-object or 6-object sequence randomly generated for each of the three object types, on a grid that was randomly selected from the four different grids. Please memorize the 4-object sequence in one minute or the 6-object sequence in two minutes.
3. After memorizing, we will ask you to enter the memorized 4-object or 6-object sequences by selecting visual objects (alphanumeric characters, dots, visual icons) on the screen by using your eye-gaze with a dwell time of 250 ms, 500 ms, 1000 ms or 2000 ms, respectively.
   - If you select a wrong object, you can use a "Clr" key to clear the selected object.
   - If the object selection was incorrect or selecting objects in the wrong order, you could retry the selection until five times. If you cannot enter the correct object on the fifth attempt, you should return to step 2 with a different password for the same dwell time, password format and grid.
   - If your sequence matches with your current sequence (correct enter), you can restart step 3
4. Finally, we will ask you to answer on a 7-point rating scale question on the screen about whether you think the dwell time used in step 3 as easy to use for object selection with eye-gaze-based inputs.

Thank you for your participation!
Yesaya Tommy Paulus, Gerard B. Remijn

Dwell time experiment, Experimenter: Yesaya Tommy Paulus - Written informed consent

Your signature on this form indicates that you understand to your satisfaction the information provided to you about your participation in this experiment, and agree to participate as a research participant.

You are free to withdraw from this experiment at any time. You should feel free to ask for clarification or new information throughout your participation.

Participant's Name: Date:

Participant's Signature:

Appendix AD. Screen interfaces of Experiment 7

Screen interfaces for selecting a sequence of visual objects (alphanumeric characters, [a pattern of] dots, and visual icons) on a 3×4 grid.
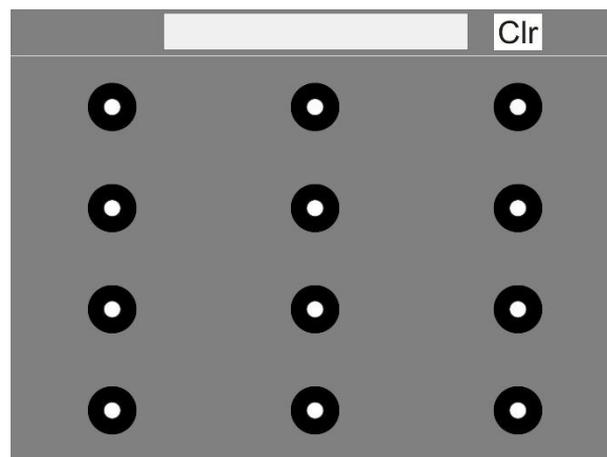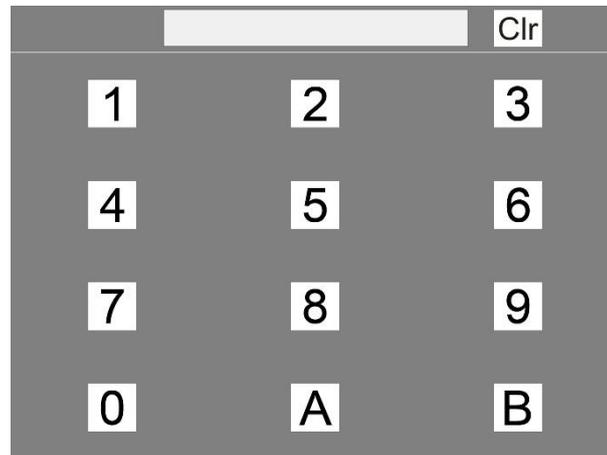
Appendix AE. Sequence memorized visual objects in Experiment 7

Appendix AF. Kruskal-Wallis and Friedman tests

Kruskal-Wallis tests between grid formations (4-object)

Test Statistics[a,b]

	AN250	PA250	PI250	AN500	PA500	PI500	AN1000	PA1000	PI1000	AN2000	PA2000	PI2000
Chi-Square	2.825	3.557	1.018	1.860	.384	1.146	3.755	.241	2.308	4.242	4.691	1.371
df	3	3	3	3	3	3	3	3	3	3	3	3
Asymp. Sig.	.419	.313	.797	.602	.944	.766	.289	.971	.511	.236	.196	.712

a. Kruskal Wallis Test
b. Grouping Variable: GRIDS

Kruskal-Wallis tests between object types (4-object)

Test Statistics[a,b]

	TIME 250 3x4	TIME 250 4x3	TIME 250 4x5	TIME 250 5x4	TIME 500 3x4	TIME 500 4x3	TIME 500 4x5	TIME 500 5x4	TIME 1000 3x4	TIME 1000 4x3	TIME 1000 4x5	TIME 1000 5x4	TIME 2000 3x4	TIME 2000 4x3	ITIME 2000 4x5	ITIME 2000 5x4
Chi-Square	3.578	2.851	.826	1.026	1.005	3.437	1.758	1.499	1.641	.599	3.573	1.326	.524	1.533	.140	.722
df	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	.167	.240	.662	.599	.605	.179	.415	.473	.440	.741	.168	.515	.770	.465	.933	.697

a. Kruskal Wallis Test
b. Grouping Variable: FORMATS

Friedman tests between object types (6-object)

Test Statistics[a]

	250				500				1000				2000			
	3x4	4x3	4x5	5x4	3x4	4x3	4x5	5x4	3x4	4x3	4x5	5x4	3x4	4x3	4x5	5x4
N	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
Chi-Square	2.667	1.167	2.000	2.667	.667	3.167	2.167	.667	.500	4.500	.167	1.167	3.167	3.500	.167	1.167
df	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	.264	.558	.368	.264	.717	.205	.338	.717	.779	.105	.920	.558	.205	.174	.920	.558

a. Friedman Test

Friedman tests between grid formations (6-object)

Test Statistics[a]

	250			500			1000			2000		
	AN	PA	PI	AN	PA	PI	AN	PA	PI	AN	PA	PI
N	12	12	12	12	12	12	12	12	12	12	12	12
Chi-Square	3.300	1.700	5.300	13.400	5.500	1.900	.100	1.500	3.700	3.700	1.500	9.900
df	3	3	3	3	3	3	3	3	3	3	3	3
Asymp. Sig.	.348	.637	.151	.004	.139	.593	.992	.682	.296	.296	.682	.019

a. Friedman Test

For a dwell time of 500 ms: **Test Statistics[a] with Holm Bonferonni**

	Z	Asymp. Sig. (2-tailed)	Rank	Holm-Bonferonni	SIG
AN5x4500 - AN4x3500	-2.432[c]	.015	1	0.008	FALSE
AN5x4500 - AN3x4500	-2.118[c]	.034	2	0.010	FALSE
AN4x5500 - AN4x3500	-1.804[c]	.071	3	0.013	FALSE
AN4x5500 - AN3x4500	-1.255[c]	.209	4	0.017	FALSE
AN5x4500 - AN4x5500	-1.020[c]	.308	5	0.025	FALSE
AN4x3500 - AN3x4500	-.235[b]	.814	6	0.050	FALSE

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.
c. Based on negative ranks.

For a dwell time of 2000 ms: **Test Statistics[a] with Holm Bonferonni**

	Z	Asymp. Sig. (2-tailed)	Rank	Holm-Bonferonni	SIG
PI4x52000 - PI4x32000	-2.589[b]	.010	1	0.008	FALSE
PI5x42000 - PI4x32000	-2.510[b]	.012	2	0.010	FALSE
PI4x52000 - PI3x42000	-2.432[b]	.015	3	0.013	FALSE
PI5x42000 - PI3x42000	-2.197[b]	.028	4	0.017	FALSE
PI5x42000 - PI4x52000	-.471[b]	.638	5	0.025	FALSE
PI4x32000 - PI3x42000	-.314[b]	.754	6	0.050	FALSE

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

Appendix AG. Regression analysis of Experiment 7 with a linear function

Regression analysis with a linear function ($y = a + bx$) for all visual objects
4-object:

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.998	.995	.995	244.711

The independent variable is DT.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	119916792.888	1	119916792.888	2002.495	.000
Residual	598836.855	10	59883.685		
Total	120515629.743	11			

The independent variable is DT.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
DT	4.717	.105	.998	44.749	.000
(Constant)	3288.059	121.466		27.070	.000

6-object:

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.998	.995	.995	328.327

The independent variable is DT.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	231149540.590	1	231149540.590	2144.276	.000
Residual	1077984.022	10	107798.402		
Total	232227524.612	11			

The independent variable is DT.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
DT	6.548	.141	.998	46.306	.000
(Constant)	5192.447	162.969		31.861	.000

Regression analysis with a linear function ($y = a + bx$) for each visual object

The continuous lines show the best-fitting a linear function of the form ($y = a + bx$), in which the correlation between object selection time (4 or 6 objects) and dwell time duration is significantly linear ($p < 0.01$) for each type of three visual objects.

DT: Dwell times, AN_4: Alphanumeric 4-object

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	1.000[a]	.999	.999	111.338350859767

a. Predictors: (Constant), DT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	37615201.278	1	37615201.278	3034.407	.000[b]
	Residual	24792.457	2	12396.228		
	Total	37639993.735	3			

a. Dependent Variable: AN_4
b. Predictors: (Constant), DT

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3363.875	95.721		35.143	.001
	DT	4.575	.083	1.000	55.085	.000

a. Dependent Variable: AN_4

DT: Dwell times, PA_4: (a pattern of) dot 4-object

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.999[a]	.998	.996	228.62068

a. Predictors: (Constant), DT

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	44295089.241	1	44295089.241	847.470	.001^b
	Residual	104534.830	2	52267.415		
	Total	44399624.071	3			

a. Dependent Variable: PA_4
b. Predictors: (Constant), DT

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2903.122	196.551		14.770	.005
	DT	4.965	.171	.999	29.111	.001

a. Dependent Variable: PA_4

DT: Dwell times, PI_4: Visual Icon 4-object

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	1.000^a	.999	.999	107.16005

a. Predictors: (Constant), DT

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	38173959.027	1	38173959.027	3324.309	.000^b
	Residual	22966.554	2	11483.277		
	Total	38196925.581	3			

a. Dependent Variable: PI_4
b. Predictors: (Constant), DT

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3597.181	92.128		39.045	.001
	DT	4.609	.080	1.000	57.657	.000

a. Dependent Variable: PI_4

DT: Dwell times, AN_6: Alphanumeric 6-object

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.998^a	.996	.995	372.53735

a. Predictors: (Constant), DT

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	75856361.254	1	75856361.254	546.578	.002^b
	Residual	277568.151	2	138784.076		
	Total	76133929.405	3			

a. Dependent Variable: AN_6
b. Predictors: (Constant), DT

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.

		B	Std. Error	Beta		
1	(Constant)	5072.193	320.280		15.837	.004
	DT	6.497	.278	.998	23.379	.002

a. Dependent Variable: AN_6

DT: Dwell times, PA_6: (a pattern of) dot 6-object

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.999[a]	.998	.997	281.25200

a. Predictors: (Constant), DT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	83718185.302	1	83718185.302	1058.348	.001[b]
	Residual	158205.373	2	79102.687		
	Total	83876390.675	3			

a. Dependent Variable: PA_6
b. Predictors: (Constant), DT

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5018.676	241.800		20.755	.002
	DT	6.826	.210	.999	32.532	.001

a. Dependent Variable: PA_6

DT: Dwell times, PI_6: Visual Icon 6-object

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.998[a]	.997	.995	344.67283

a. Predictors: (Constant), DT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	71810246.828	1	71810246.828	604.467	.002[b]
	Residual	237598.723	2	118799.361		
	Total	72047845.551	3			

a. Dependent Variable: PI_6
b. Predictors: (Constant), DT

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5486.471	296.325		18.515	.003
	DT	6.322	.257	.998	24.586	.002

a. Dependent Variable: PI_6

Appendix AH. Regression analysis of Experiment 7 with a quadratic function

Regression analysis with a quadratic function ($y = ax^2 + bx + c$) for each visual object
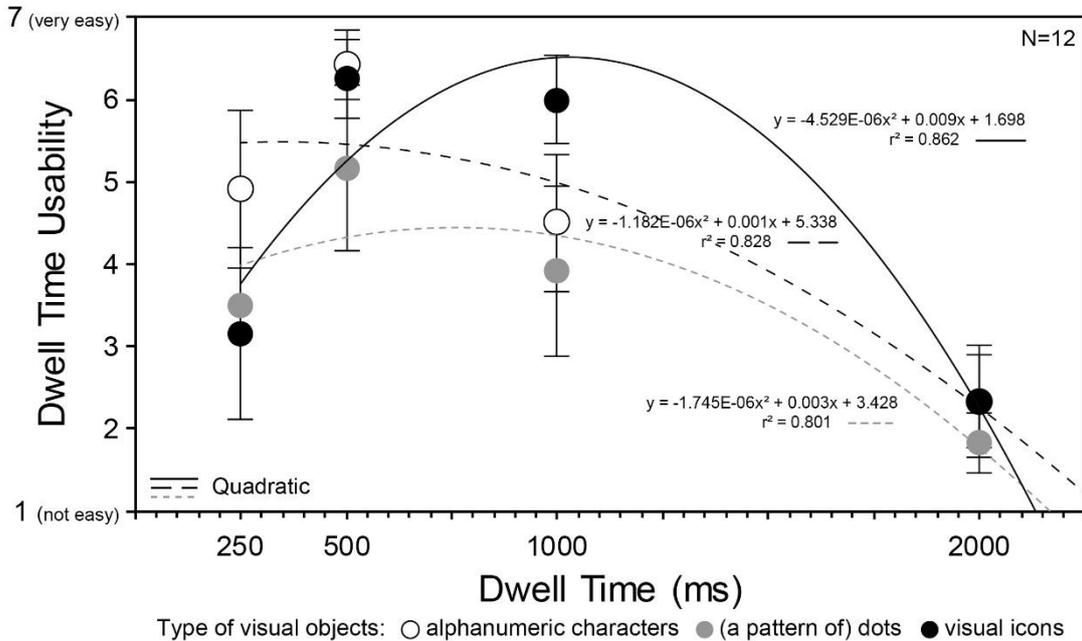
alphanumeric characters

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.910	.828	.485	1.213

The independent variable is DT.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	7.094	2	3.547	2.412	.414
Residual	1.470	1	1.470		
Total	8.564	3			

The independent variable is DT.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
DT	.001	.005	.378	.178	.888
DT ** 2	-1.182E-06	.000	-1.278	-.600	.656
(Constant)	5.338	1.959		2.724	.224

(a pattern of) dots

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.895	.801	.403	1.066

The independent variable is DT.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	4.574	2	2.287	2.012	.446
Residual	1.136	1	1.136		
Total	5.710	3			

The independent variable is DT.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
DT	.003	.004	1.492	.651	.633
DT ** 2	-1.745E-06	.000	-2.310	-1.008	.498
(Constant)	3.428	1.723		1.990	.296

visual icons

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.928	.862	.585	1.276

The independent variable is DT.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	10.146	2	5.073	3.115	.372
Residual	1.629	1	1.629		
Total	11.775	3			

The independent variable is DT.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
DT	.009	.005	3.648	1.909	.307
DT ** 2	-4.529E-06	.000	-4.176	-2.185	.273
(Constant)	1.698	2.062		.824	.561

Regression analysis with a quadratic function ($y = ax^2 + bx + c$) for each visual object through three dwell time durations

alphanumeric characters

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
1.000	1.000		

The independent variable is DT.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	2.038	2	1.019		
Residual	0.000	0			
Total	2.038	2			

The independent variable is DT.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
DT	.016	0.000	5.993		
DT ** 2	-1.312E-05	0.000	-6.448		
(Constant)	1.780	0.000			

(a pattern of) dots

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
1.000	1.000		

The independent variable is DT.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	1.509	2	.755		
Residual	.000	0			
Total	1.509	2			

The independent variable is DT.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
DT	.016	0.000	6.972		
DT ** 2	-1.224E-05	0.000	-6.990		
(Constant)	.300	0.000			

Visual icons

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
1.000	1.000		

The independent variable is DT.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	5.853	2	2.926		
Residual	.000	0			
Total	5.853	2			

The independent variable is DT.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
DT	.025	0.000	5.612		
DT ** 2	-1.709E-05	0.000	-4.957		
(Constant)	-2.047	0.000			

Appendix AI. The Linex loss function

The asymmetric Linex loss function *f(x)* is given by:

$$f(x) = -\exp(ax) + ax - 1$$

where *x* is the loss associated predictive error, *a* is given parameter, and *exp* is the exponential function.

Small (large) curves can be arranged by determining the value of *a*.
If the value of *a* is negative, the continuous curve will be plotted to the right direction, and vice-versa if the value of *a* is negative.

The following figure is an example that shows a continuous curve through the four dwell time durations for three types of visual objects. Furthermore, this function can be used to estimate the peak points of dwell time evaluations, as shown in this figure.

Type of visual objects: ○ alphanumeric characters ● (a pattern of) dots ● visual icons
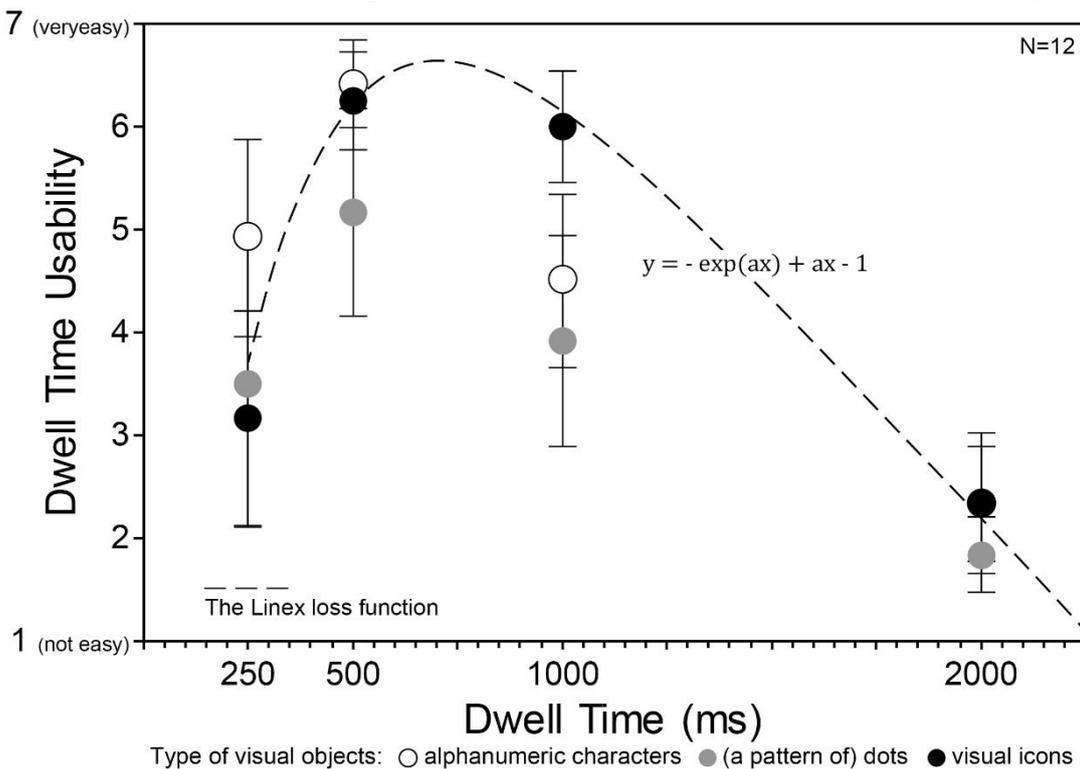
Appendix AJ. Friedman tests for dwell time evaluations

For alphanumeric characters,

Test Statistics[a]

N	12
Chi-Square	25.226
df	3
Asymp. Sig.	.000

a. Friedman Test

Test Statistics[a] with Holm Bonferonni

Alphanumeric	Z	Asymp. Sig. (2-tailed)	Rank	Holm-Bonferonni	SIG
AN2000 - AN500	-3.134[c]	.002	1	0.008	TRUE
AN1000 - AN500	-3.100[c]	.002	2	0.010	TRUE
AN2000 - AN1000	-3.093[c]	.002	3	0.013	TRUE
AN2000 - AN250	-2.662[c]	.008	4	0.017	TRUE
AN500 - AN250	-2.313[b]	.021	5	0.025	TRUE
AN1000 - AN250	-.628[c]	.530	6	0.050	FALSE

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.
c. Based on positive ranks.

For (a pattern of) dots,

Test Statistics[a]

N	12
Chi-Square	18.471
df	3
Asymp. Sig.	.000

a. Friedman Test

Test Statistics[a] with Holm Bonferonni

Pattern	Z	Asymp. Sig. (2-tailed)	Rank	Holm-Bonferonni	SIG
PA2000 - PA500	-2.971[c]	.003	1	0.008	TRUE
PA2000 - PA1000	-2.821[c]	.005	2	0.010	TRUE
PA1000 - PA500	-2.539[c]	.011	3	0.013	TRUE
PA500 - PA250	-2.215[b]	.027	4	0.017	FALSE
PA2000 - PA250	-2.099[c]	.036	5	0.025	FALSE
PA1000 - PA250	-.669[b]	.503	6	0.050	FALSE

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.
c. Based on positive ranks.

For visual icons,

Test Statistics[a]

N	12
Chi-Square	26.838
df	3
Asymp. Sig.	.000

a. Friedman Test

Test Statistics[a] with Holm Bonferonni

Picture	Z	Asymp. Sig. (2-tailed)	Rank	Holm-Bonferonni	SIG
PI2000 - PI1000	-3.089[c]	.002	1	0.008	TRUE
PI2000 - PI500	-3.088[c]	.002	2	0.010	TRUE
PI1000 - PI250	-2.954[b]	.003	3	0.013	TRUE
PI500 - PI250	-2.816[b]	.005	4	0.017	TRUE
PI2000 - PI250	-1.273[c]	.203	5	0.025	FALSE
PI1000 - PI500	-.749[c]	.454	6	0.050	FALSE

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.
c. Based on positive ranks.

Appendix AK. Dwell time usability between object types (Friedman tests)

For dwell time of 250 ms,

Test Statistics^a

N	12
Chi-Square	13.317
df	2
Asymp. Sig.	.001

a. Friedman Test

s

Test Statistics^a with Holm Bonferonni

	Z	Asymp. Sig. (2-tailed)	Rank	Holm-Bonferonni	SIG
PI250 - AN250	-2.969^b	.003	1	0.017	TRUE
PA250 - AN250	-1.853^b	.064	2	0.025	FALSE
PI250 - PA250	-.497^b	.619	3	0.050	FALSE

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

For dwell time of 500 ms,

Test Statistics^a

N	12
Chi-Square	6.200
df	2
Asymp. Sig.	.045

a. Friedman Test

Test Statistics^a swith Holm Bonferonni

	Z	Asymp. Sig. (2-tailed)	Rank	Holm-Bonferonni	SIG
PA500 - AN500	-2.200^b	.028	1	0.017	FALSE
PI500 - PA500	-1.980^c	.048	2	0.025	FALSE
PI500 - AN500	-.707^b	.480	3	0.050	FALSE

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.
c. Based on negative ranks.

For dwell time of 1000 ms,

Test Statistics^a

N	12
Chi-Square	13.282
df	2
Asymp. Sig.	.001

a. Friedman Test

Test Statistics^a with Holm Bonferonni

	Z	Asymp. Sig. (2-tailed)	Rank	Holm-Bonferonni	SIG
PI1000 - PA1000	-2.969^c	.003	1	0.017	TRUE
PI1000 - AN1000	-2.448^c	.014	2	0.025	TRUE
PA1000 - AN1000	-.917^b	.359	3	0.050	FALSE

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.
c. Based on negative ranks.

For dwell time of 2000 ms,

Test Statistics^a

N	12
Chi-Square	3.379
df	2
Asymp. Sig.	.185

a. Friedman Test

Note:

AN: alphanumeric characters, PA: (a pattern of) dots, and PI: visual icons