LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
Department “Institut für Informatik”
Lehr- und Forschungseinheit Medieninformatik
Prof. Dr. Heinrich Hußmann

Bachelorthesis

Comparing the performance of touch, mid-air gestures, and gaze for cue-based and classical authentication on public displays

Hai Ngoai Jens Le
jens.le@campus.lmu.de

Bearbeitungszeitraum:	15.09.2017 bis 07.03.2018
Betreuer:	Mohamed Khamis
Verantw. Hochschullehrer:	Prof. Dr. Floarian Alt

Zusammenfassung

Fortschritte in der Sensortechnologie haben neue Wege der Interaktion mit Public Displays möglich gemacht und Forscher nutzen diese um die Benutzerauthentifizierung zum Beispiel auf Bankautomaten sicherer zu machen. Aber es ist unklar, wie diese Vielfalt an Modalitäten sich miteinander vergleichen lassen im Bezug auf Usability und Security. Wir wollen diese Lücke schließen, indem wir moderne Authentifzierungsmethoden und traditionelle Interaktionsmethoden für Pineingaben auf Situated Displays vergleichen, die Touch, Mid-air gestures und Gaze als Eingabemodalität benutzen. Dazu haben wir zwei Laborstudien durchgeführt. Unsere Ergebnisse zeigen, dass die Studienteilnehmer skeptisch gegenüber Mid-air gestures für Benutzerauthentifizierung in der Öffentlichkeit sind. Die Eingabezeit durch Gaze hängt stark von der Implementierung und der Interaktionstechnik ab. Wir haben SwiPIN, ein modernes und sicheres Authentifizierungsschema, benutzt für den Touch input. Es zeigte sich, dass diese zwar langsamer als die traditionelle Pineingabe war, das Feedback von den Studienteilnehmer aber durchweg positiv war, da es einfach, schnell und sicher war.

Abstract

Advancements in sensing technology have made new ways of interaction with public displays possible and researches use them to make user authentication more secure, for example, on ATMs. But it is unclear, how this variety of modalities compare to each other in terms of usability and security. We want to close this gap by comparing modern authentication methods with traditional interaction methods for PIN entry on situated displays, which use touch, mid-air gestures and gaze as input modality. Therefore, we conducted two lab studies. Our results show that participants are skeptical using mid-air gestures for user authentication in public. Entry time via gaze depends very much on the implementation and the interaction technique. We used SwiPIN, a modern and secure authentication scheme, for the touch input. Entry time of SwiPIN was slower than traditional PIN entry, but the feedback of participants was throughout positive, because it was easy, fast and secure.

Aufgabenstellung

Task Description

A variety of modalities have been proposed for secure user authentication in the past but it still remains unclear how they compare in terms of usability and security.
The aim of this project is to compare several modalities for authentication on public displays.

Tasks

- Conducting a study of an already-implemented system
- Slightly adapting the system according to the outcome of the study
- Conducting a usability study and analyzing the data

Minimum requirements

- Previous experience in conducting user studies
- Basic knowledge in C#
- Previous experience with Eye Trackers and/or Kinect is a plus

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig angefertigt, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

München, March 7, 2018

. . . .

Contents

1 Introduction

An easy and convenient way for user authentication is by traditional personal identification number (PIN) entry, for example, on touch displays of mobile devices. But also automatic teller machines (ATM) or Point-of-sale terminals still rely on PIN entry to approve electronic transactions. But it is known that this authentication method is prone to diverse side-channel attacks like shoulder-surfing [14], thermal attacks [1] or smudge [45]. In the mean time, advancements in technology and sinking costs have made new ways of interaction possible. For example, in addition to touch input, equipping public displays with different cameras and sensors make interaction using the eyes or mid-air hand gestures possible. Prior works tried to design and build secure, human-centric authentication systems by using new input methods for user authentication, for instance, via gaze[13] or mid-air hand gestures [5]. Novel interaction methods might increase the security for user authentication on public situated displays. But this can happen at the cost of usability [33] and eventually be less likely to be accepted by the user if a system is too complicated to handle. Therefore, the choice of input modality for public displays depends on, for example, the use cases, the environment and evaluating the advantages and drawbacks of each modality. But a literature research showed that no prior works give an overview comparison for input modalities via touch, mid-air gestures and gaze in the context of performance for user authentication on public displays.

We address this problem by comparing the performance of cue-based modalities for secure authentication and traditional PIN entry for classical authentication. We explore how the modalities compare to each other and their alternative implementations and present our results and findings.

The cue-based authentication scheme for this work is based on SwiPIN [40] and used touch, mid-air gestures and gaze for input. Originally developed for mobile devices for touch gestures, it was ported to situated displays and was also extended for the use by mid-air gestures and gaze. The detection for mid-air gestures used the Microsoft Kinect, whereas the Tobii eye tracker detected eye movements.

First, we conducted a study with 20 participants of an already implemented cue-based authentication system. To compare it with established authentication schemes via traditional PIN entry, we adjusted the software implementation to classical authentication using aforementioned modalities and conducted a second study with 20 participants. For touch, users would have touched the numbers on the touch screen, for mid-air gestures they would have wave their hands to move a cursor on the display and for gaze users would have fixated a digit with the eyes for a certain dwell time.

The evaluation of the results of both studies show:

- Authentication via SwiPIN is slower than traditional PIN entry but is still a viable option due to acceptable entry times and error rates. Feedback of participants for SwiPIN were also positive, saying it was fast, secure and easy.

- Cue-based mid-air gestures are more secure and perform similar compared to classical mid-air gestures. But participants are skeptical of using mid-air gestures for authentication on public displays, because performing them makes them feel embarrassed.

- Cue-based gaze is not viable yet due to high entry times but gaze using dwell time can be an alternative

The thesis is structured as follows: Some background knowledge on public displays and authentication is provided in chapter 2. Chapter 3 is dedicated to cue-based authentication using touch, mid-air gestures and gaze. In this chapter we present the concepts of the input modalities, followed by a description of the study. The results of the evaluation of the study are then reported. The next chapter 4 uses the same methodology as in previous chapter but now for classical authentication. Chapter 5 report significant differences of cue-based authentication and classical authentication. Afterwards, chapter 6 discusses the results and insights gained from both studies.

Based on the challenges we faced during the studies, chapter 7 proposes future directions and some research questions, which might be of interest. The thesis ends in chapter 8 with a conclusion and a summary of the insights.

2 Related Work

This section should give an overview about background and related work.

2.1 Public Displays

In this subsection we discuss what public displays are and how we can interact with them.

2.1.1 Definition

Public or situated displays are screens varying from small to large. They are deployed mostly in public spaces or in public institutions. They can be used for authentication by using them as touch screen monitors, like ATMs or vending machines. Another main purpose is the display of information. They are used, for example, at airport terminals to show arrival and departure times, or for advertisement purposes. With recent trends in technology like the Internet of Things (IoT) or cloud computing, devices are getting more and more connected. For instance, public displays in different places can be connected with each other to connect two locations [17] or content on public displays can be transferred to mobile devices [28].

2.1.2 Interaction with Displays

Through new innovations and rapid advances in technology and with sinking costs at the same time, new hardware sensors become affordable to the public. Cameras and sensors like eye trackers from Tobii or the Microsoft Kinect One motion sensor, make new interaction modalities possible for public displays [2, 31, 18].

The most common way to interact with public displays is by touch gestures. Previous work [34, 3] explored people's behavior in interacting with large multi-touch displays. Other work focused on guiding users' touch interaction [9] or attracting [27] users on public touch displays.

Due to low costs and wide availability of Microsoft Kinect, this motion sensor has gotten attractive not only for consumers but also for researchers [46]. With the Kinect, interactive items on large displays could now be manipulated not only by touch but also by mid-air hand gestures [42, 41, 44]

Another interaction modality for public displays are gaze-based techniques. Users interact by fixating on objects with their eyes for a certain dwell time [37].

An overview of other possible interaction modalities is provided, for example, in a paper by Müller et al. [31]

2.2 Authentication

The definition of *authentication* in computer systems varies depending on the context used for this term. One definition provided by the IAEA Safeguards Glossary focuses on the aspect of the source of the data and if the data gathered is trustworthy [19].

Another definition focuses on the user: the process of *authentication* is required to assure that the user is the one he claims the be by using mechanism to support that claim. According to [6] *user authentication* on computer systems is described primarily by three factors:

- Knowledge-based factors (*something you know*), which includes personal identification numbers (PIN), passphrases, etc.

- Ownership-based factors (*something you have*), for example one-time password tokens, hardware tokens, bank cards, etc.

- Inherence-based factors (*something you are*), this include biometrics like fingerprints, iris scans, facial recognition, voice recognition, or also behavioral biometrics like keystroke dynamics.

A fourth factor was introduced by Brainard et al. [6]: *somebody you know*, which uses human relationships for authentication. If the provided credentials match with the data in the computer system, the user is granted access to the system. This permission is also called *authorization*.

For this work, we only focus on knowledge-based authentication schemes.

2.3 Challenges

The choice of input modalities for public displays has an influence on the user experience but furthermore, each input modality poses its own challenges. PIN entry by touch proved to be fast and easy in the daily use for authentication on situated displays like ATMs or other devices for electronic payments [40]. But in terms of security and privacy it is known to be prone to different threat models. One of them is shoulder-surfing. It describes a threat scenario, in which the attacker tries to obtain private and sensitive data like passwords or PINs by observing the user. For instance, a user is trying to obtain money from an ATM and has to log in with his credentials. During the authentication process, an adversary is standing behind the him and is looking over his shoulder and can see what the user is typing. The topic of shoulder-surfing was investigated, for example, by Eiband et al. [14]. Other more sophisticated attacks include thermal attacks [1] or smudge [45]. Another aspect is the usability. Sometimes public displays are far away and not easily to reach for the user, or are blocked, for example, by glass windows. So, touch input is in this scenario not suitable.

Interaction via mid-air hand gestures can be a viable option in those cases because users do not need to touch the display. But as stated in [11] there are no standardized gestures for interacting with public displays, thus resulting in many creations of gesture sets. Some of them would lead to embarrassment [15].

Another alternative to touch is gaze-based interaction. But gaze-based input using dwell time is not calibration-free and is known to have the "Midas Touch problem" [30]. For gaze-based interaction, the eyes need to look at an object and manipulate it, thus the distinction of both actions is important. If not, looking at a button, for instance, can simultaneously trigger a button, which leads to unintended action.

But not only the choice of input modality, also deployment of the public displays presents a challenge [29]. Mäkelä et al. proposed to take external factors including weather, events, surroundings, space, inhabitants and vandalism into consideration for deployment [29].

2.4 Comparing Touch, Mid-air Gestures and Gaze

Since our authentication schemes used in this thesis were executed via single-modal input like touch, mid-air gestures or gaze, we wanted to see if previous work already has explored and compared them. A literature research indicates the following:

- Touch vs. Mid-air Gestures: Selection using mid-air gestures is slower and has higher error rates than using touch input [20].

- Touch vs. Gaze: Interaction by touch seem to have better performance and to be more robust than gaze [21], but this was reported in 2009 and technology has advanced since then.

- Mid-air Gestures vs Gaze: Gaze has a better performance than mid-air gestures, but at the cost of more miss clicks and more failures [7].

In this work, we compare the performance of touch, mid-air gestures and gaze in the context of authentication. We also investigate, how cue-based implementations compare to traditional of implementions aforementioned input modalities.

3 Cue-based Authentication

In this section the concept for cue-based authentication are explained, followed by a description of the study and an evaluation of the results

3.1 Concept

The concepts used in this thesis for cue-based authentication using touch, mid-air gestures and gaze are based on prior work [24, 40, 39]. We used an already implemented software implementation, which is also described in [24]. The following subsections give an overview on how the three modalities work.

3.1.1 Touch-based Input

The concept for touch input used is based on SwiPIN. SwiPIN is a state-of-the-art authentication scheme proposed by von Zezschwitz et al. [40] and was originally developed and optimized for mobile devices. It is an approach to make classical PIN entry more resilient against shoulder-surfing and does so by introducing visual cues, which include arrow-like symbols and colors, and simple touch-based gestures. Unlike tapping the PIN numbers directly on buttons as in classical PIN entry, users are required to keep track of the cues first. Then, they have to perform a gesture at specific area depending on the cues.

There are four different arrow symbols: up, down, left and right. These arrows are integrated into the buttons containing the digits. The block of PIN digits is divided into two areas, which are marked by colors red and yellow. Two of the ten digits are colored, but have no arrows. The combination of arrows and colors are then mapped to each PIN digits and they indicate which gesture to perform. The design of SwiPIN allows five distinct gestures: swipe to the left, right, up, down and tap. Figure 3.1 shows an example of a possible assignment of visual cues to PIN digit. To enter number 1, for instance, the user has to make a swipe gesture from the left to the right starting in the red box beneath the PIN keypad. If there are no arrows shown, the user is required to tap the correct box. For instance, for number 5 the user has to tap the red box. After entering each digit, the cues are shuffled and assigned again randomly to the PIN digits.

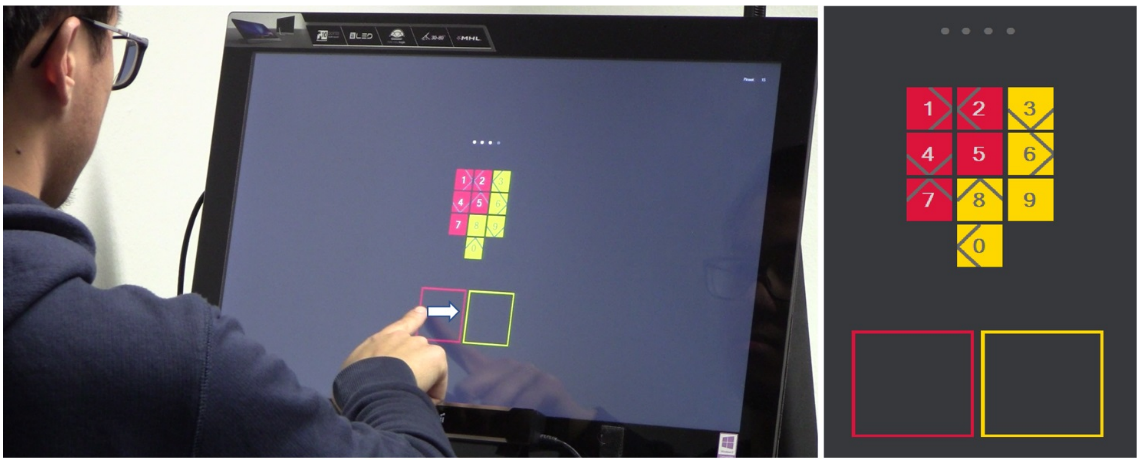


Figure 3.1: Touch-based authentication

3.1.2 Mid-air Gestures-based Input

Prior work like [41, 42] have already explored mid-air gestures as an interaction modality for public displays. Mid-air gestures have inherent advantages in scenarios, where only distant interaction with displays are possible or where user have no direct access to the screen. Barriers like glass windows are, for instance, placed between user and display. Another benefit is the hygienic aspect, because no touching on the display surface is needed.

To use mid-air gestures for authentication, the concept of touch-based SwiPIN has been ported to and adjusted for the use of mid-air hand gestures. At the beginning before entering any number, the user has to hold their hands in a starting position, (see Figure 3.2). Like in touch-based input, the digits contain cues in form of arrows and colors, which separate the PIN pad in two areas. Red indicates that the user has to use the left hand, and the color yellow to use the right hand. Furthermore, the arrows give a hint in which direction the user has to perform the mid-air swipe gesture with the hand. If there are no arrows, the user is required to push the hand forward. For example, if the user wants to enter number 5, he is required to use the left hand and make a mid-air gesture upwards. After performing the gesture, the user has to move the hand back to the starting position. Also, the visual cues of arrows are randomly assigned to the digits again.

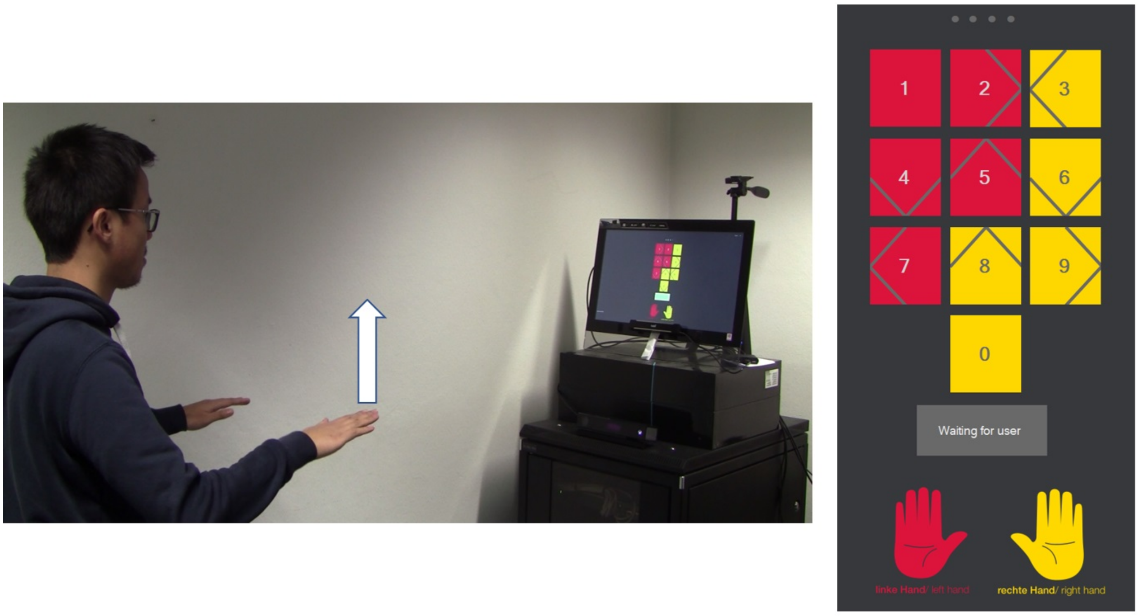


Figure 3.2: Mid-air gestures-based authentication

3.1.3 Gaze-based Input

In order to detect eye-based input for cue-based authentication, an interaction technique called Pursuits [39] was used for this modality. Eye movements can be distinguished into different types of movements, Møllenbach et al. [30] categorized eye movements in three categories: Fixations, Saccades and Smooth Pursuits. Our authentication method uses last-mentioned category. In principle, Pursuits tries to match the trajectory of smooth pursuit eye movements with the trajectory of the moving stimulus. A correlation function is then used to determine how similar gaze and stimulus trajectories are. The challenge is to determine a threshold value, that balances accuracy and speed. A low threshold value leads to faster interaction, but also a higher rate of false detection. A higher threshold has the opposite effect: slower recognition time, but higher accuracy. A side benefit of using Pursuit is, that no calibration is needed.

Therefore, unlike the aforementioned touch-based and mid-air based modalities, authentication by gaze does not use arrows or colors as cues, but instead it uses moving objects (the digits itself) as visual cues, similar to the work by Cymek et al. [10]. The digits are moving in linear, circular and sinusoidal trajectories 3.4. The Linear and the circular ones in particular were also studied in prior work by Khamis et al. [22]. A variation of those basic trajectories in direction path make it possible to distinguish the ten PIN digits. In order to select and enter a number, user would have to follow the desired digit with their eyes. For instance, if the user wants to enter digit 5 like shown in Figure 3.3, he has to follow digit 5 in a circular trajectory.

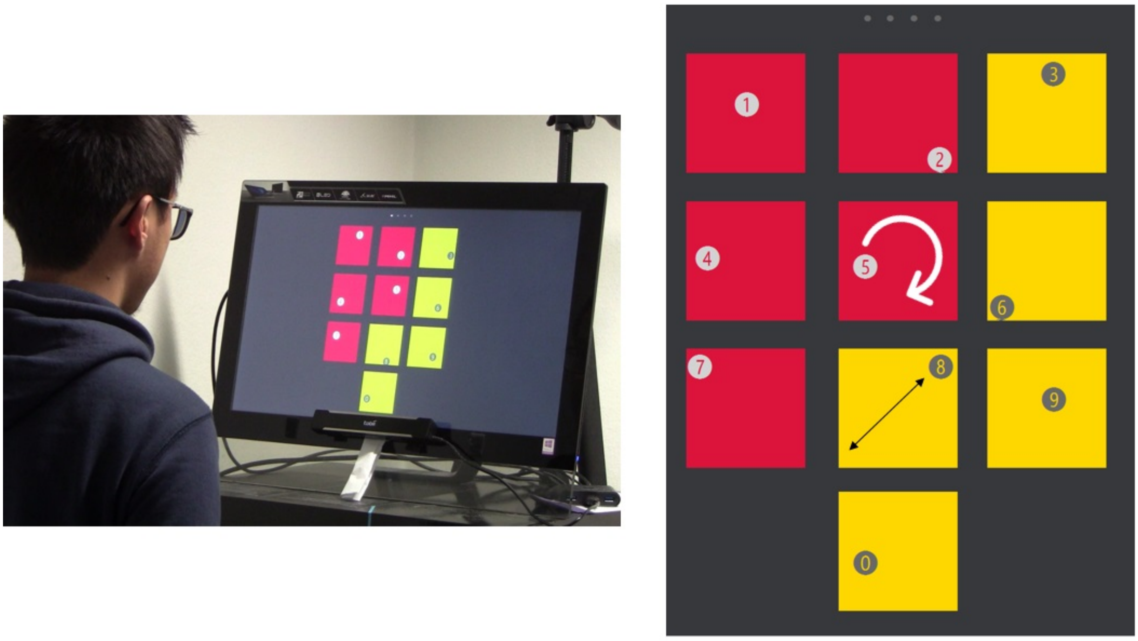


Figure 3.3: Gaze-based authentication. Note: Arrows are not part of the user interface and are only shown for illustration.

3.2 Study

In order to understand the performance of the different aforementioned authentication methods, we conducted a usability study. The study was conducted between the September 1st 2017 and September 6th, 2017 in the laboratory of the media informatics institute of Amalienstrasse 17. The aim of the study was to reconduct the experiment from a previous work [24] using a more advanced and more accurate eye tracker.

3.2.1 Experimental Design

The field study was conceptualized as a mixed design, which means two studies were conducted, one for cue-based authentication and another for classical authentication (see Chapter 4.3) with two different groups of participants. There are two independent variables: On the one hand there is *modality* as the within-subject factor. Modality had three characteristics, which could either be *touch, mid-air gestures* or *gaze*. On the other hand, the two groups were distinguished by the between-subjects factor *mode* with the characteristics *cue-based* or *classical*. Each study itself was carried out as a repeated measure experiment. Every participant had to perform all three conditions: touch, mid-air gestures and gaze. Only one of the modes (for the first study only cue-based) had to be performed. Therefore, a session was divided into three blocks. In each

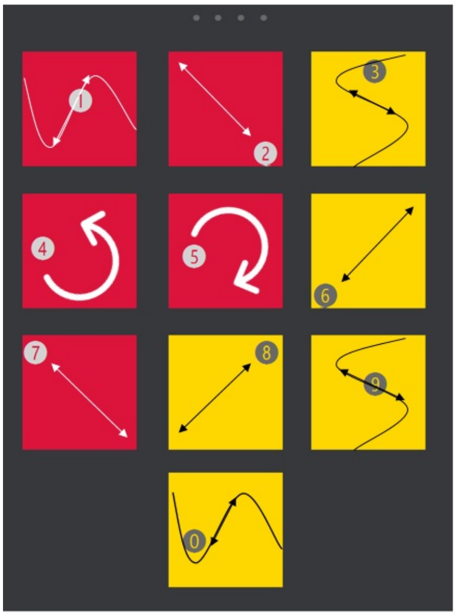


Figure 3.4: Trajectories used for gaze-based interaction using Pursuits. Note: Arrows are not part of the user interface and are only shown for illustration.

block, one of the three authentication methods was chosen. The selection was counterbalanced by Latin square 3.1 to avoid influence from scheduling of the modalities, which can potentially distort the results of the quantitative analysis. Participants had to complete all three blocks and for each modality, 16 PINs had to be entered. After completing a block, the participant filled in a non-weighted NASA Task Load Index (TLX) questionnaire to measure the perceived workload. Once the participant had finished all three blocks of modalities they had to fill in a questionnaire with 5-point Likert scale and open-ended questions about the three input methods. If there was sufficient time left, an optional short semi-structured interview was conducted to gather more specific insights about the performed authentication methods.

mid-air gesture	gaze	touch
gaze	touch	mid-air gesture
touch	mid-air gesture	gaze
touch	gaze	mid-air gesture
mid-air gesture	touch	gaze
gaze	mid-air gesture	touch

Table 3.1: Selection of modalities was counterbalanced by Latin square. Each participant was assigned to one of the rows.

3.2.2 Apparatus

For detecting gestures from input sources like touch, mid-air hand gestures and eye movements, different input devices were used to perform the three conditions of the study. An Acer T232HL monitor was used as an authentication display, which showed the user interface. The resolution of the LCD was 1920x1080 pixels with measurements of 23-inch widescreen, 60Hz refresh rate and 5ms reaction time. Another feature of the screen was the 10-point multitouch capability. Hence, it also served as a device for touch-based inputs. In order to trace the eye movements of a person for gaze inputs, the Tobii REX eye tracker with a sampling rate of 30Hz was used. It was attached and fixated beneath the screen with a velcro fastener. A cardboard also helped to adjust the angle of the eye tracker to different heights of a person to ensure correct detection of the eyes, since different users have different body heights and the tracking device having a limited operating distance. In addition, a Microsoft Kinect One for Windows motion sensor with 30Hz sampling rate was utilized for mid-air hand gesture detection. A laptop of the Lenovo ThinkPad T540p series with Windows 8 was connected to a to the monitor via VGA and USB 3.0 connection. It was equipped with the Microsoft Visual Studio 2015 as integrated development environment (IDE) and had all device drivers and software development kits (SDK) already installed, like the Tobii Core SDK for the eye tracker and the Kinect for Windows SDK 2.0. The software implementation was taken from [24], which used C# as the programming language and was executed with the help of the IDE. The technical setup can be seen in Figure 3.5.

Participants had to perform interactions at a distance of 60 cm for touch and gaze condition, 145 cm for the mid-air gestures. Markers were put on the floor to ensure that the participant sticked to these distances. Predefined button sizes containing the PIN digits (64 pixels for touch, 152.3 pixels mid-air gesture) combined with the interaction distances resulted in 1.59 degree in visual angles. For the gaze modality the visual angle of 4.97 degree was higher for the button size, which was 200 pixels. The reason was to make sure that there was enough space for the trajectories [24].

In addition, a second laptop was used to fill in the NASA TLX and the questions in Google Forms via a browser by the participant.

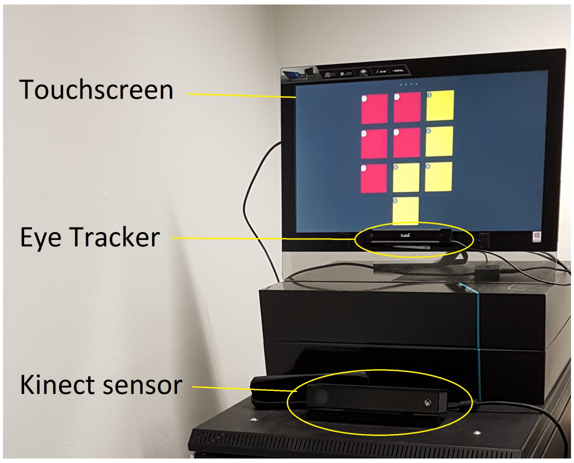


Figure 3.5: Technical setup for the study

3.2.3 Participants

Twenty people participated in the first study, seven of them were male and 13 were female. They were recruited mainly through a mailing list for user studies by the university. The age ranged from 18 to 33 years with an average of 24.1 years (SD = 3.9) and the body height ranged from 147cm to 183cm (M = 169.2cm, SD = 9.0). All participants had normal vision or had been corrected to normal vision by visual aid. In detail, this means that seven of them wore eye glasses during the experiment, three people used contact lenses and the rest did not use any eye-sight correction.

Eight out of 20 participants had no prior experience with motion sensors, the rest had already used them from video game consoles like Nintendo Wii, Microsoft Kinect or through other studies. Half of the 20 subjects had no previous experience with eye trackers, the other half had experience through devices like Google Glass or Tobii eye tracker from previous studies.

Participants received a 10 Euro Amazon Voucher or course credits for the 60-minute-long study as compensation for their participation.

3.2.4 Procedure

Upon arrival participants were welcomed to the study and were requested to sign a consent form. Before they could start testing and using the system, the instructor explained why the study was conducted. It was also put emphasis on understanding the general idea of SwiPIN (see 3.1.1), which all of the three cue-based authentication modalities are based on. Not apprehending the idea of switching the cues would skew the results of the qualitative evaluation, because participants would miss the aspect of added security.

Afterwards, a modality was chosen by using Latin square 3.1. Depending on the selection of the condition, the instructor presented the authentication method to the participant first by opening the application window and showed them the user interface. He then explained how to interact with it in order to insert numbers. In case of the gaze condition, an additional calibration step for the eyes was added to improve the performance. This was optional, because the software implementation used smooth pursuit eye movements to enter digits, which made calibration theoretically unnecessary. But during the study, it turned out that if not calibrated, the selection time for digits would have increased significantly or even led to no selection at all. In most cases this issue improved after a new calibration.

Participants entered about three to five PIN codes for training to get familiar with the respective modality. During these test runs, the participant was also corrected, if they performed the method not accordingly. Data points gathered from these runs were excluded from the analysis.

Followed by the training was the actual task with 16 authentication attempts for each modality. The typical authentication attempt is described as follows: The application announced a four-digit PIN code through a speech API verbally. The PINs were chosen by a predefined set of random PINs. After the announcement, a beep sounded, which signaled the start for the user to enter the password. Errors were only shown after having entered all four digits. This means that they had only one chance to enter the PIN code and could not correct wrong digits. If a wrong number was entered, participants had to continue with the next digit.

When participants finished a block, they had to fill in a NASA TLX to evaluate and measure the perceived workload. The template used for the NASA TLX can be found online [16]. This procedure was repeated for all three modalities.

After being exposed to all conditions, participants had to fill in a questionnaire with a 5-point Likert scale and open-ended questions about the three modalities. If there was sufficient time left, a short semi-structured interview was performed, where users could add additional thoughts and comments.

3.2.5 Measures

For each participant, three logfiles were created for each modality. These logfiles contained, for example, participant identification numbers, position of the current PINs, expected PIN and entered PIN, start and end timestamps for an authentication attempt. From these data entries, the Levenshtein distance was calculated to see how far off the correct and the entered PINs were. Furthermore the successful entry rate, which gives the percentage of correct PINs of all entered PINs, was measured. Both, the NASA TLX scores and the questionnaire with 5-point Likert scale and open questions were saved on excel sheets.

3.2.6 Limitations

Due to calibration problems with participant 15 in the gaze condition and unusual high error rates in the mid-air gesture condition from participant 8 compared to the others, two data sets were

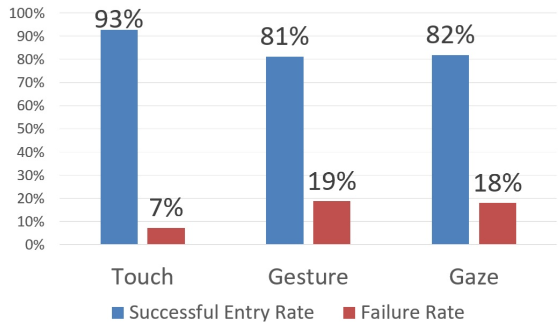


Figure 3.6: Success rate of cue-based authentication

excluded from further quantitative analysis. These data sets were excluded for successful entry rate, Levenshtein distance and entry time.

3.3 Analysis and Results

For chapters 3.3 and 4.5 we used a repeated measure Analysis of Variance (ANOVA) test, whereas for chapter 5 we used a Two-way mixed ANOVA. The ANOVA tests used SPSS for statistical analysis. In the following the results of the study are presented.

3.3.1 Successful Entry Rate

The successful entry rate of a modality is the percentage of all authentication attempts that users completed successfully. In order to calculate the success rate, we compared the entered PIN with the expected PIN. Then all successful entries were summed up. Then the resulting sum was divided by the total number of entered PINs. After removing two participants. For each modality we collected 288 valid data points (18 participants, 16 PIN entries each). Touch contained 21, mid-air gestures 54 and gaze 52 authentication failures.

After pairwise comparison of the results (see Figure 3.6) we can see that authentication attempts using the touch-based method yielded in the best success rate with 93%, followed by gaze with 82% and mid-air gestures with 81%.

Two key factors influenced the outcome of the success rate. On the one hand software implementation and hardware used in the studies played an important role, because recognizing the users intended input correctly depended on these devices working and being calibrated appropriately. On the other hand, it was also crucial that user performed the conditions properly. Both factors were tested in the training runs before the actual task and, if it was possible, improved.

3.3.2 Levenshtein Distance

The Levenshtein distance is a distance function to measure, how similar two strings are. In our case, we wanted to measure, how similar the entered PIN with the expected PIN was. For a 4-digit code a Levenshtein distance of 0 means that both PINs match each other and therefore are equal, whereas a Levenshtein distance of 4 means that no digit is matching. So for example, the expected PIN code is 1337. If the user entered 1347 or 1338, we would have a Levenshtein distance of 1. This means that the entered PIN is only 1 digit off compared to the expected PIN. A sequence of 1447 or 2338 would yield in a Levenshtein distance of 2, with two digits off the expected sequence.

Table 3.2 shows the relative and absolute distribution of errors categorized in the four possible classifications of the Levenshtein distance from 1 to 4. The majority of unsuccessful authentication attempts have a Levenshtein distance of 1 for all cue-based modalities. When entering a PIN code

Modality \ Distance	1	2	3	4
Touch	62%	38%	0%	0%
Mid-air Gestures	52%	24%	19%	6%
Gaze	79%	19%	0%	2%

Table 3.2: Distribution of incorrect PIN entries for cue-based authentication. Levenshtein distance is a measure to compare entered PIN to expected PIN, for example, expected PIN of "1234" and entered PIN of "1134" (or "1324") results in a distance of 1 (or 2).

Figure 3.7: Entry time of cue-based authentication. It should also be noted that the outliers were removed using excel function to remove outliers. Otherwise, the results were more drastic. *** means significance of $p < 0.001$

62% of the errors when authenticating with touch, over half of the wrong inputs of mid-air gesture with 52% and 79% for gaze were only one digit off.

To see, if there was any significant difference in the Levenshtein distance between the modalities, we ran a statistical analysis. ANOVA showed that there was a significant main effect of the within-subject factor modality for Levenshtein distance, $F(2, 34) = 4.99$, $p = 0.013 < 0.05$. A pairwise comparison using Bonferroni correction showed only for the pair Touch ($M = 0.10$, $SD = 0.38$) and Mid-air Gestures ($M = 0.33$, $SD = 0.80$) with $p = 0.04 < 0.05$. No significant difference was found for Touch and Gaze ($M = 0.23$, $SD = 0.53$) with $p = 0.07$ and Mid-air Gestures and Gaze with $p = 0.64$.

3.3.3 Entry Time

The mean entry time for Touch was 3.7s ($SD = 1.0$), which was faster than the mid-air gesture condition with a mean entry time of 5.5s ($SD = 3.9$). The gaze-based input method performed worst compared to the previous two methods with a mean of 26.3 ($SD = 22.1$). This is a factor of about 7x or 4.7x compared to touch or mid-air-gestures respectively.

Having a look at the sizes of the box plot (see Figure 3.7), the relatively small box plots of touch and mid-air gesture suggest that they perform relatively reliable in terms of entry time, whereas the gaze condition has a box plot, which is comparatively tall. This suggests that the entry time varies greatly from minimum to maximum, thus making it more unreliable compared to the other modalities.

A repeated measure ANOVA with Greenhouse-Geisser correction revealed that there was a statistically significant difference of entry time between modalities, $F(1.02, 17.41) = 83.05$, $p < 0.001$. Bonferroni-adjusted post-hoc analysis revealed a significant difference in Levenshtein distance between all pairs:

- Touch and Mid-air gestures ($p < 0.001$),
- Touch and Gaze ($p < 0.001$),
- Mid-air gestures and Gaze ($p < 0.001$)

This means for cue-based authentication: touch modality is significantly faster than mid-air gestures and gaze, and mid-air gesture is significantly faster than gaze.

3.3.4 NASA TLX

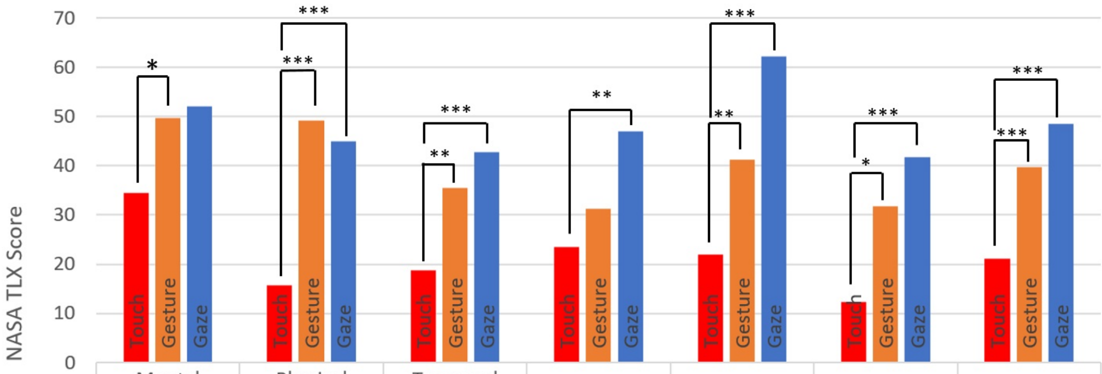
The NASA Task Load Index is an assessment tool in order to assess the perceived workload of a user. The NASA TLX questionnaire consists of six different subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration. We used a non-weighted version for the study. It also provided an overall score, which was calculated by the mean value of all six subscale scores. The questionnaire had to be filled in after finishing a task, this means after finishing each 16 authentication attempts. The summary of the results is depicted in Figure 3.8.

The touch-based authentication method performed best with the lowest subjective workload in all six subscales with an overall score of 21.1 followed by mid-air gesture (score: 39.8) and gaze (score: 48.5). A repeated measure ANOVA also revealed a significant difference in the overall score between modalities, $F(2, 38) = 20.1$, $p < 0.001$. Post-hoc analysis using Bonferroni correction showed that the differences between touch and mid-air gesture ($p < 0.001$), and between touch and gaze ($p < 0.001$) were significant. No significant difference was found between mid-air gesture and gaze ($p = 0.37$). For overview of comparison of the subscales, see Figure 3.8.

3.3.5 Qualitative Feedback

After completing all tasks, we asked participants to fill in a questionnaire about all three modalities. For each one we provided a set of questions with 5-point Likert scale, which they could rate from 1 (strongly disagree) to 5 (strongly agree), and a set of questions with an open format. A short interview was conducted, if there was sufficient time left in the end. The overall results can be seen Figure 3.9

The evaluation of the results revealed that, all things considered, participants rated the touch condition positively. As shown in Figure 3.9 they agreed or strongly agreed that, for instance, touch-based input was easy, natural, pleasant and fast. They also would use it in public and feel comfortable using it in public. The same answers were repeated again in the open questions section. The majority especially emphasized that it was easy and fast to use, because it was "similar to the old fashioned way" (participant P18) and similar to PIN entry or pattern lock on mobile devices (P6). Furthermore, participants pointed out that it was secure (for example stated from P1, P2, P8, P14, P17) due to the combination with the cues (P8). Participant 10 found the touch modality novel and unnatural, but one can get used to the method quickly. On the downside,



	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration	Overall
Touch	34.5	15.75	18.75	23.5	22	12.25	21.13
Gesture	49.75	49.25	35.5	31.25	41.25	31.75	39.79
Gaze	52	45	42.75	47	62.25	41.75	48.46

Figure 3.8: Results of the NASA TLX questionnaire for cue-based authentication. * means significance of p < 0.05, ** means significance p < 0.01 and *** means significance of p < 0.001.

participants, did not like the touchscreen, because it needed a little more pressure for input to avoid errors, otherwise the system would register an incorrect gesture and therefore an incorrect number. For improvement participants suggested a change of colors after each entry (P13) and a feature "to revert the last input" (P16).

Participants were skeptical about authenticating using mid-air gestures. On the one hand, they reported advantages in terms of hygiene, because no touching was needed (P16). In addition several participants pointed out that it was easy to use and also fast. But on the other hand, the majority stated that using mid-air gestures was unnatural. When asked why, participants said that performing the mid-air gestures involved too big and expansive movements (P4, P6, P14, P15). Participant 5 stated that this "might draw attention, because it is not the common way to enter PIN". Some felt uncomfortable using the modality (P3, P5). They also reported that it would make them feel "awkward" and "goofy" (P9), look "ridiculous" (P6), "strange"(P11) or "stupid"(P16) in public. Another point was the interaction distance from user to authentication display, which was too far and would require more space (P10, P11). These disadvantages led the participants to disagree to use mid-air gestures in public (see Figure 3.9. But they saw potential improvements, when smaller and more subtle gestures (P4, P5, P7, P11) were to be implemented and if the interaction distance was lowered (P10).

Participants had a mix opinion on input using gaze. Because of long entry times (see 3.3.3) the majority complained that the authentication process took too long. Participant 5 for example reported that it was "time consuming" and that the "eye tracking needed much time". Participant 10 stated that the long entry time also led him to doubt, if he was watching the correct number. Some criticized the low accuracy of the eye tracker, which led to higher error rates (P6, P8, P13). Several participants also stated that authenticating using gaze was exhausting and straining for the eyes due to long entry times (for example P12). Participant 17, who wore contact lenses, reported that the red colors were difficult to see and that staring too long made their eyes to water, which made the eye tracking even harder. But they also saw benefits using the gaze modality. Several participants mentioned that using gaze make them feel more safe and secure, because it is hard to identify the PIN. They also liked the fact that no hands or some kind of physical activity was need (P2, P10, P11, P14). However, the majority stated that the entry time has to be lower to be accepted, so that they would use it in public and feel comfortable with it.

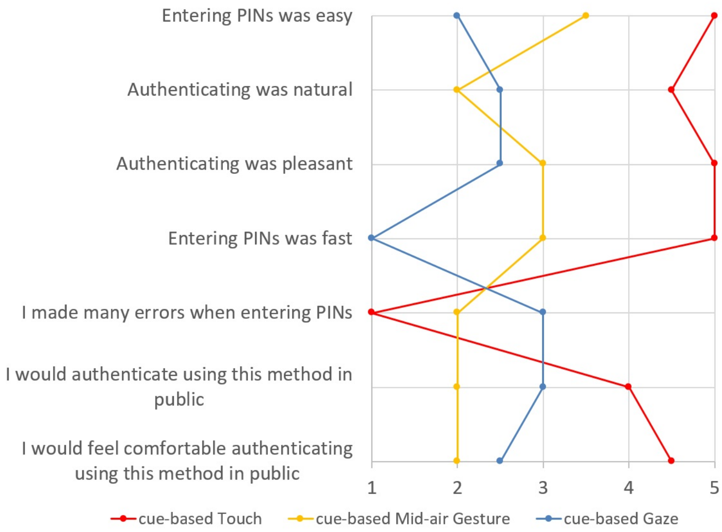


Figure 3.9: Result of the 5-point Likert scale questions for cue-based authentication. Median of 20 participants. 1 means strongly disagree and 5 means strongly agree

As shown in Figure 3.10 participants rated touch-based input as the best. Mid-air gestures and gaze are close to each other.

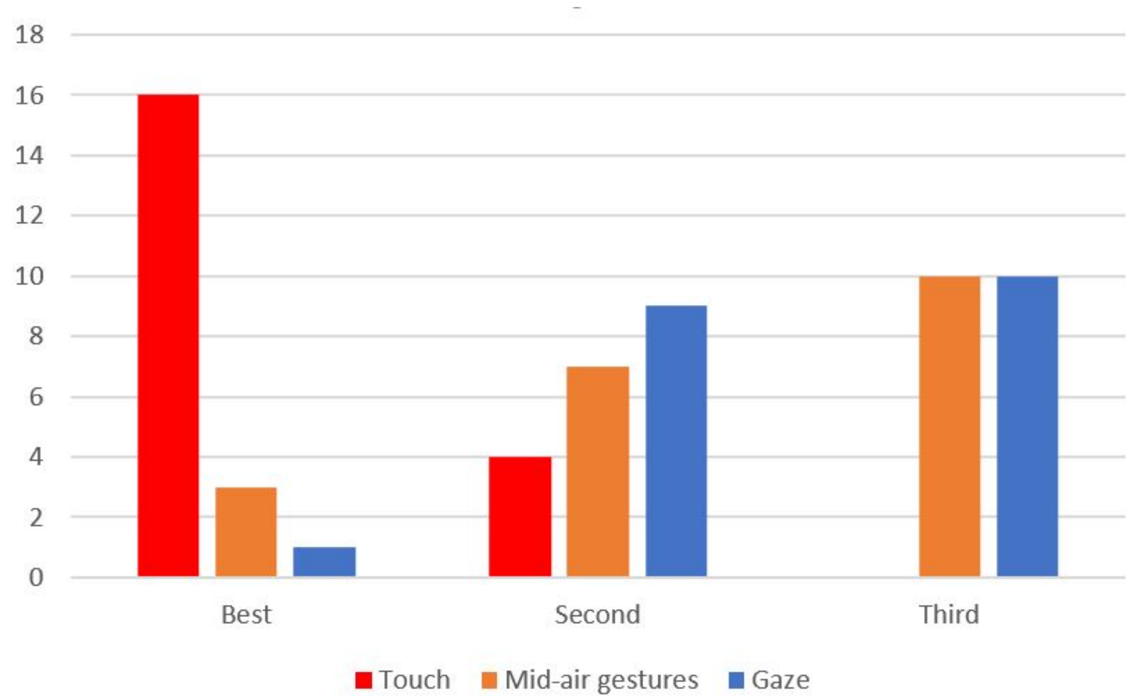
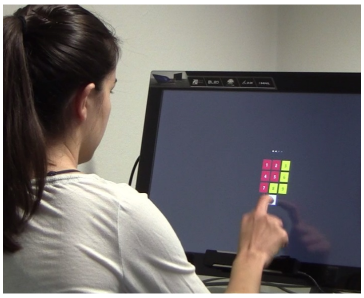
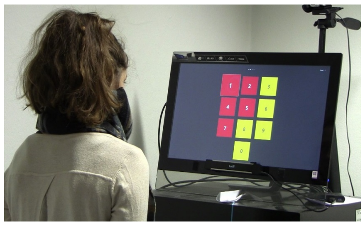


Figure 3.10: Rating for the cue-based authentication methods



(a) Touch input

(b) Mid-air gestures input



(c) Gaze input

Figure 4.1: Traditional PIN entry via different input modalities

4 Classical Authentication

In this section the concept for classical authentication are explained, followed by a description of the study and an evaluation of the results

4.1 Concept

Touch. Authentication on mobile devices or terminals in public spaces like ATMs still use classical PIN entry by touch, despite the fact that this authentication scheme is known to be susceptible to side-channel attacks like shoulder-surfing [14], thermal attacks [1] or smudge [45].

For our study, users would touch the button containing the desired digit for PIN entry on the situated display.

Mid-air Gestures. Previous work [43, 42] have investigated mid-air gestures-based interaction with public displays including gestures like push or grab/grip.

In order to authenticate by mid-air gestures, participants have to wave their hand to move a cursor on the screen. In order to select a number this cursor has to be on top of the desired digit. If a number has been chosen, the user can either (1) push the hand forward or he can (2) make a grab gesture by forming the hand into a fist and then releasing it again in order to enter the number. Version (2) was used for the main study.

Gaze. Previous work by De Luca [13] used gaze-based interaction to make classical PIN entry more secure, for example, against shoulder-surfing. For this thesis, we used the same concepts.

In order to select and enter a digit, the user has to look at the digit and fixate it for a certain amount of time to indicate attention. After literature research [13] and by trying out several dwell time values, we found 800 milliseconds optimal for our setup. After these 800 milliseconds the system would respond by extending briefly the chosen button containing the digit to provide feedback for the user. Another feedback was provided by showing a cursor in form of a bubble, which shows where the user is looking at (see Figure 4.2).

4.2 Implementation

The software implementation we used for classical authentication was written in C# and used the implementation of the cue-based version as a basis. The graphical user interface (GUI) framework for the creation of the buttons, labels, etc. was switched from Windows Forms to Windows Presentation Foundation (WPF). Documentations and code samples of both, the Windows SDK 2.0 for the Kinect One sensor and also the Tobii SDK for the Tobii Rex eye tracker, used WPF, which uses XAML (eXtensible Application Markup Language) to describe the GUI.

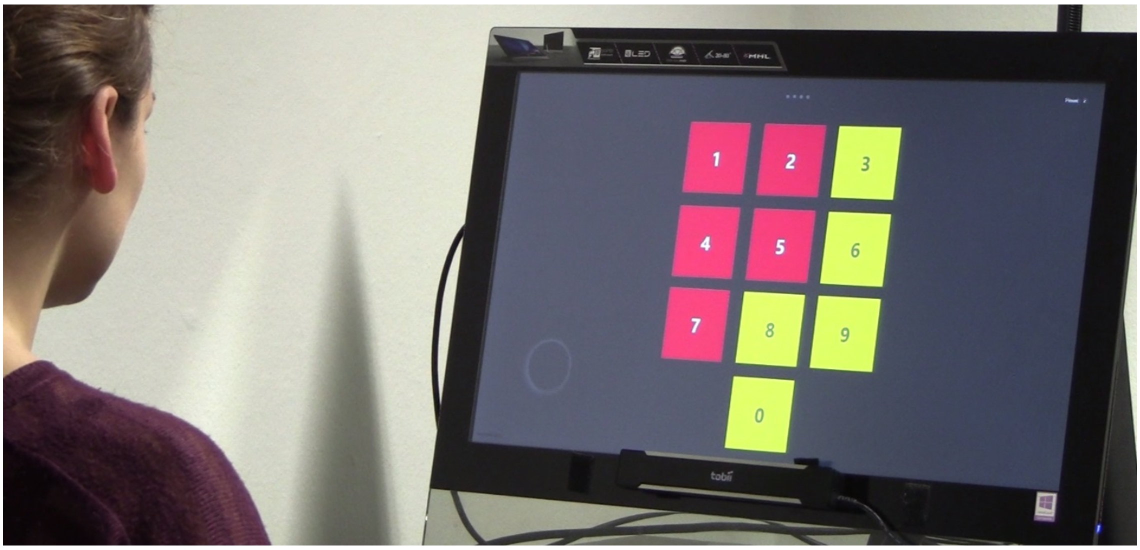


Figure 4.2: Depicts the cursor in form of a bubble.

```xml
<Style x:Key="ButtonStyle" TargetType="Button">
    <Setter Property="MaxWidth" Value="112" />
    <Setter Property="MaxHeight" Value="122" />
    <Setter Property="Margin" Value="10" />
    <Setter Property="Padding" Value="10" />
</Style>

<Style x:Key="EyeXGazeAwareElement" TargetType="FrameworkElement">
    <Setter Property="eyeX:Behavior.GazeAware" Value="True" />
    <Setter Property="eyeX:Behavior.GazeAwareDelay" Value="800" />
</Style>

<Style x:Key="EyeXGazeAwareButton" BasedOn="{StaticResource EyeXGazeAwareElement}" TargetType="Button">
    <Setter Property="Background" Value="LightSteelBlue" />
    <EventSetter Event="eyeX:Behavior.HasGazeChanged" Handler="Button_OnEyeXActivate" />
    <Style.Triggers>
        <Trigger Property="eyeX:Behavior.HasGaze" Value="True">
            <Setter Property="BorderBrush" Value="DarkGray" />
            <Setter Property="RenderTransformOrigin" Value="0.5, 0.5"/>
            <Setter Property="RenderTransform">
                <Setter.Value>
                    <ScaleTransform ScaleX="1.2" ScaleY="1.2"/>
                </Setter.Value>
            </Setter>
        </Trigger>
    </Style.Triggers>
</Style>
```

Figure 4.3: Implementation for gaze using dwell time of 800ms.

The user interfaces for touch, mid-air gestures and gaze can be seen in Figure 4.1.

Touch. The user interface was modified to be able to click the buttons with the digits by touch.

Mid-air Gestures. For the push gesture method, the SDK for the Kinect already provided a built-in feature for this gesture in order to click a button. This was done by adding a "Kinect Region" in the XAML file. Since the SDK did not provide a grab feature for clicking, we had to implement it separately (see source code).

Gaze. The Tobii SDK provided a framework for the layout of the GUI, which used the Tobii eye trackers as input. The dwell time was configurated by setting the properties in the XAML file. In this case we set "GazeAwareDelay" - Property to a value of 800, which equals to a dwell time of 800 milliseconds (see Figure 4.3).

4.3 Study

The second usability study took place between November 10th, 2017 and November 13th 2017 in the same location as in the previous study. The aim was to rerun the experiment with a software implementation adapted to classical authentication methods with similar conditions to gather data for later comparison between both authentication modes.

4.3.1 Experimental Design

The experimental design is described in chapter 3.2.1. A repeated measure was used again for the experiment to ensure comparability. Instead of cue-based methods, users had to perform all three classical versions of the modalities touch, mid-air gestures and gaze in the second study.

4.3.2 Apparatus

The technical setup consisted of the same devices like in the cue-based authentication study (see Figure 3.5). Only the software implementation was adjusted to the classical mode.

4.3.3 Participants

Twenty participants were recruited again through Facebook and mailing list services of the university. Ten of them were female, nine were male and one other. The ages ranged from 18 to 35 years, with a mean of 23.7 and a standard deviation of 4.1. The body height of the participants was between 156 cm and 196 cm (M =174.1 cm, SD = 9.5). Fourteen out of 20 did not use any eyesight corrections, the rest wore eye glasses or used contact lenses for vision correction. Participant 7 had strabismus or a slight eye misalignment of the right eye. When calibrating both eyes and testing the gaze input condition, the system registered a high error rate in recognizing the correct selected digits. After tweaking the system by only calibrating the left eye, the eye tracker selected the intended numbers correctly and reliably. Out of the 20 subjects, 35% had no previous experience with motion sensors, 65% had already experience, for instance, through gaming consoles like the Microsoft Xbox, Nintendo Wii or through other studies. When asked, if they had previous experience with eye trackers, 65% answered with no, and 35% with yes, mainly through other experiments. As compensation for participating in the 45-minute-long study, a 7.50 Euro Amazon Voucher or course credits were granted for each subject.

4.3.4 Procedure

The procedure was similar to the experiment for cue-based authentication (see section 3.2.4 to ensure comparability between cue-based and classical authentication methods. On arrival of the participants, they were requested to sign a consent form. Next, the instructor chose a modality, which was counterbalanced by Latin square (see table 3.1). He then opened the application and explained it to the participant. Like in 3.2.4 the task was again to enter the PIN after a announcing PIN through a speech API using the chosen modality. A training run of three to five authentication attempts was granted, followed by the actual task of 16 PINs for each of the three modality. After completing the 16 PINs, participants had to fill in a NASA TLX questionnaire. In case of mid-air gestures and the gaze condition, participant could also try out the alternative implementations, but these data points were excluded in the main analysis. Finally they had to fill in a last questionnaire about the modalities and if there was sufficient time left, an interview was conducted.

4.3.5 Measures

We collected data for the same variables for cue-based authentication (see measures in 3.2.5).

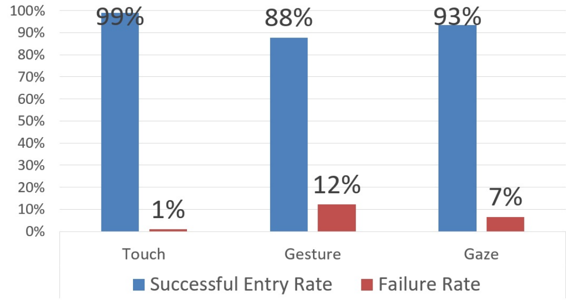


Figure 4.4: Success rate of classical authentication

Modality \ Distance	1	2	3	4
Touch	67%	33%	0%	0%
Mid-air Gestures	62%	23%	10%	5%
Gaze	38%	38%	0%	24%

Table 4.1: Distribution of incorrect PIN entries for classical authentication. Levenshtein distance is a measure to compare entered PIN to expected PIN, for example, expected PIN of "1234" and entered PIN of "1134" (or "1324") results in a distance of 1 (or 2).

4.4 Limitations

During the study we faced another problem. Due to a bug in the software implementation, sometimes an entry of a digit was registered if the user looked at a number right before the announcement was over, which could lead to a failed authentication attempt. Unable to fix the problem in the software code, we asked the participants to look away right before the announcement ended as a workaround. By that we could still collect valid data. Another minor bug was that following two identical numbers, participants had look away or blink to enter the other digit.

4.5 Analysis and Results

In the following the results of the study are presented.

4.5.1 Successful Entry Rate

We collected data of 320 authentication attempts (20 participants with 16 attempts each) for each modality, so in total 960 PIN entries for touch, mid-air gestures and gaze input together. The touch condition only registered three failed attempts, whereas mid-air gesture and gaze had 39 and 21 failed authentication attempts respectively. As shown in Figure 4.4 Touch input was most successful with a successful entry rate of 99%, followed by gaze with a success rate of 93%. Mid-air gestures were not as successful as touch and gaze with 88%. Figure 4.4 shows the results.

4.5.2 Levenshtein Distance

As seen in Table 4.1 the large proportion of errors of the corresponding modality were only one digit off for touch-based input with 67% and for Mid-air gestures with 62%. The percentage of errors when using gaze were distributed in the distances 1 and 2 with 38% and 4 with 24%.

Greenhouse-Geisser correction was used, because Mauchly's Test of Sphericity was violated. A repeated measure ANOVA with Greenhouse-Geisser correction determined that there was a statistically significant difference in Levenshtein distance between modalities, F(1.46, 27.81) =

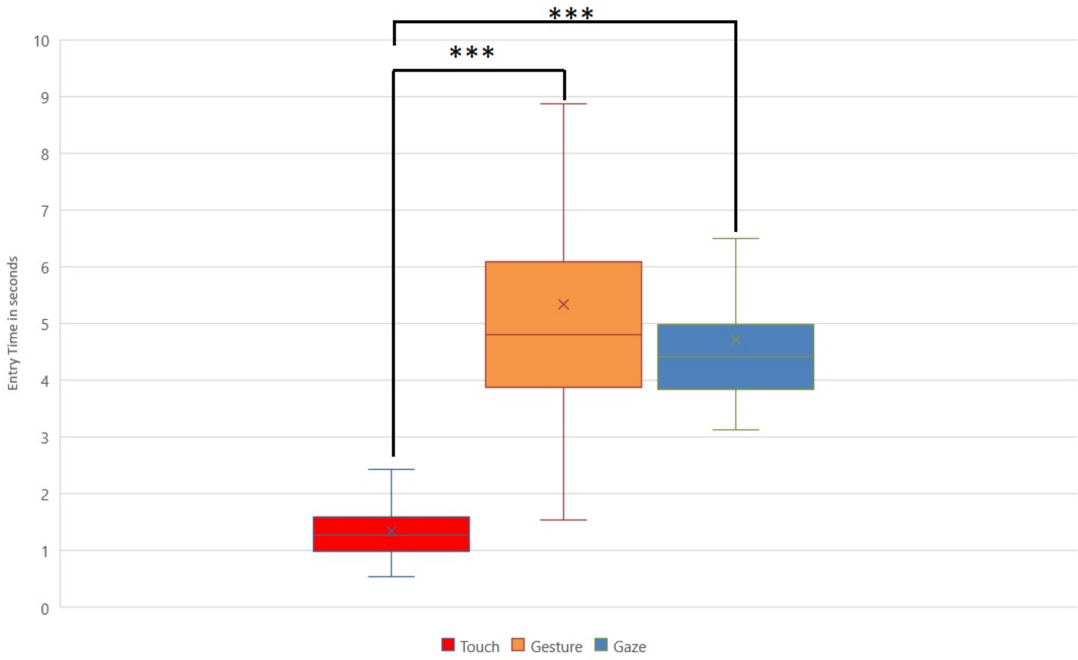


Figure 4.5: Box plot of entry time of classical authentication, *** means significance of $p < 0.001$

7.62, $p = 0.005 < 0.01$. Bonferroni-adjusted post-hoc analysis revealed a significant difference in Levenshtein distance between the pairs Touch (M = 0.01, SD = 0.14) and Mid-air gestures (M = 0.19, SD = 0.60) with $p < 0.01$, Touch and Gaze (M = 0.14, SD = 0.60) with $p < 0.01$. No significant difference was found between Mid-air gestures and Gaze ($p = 1.0$).

4.5.3 Entry Time

The mean entry time for touch-based input was 1.3s (SD = 0.5), which was faster than gesture with 5.3s (SD = 2.5) and gaze with 4.7s (SD = 1.5). Figure 4.5 shows that the data points of touch are comparatively closer to each other than mid-air gesture or gaze.

A repeated measure ANOVA with Greenhouse-Geisser correction revealed that there was a statistically significant difference of entry time between modalities, $F(1.19, 21.43) = 113.04$, $p < 0.001$. Bonferroni-adjusted post-hoc analysis revealed a significant difference in Levenshtein distance for:

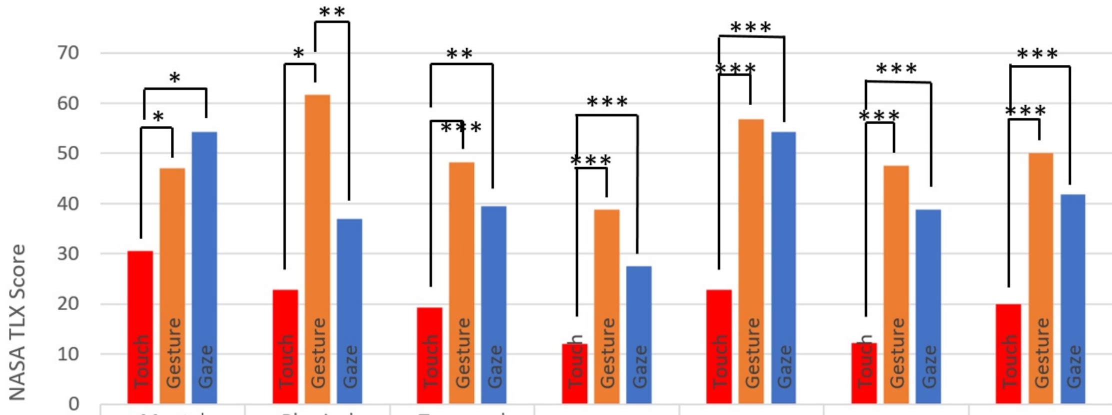
- Touch and Mid-air gestures ($p < 0.001$),
- Touch and Gaze ($p < 0.001$).

No significant difference was found for Mid-air gestures and Gaze ($p = 0.11$)

For classical authentication, touch performed significantly faster than mid-air gesture and faster than gaze. Mid-air gesture and gaze performed similar in terms of entry time.

4.5.4 NASA TLX

Figure 4.6 shows the results of the NASA TLX questionnaire. The touch-based input method had with the lowest subjective workload in all six subscales with an overall score of 19.9 followed by mid-air gesture (score: 50.0) and gaze (score: 41.9). A repeated measure ANOVA also revealed a significant difference in the overall score between modalities, $F(2, 38) = 26.6$, $p < 0.001$. Pairwise



	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration	Overall
Touch	30.5	22.75	19.25	12	22.75	12.25	19.92
Gesture	47	61.75	48.25	38.75	56.75	47.5	50.00
Gaze	54.25	37	39.5	27.5	54.25	38.75	41.88

Figure 4.6: Results of the NASA TLX questionnaire for classical authentication. * means significance of $p < 0.05$, ** means significance $p < 0.01$ and *** means significance of $p < 0.001$.

comparison with Bonferroni correction showed significant differences between touch and mid-air gesture ($p < 0.001$), and between touch and gaze ($p < 0.001$). No significant difference existed between mid-air gesture and gaze in the overall score ($p = 0.45$). For overview comparison of the subscales, see Figure 4.6.

4.5.5 Qualitative Feedback

The results of the feedback can be seen in Figure 4.7.

The majority of the participants agreed or strongly agreed that authenticating using classical touch was easy, natural, pleasant, fast and not error-prone. One of the reason for this rating was the familiarity with touch-based click input. Participant 1, for instance, reported that it works fast and that they are accustomed through daily use of classical touch input. Participant 15 added that this was the previous practice and that he felt quite confident using this modality. It was also reliable to use (P5, P8). But there were concerns about the security. Many stated the problem of authenticating using touch was more likely to be exposed to shoulder-surfing (for example P1, P2, P9) or by leaving fingerprints on the touchscreen (P8).

The score for authentication via mid-air gestures in Figure 4.7 showed a tendency that users were skeptical about using this modality in public. The majority disagreed that it was easy, natural, pleasant or fast. Participants liked the hygienic aspect of not needing to touch anything (P2, P3, P17). But similar to the cue-based version, many complained about the expansive and big movements needed to enter a number. This would lead to several downsides: Performing the authentication in public would make them feel awkward (P2) or stupid (P9). Some said it was very conspicuous and therefore would possibly draw attention from nearby people (P10). They also reported that it was very difficult to perform and to move the cursor on the display (P5, P14). Participant 5 also had the feeling he did not have control and so made many mistakes, making it not working reliable enough (P8). For some, another problem was that entering a number was relatively slow and that everybody could see what users were typing (P7, P8, P19). Also using mid-air gestures were physically demanding and exhausting (P10, P11, P12). As seen on the ratings 4.5 participants would neither use mid-air gestures in public nor feel comfortable in public with the aforementioned problems.

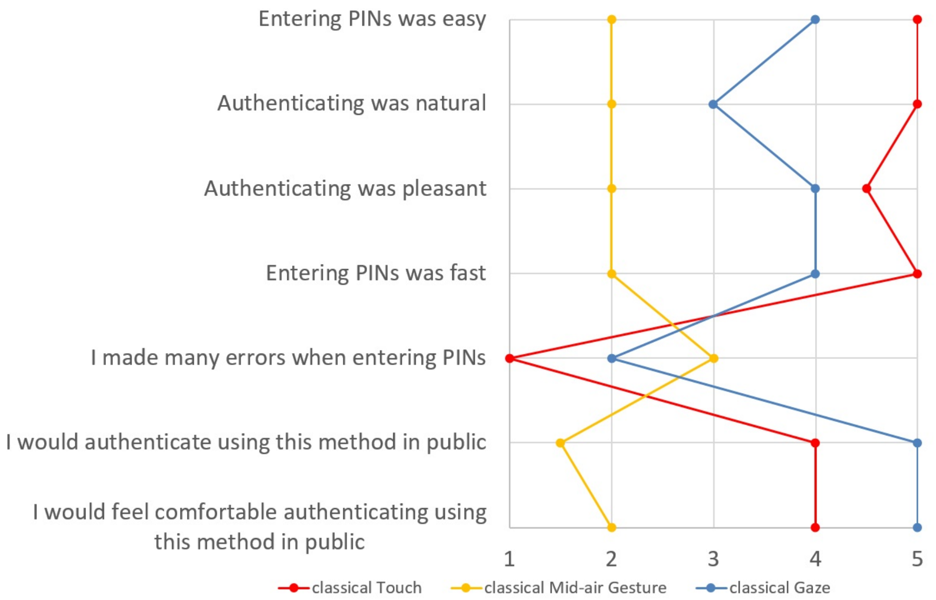


Figure 4.7: Result of the 5-point Likert scale questions for classical authentication. median of 20 participants. 1 means strongly disagree and 5 means strongly agree

There were few downsides using gaze-based input. On the one hand, some participants reported that using the gaze trace feature of the Tobii SDK in form of a bubble as feedback was sometimes irritating, confusing and some did not like it at all (P9, P15, P17, P18). This irritation became worse, when eyes were not calibrated correctly. The bubble would have an additional offset to where one was looking at. Participants then would try to look at the bubble, which caused the bubble to move further away. Participant 17 did not like the bubble during the announcement of the PIN, but also said that it was helpful in entering the PIN. Participant 9 also stated that without gaze trace feedback, it felt more natural, because you were not trying to move the cursor, but the eyes could move between the numbers comfortably. Besides, due to a bug in the software implementation, participant had to look away from the numbers right before the announcement of the PIN ended. This had an impact on the user experience. Several participants stated that they were bothered about not being able to look around the PIN numbers and being required to watch out what they were looking at (P4, P6, P7, P11, P12, P15). This problem is also known as *Midas Touch problem*. On the other hand, there were aspects that users like about gaze-based input. Several participants liked the aspect of additional protection of privacy and protection against shoulder-surfing. It was more difficult to track eye movements (P1), but with the bubble it was easier to detect, what the user was entering (P3). Furthermore there were no fingerprints (P8). Authenticating using gaze was convenient and did not require much physical demand (P15, P17, P19). There were also no conspicuous movements that would draw attention from nearby people (P3). If we look at the ratings of the 5-point Likert scale questions 4.7, we can see that the majority agreed that authenticating using gaze was easy, pleasant and fast and strongly agreed that they would use it in public and would feel comfortable using it in public.

As shown in Figure 3.10 participants rated touch-based input as best, followed by gaze and then mid-air gestures.

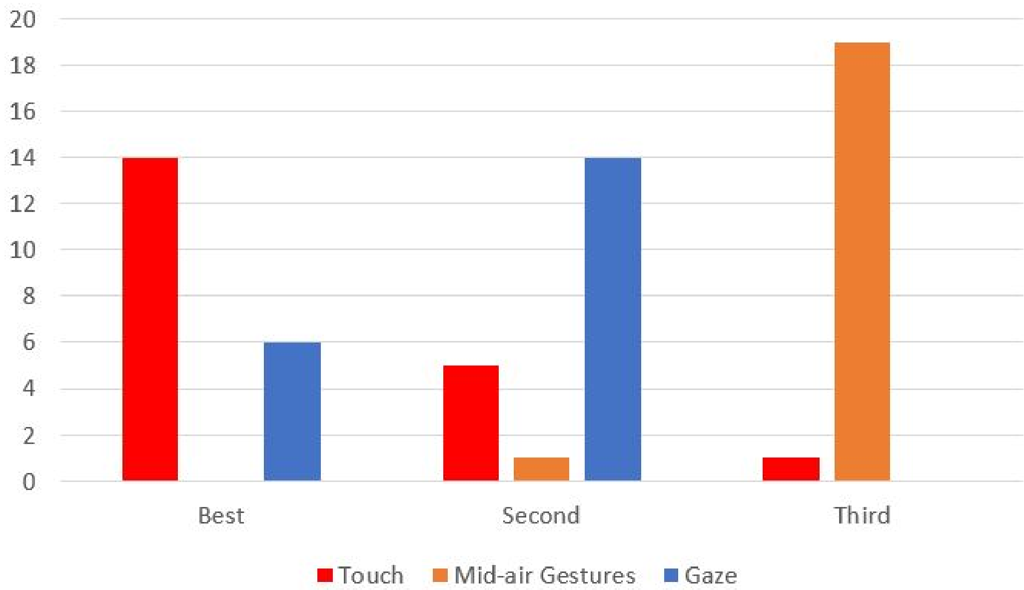


Figure 4.8: Rating for the classical authentication methods.

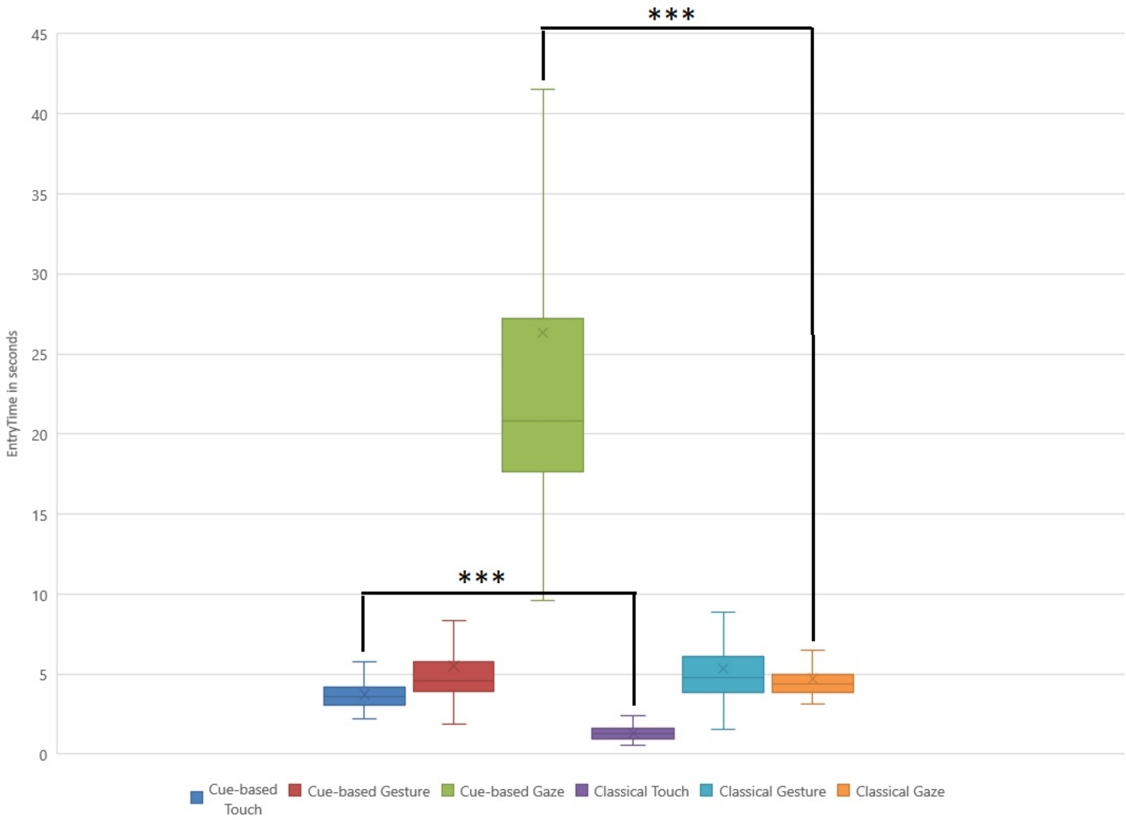


Figure 5.1: Entry time of cue-based and classical authentication

5 Comparison of Cue-based and Classical Authentication

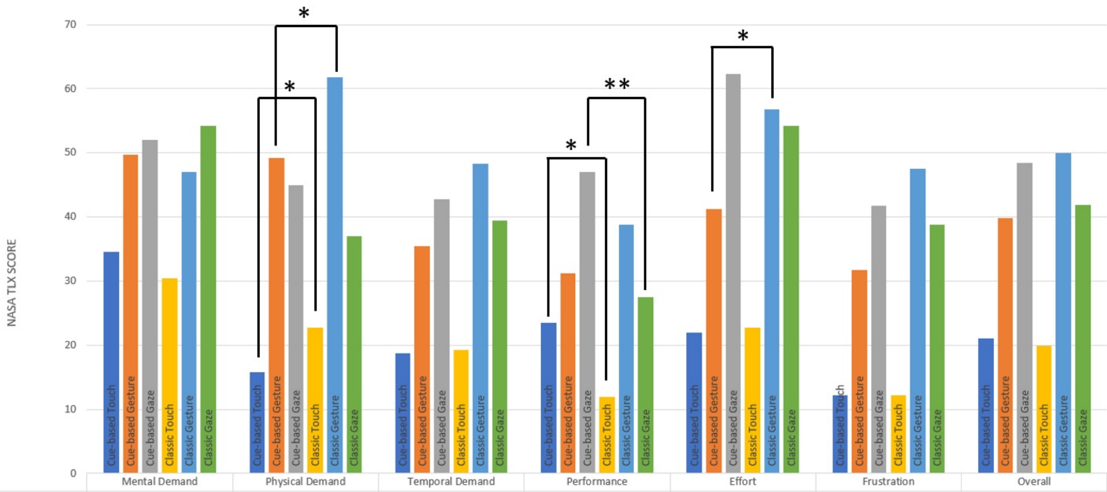
In this section, we compare cue-based authentication methods with classical authentication methods and report the results.

5.1 Levenshtein Distance

Mixed ANOVA revealed that there was no statistically significant interaction between modality and mode in Levenshtein distance, Greenhouse-Geisser $F(1.54, 55.26) = 0.24$, $p = 0.73$. But there was significant main effect for modality in Levensthein distance, Greenhouse-Geisser $F(1.54, 55.26) = 11.75$, $p < 0.001$. There was a significant main effect for mode, meaning that the modes differed significantly, $F(1, 36) = 7.87$, $p = 0.008 < 0.01$. A oneway ANOVA test showed that the Levenshtein distance of cue-based touch ($M = 0.38$) was significantly higher than classical touch ($M = 0.13$) ($p < 0.01$). No significant difference was found for cue-based mid-air gestures ($M = 0.81$) and classical mid-air gestures ($M = 0.60$) ($p = 0.14$), or cue-based gaze ($M = 0.54$) and classical gaze ($M = 0.14$) ($p = 0.08$).

5.2 Entry Time

Mixed ANOVA revealed that there was a statistically significant interaction between modality and mode for entry time, Greenhouse-Geisser $F(1.06, 36.98) = 76.47$, $p < 0.001$. Also there was a significant main effect for mode for entry time, meaning that modes differed significantly, $F(1, 935) = 135.66$, $p < .001$, One way ANOVA showed that cue-based touch ($M = 3.7$s) was significantly slower than classical touch ($M=1.3$s) ($p < 0.001$), also cue-based gaze ($M = 26.3$s)



	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration	Overall
Cue-based Touch	34.5	15.75	18.75	23.5	22	12.25	21.13
Cue-based Gesture	49.75	49.25	35.5	31.25	41.25	31.75	39.79
Cue-based Gaze	52	45	42.75	47	62.25	41.75	48.46
Classic Touch	30.5	22.75	19.25	12	22.75	12.25	19.92
Classic Gesture	47	61.75	48.25	38.75	56.75	47.5	50.00
Classic Gaze	54.25	37	39.5	27.5	54.25	38.75	41.88

Figure 5.2: Results of the NASA TLX questionnaire. * means significance of $p < 0.05$, ** means significance $p < 0.01$ and *** means significance of $p < 0.001$. Only the significant difference of between-subjects factor mode is depicted

was significantly slower than classical gaze (M = 4.7)($p < 0.001$). No significant difference in entry time was found for cue-based mid-air gestures (M = 5.5s) and classical mid-air gestures(M = 5.3s)($p = 0.78$).

5.3 NASA TLX

The results for the subscales are depicted in Figure 5.2.

6 Discussion

The results of the analysis and the feedback of the participants show that cue-based authentication via touch gestures using SwiPIN is a viable alternative to classical PIN-entry in terms of usability and was positively perceived by participants. Cue-based touch was considered as easy, natural, fast and less error-prone. Also the migration from mobile device to situated display does not seem to have a negative impact on the performance. If anything, the mean entry time of 3.73 seconds measured in the study is nearly the same as the reported value of 3.7 seconds by von Zezschwitz et al. [40]. Maybe, this can also be applied to other cue-based authentication schemes.

PIN entry via gaze using Pursuit is not viable yet due to long entry times. The mean entry (M = 26.3) is five times higher than using gaze with dwell time(M = 4.7), and is 20 times higher than classical touch (M = 1.3). A field study of ATM use by De Luca et al. [12] showed that the average time spend on ATMs are 54.9s (about 2 seconds for PIN-entry), which means that authentication schemes with significantly slower authentication times are less likely to be accepted by the user. Since cue-based authentication relies on constant movement of objects, it becomes straining for the eyes resulting in worsened focus on the stimuli. A reason for the poor performance could be the trajectories used in the implementation. As Vidal et al. [39] reported Pursuit uses Pearson's product-moment correlation to calculate a coefficient (from 0 to 1), which determines how similar the trajectory of the eyes compared to the trajectory of the objects. Therefore the correlation coefficient for each of those objects has to be significantly different in order to distinguish the stimuli accurately. Cymek et al. [10] used only linear trajectories limited to horizontal or vertical direction and yielded good performance. A field study by Khamis et al. [22] made the observation that the selection time of Pursuits using linear trajectories were significantly faster than circular trajectories. During our study, we could confirm this observation and also noticed, that there were differences in the selection time between linear, circular and sinusoidal trajectories. We observed that especially the sinusoidal were difficult to detect. The finding is that the choice of trajectory could have a huge impact on the selection time.

On the other hand, gaze-based authentication by fixation with dwell time yielded in much better entry times and less error rates than Pursuits. However it poses other challenges. We made the observation that participants tried to remember their PIN code by looking at the PIN pad, which led sometimes to unwanted PIN entries, if the eyes are staring at the digits longer than the dwell time value. Furthermore, participants could made unintended selections shortly before the announcement of the numbers ended. Since the study already started and we did not have time to fix this bug, we hoped to fix the problem by telling them not to look at the PIN pad shortly before the announcement ended. Some did not look at the pads at all during the announcement. Asking why, they said that they would forget it otherwise to look away. But this made it more difficult for the user to remember the PIN, because the pad also acted for some as a reminder. One way to solve or minimize the Midas Touch problem is to increase the dwell time needed for the selection for the cost of worse performance. Proposed by Penkar et al. [35], another solution is to redesign the buttons. They recommend placing the labels, in our case the digits, outside the buttons. By this, users would be able to watch the digits without worrying to trigger the buttons. Penkar's observation was a significantly improvement of the Midas Touch problem with accidental clicks.

For gaze-based input we used a cursor in form of a bubble as a feedback in order to have similar conditions between mid-air gestures and gaze for classical authentication. The results of the interviews indicate, that users do not need feedback to guide the users gaze. On the contrary, some of the participants felt irritated and distracted by the cursor, and even worsened their experience, when the eye tracker was not calibrated well. We also let the participant try out gaze input without the cursor (about ten attempts per participant). We logged the data but did not use it for the main evaluation. Their response was that it felt more natural without cursor. Some would prefer it without the cursor if the system works reliable. Also it seems, that there was little to no difference between cursor (M = 4.7s) and no cursor (M = 4.8s). In the classical mid-air hand gestures condition, users do not know where to move the hands to select a button. As stated by Walter et al. [42], most of the time mid-air interaction techniques require visual feedback to represent the user. Also, a visual feedback in form of a cursor can help the user to guide their hand to the buttons so that they can click the button by grab or push. With the gaze condition on the other hand, users do not need guidance for their eyes, they already know where to look at and the click action is performed by dwelling.

A disadvantage of gaze by fixation is the requirement of calibration, which takes more time for authentication process and therefore is less convenient for the use in public. However, gaze using smooth pursuit eye movements is calibration-free. But during the study, we also noticed that for gaze using Pursuits the recognition for non-linear trajectories were better with calibration.

Unlike touch-based input, authentication by mid-air gestures or gaze does not need physical contact between user and the display. This is useful, for example, for hygienic reasons or if there is a barrier between user and public display.

Interestingly, cue-based mid-air gestures is more secure than classical mid-air gestures, but there were no significant differences in entry time. But the results of the questionnaire and the interview show, that participants are skeptical using mid-air gestures in public. They argued that by performing expansive movements would attract attention from nearby people, thus feeling uncomfortable. When we asked them if they try to enter their PIN as discreetly as possible at ATMs, 82% of them agreed or strongly agreed. This lack of privacy and embarrassment would result in less acceptance to use mid-air gestures in public. Several participants suggested to use more subtle gestures, for example, just using the finger tips to avoid expansive movements.

7 Future Work

Further investigation. During our study, we could observe differences in selection times for different stimuli for cue-based gaze. An interesting research question would be how much impact the choice of certain trajectories has on gaze-based authentication using Pursuit. For our work, we used linear and non-linear trajectories including circular and sinusoidal movements. With the camera footage of the screen and the PIN entry logs, we can reconstruct and map each digit with its trajectories and its needed selection time. Then we would compare them in how much they differentiate from each other in terms of performance in the aspect like accuracy and selection time. In addition, we could elaborate other possible trajectories and also examine the optimal size of the moving objects with varying movement speed. Another approach would be to reduce the number of cues and therefore the trajectories used at a time. This would increase the accuracy, because the trajectories can be distinguished better.

Comparison of Multimodal Approaches. This work only focused on comparing each modality on its own. But it would be also interesting to compare multimodal authentication schemes by using a combination of touch-, mid-air gestures- and gaze-based input. The combination of touch and mid-air gestures was investigated, for example, by Müller et al. with MirrorTouch [32]. Previous work also explored touch and gaze, for instance gaze-touch [36] or GazeTouchPass [23]. Khamis et al. reported that by using the authentication scheme GazeTouchPass, it was significantly more secure than using only touch or only gaze against shoulder-surfing, but is still usable. Like SwiPIN we could also port GazeTouchPass, which was introduced for mobile devices, to situated displays. Arcade+ by Velloso et al. [38] used multimodal input for interaction for their arcade game machine, including the combination mid-air gestures and gaze. For our purpose of authentication, we could, for example, use gaze to select an object and a push gesture to confirm the selection. Results of Chatterjee et al. [8] indicate, that the combination of gaze and gesture can be faster than single-modal approach for certain interactions.

Improving the modalities. For the cue-based touch, participants suggested to not only change the arrow symbols after each digit entry, but also assign randomly the colors to the buttons containing the digits. In theory, this would make it more robust against shoulder-surfing. Based on the feedback by participants, we could improve, for instance, cue-based mid-air gestures input. The majority found that using large movements would attract too much attention in public. An approach to solve this problem of embarrassment is the use of just the finger tips instead of tracking the whole hand or wrist joints. Lekova et al. [26], for example, proposed a solution for finger tracking and hand gestures recognition using the Kinect v2 sensor. For instance, we could use the index finger tip to perform swipe gestures similar to SwiPIN. Instead of touching the screen, we would then perform the gesture mid-air.

The Midas Touch problem for gaze-based input by fixation can be solved by placing the content of the buttons outside the triggering area, which proved to be very effective according to [35].

Another improvement suggested by participants would be to implement an undo feature. Khamis et al. [25] reported that when interacting with public displays users are willing to correct system errors, if the correction is fast. In theory, an undo button should enhance the user experience since most of the errors are only one or two digits of the expected one.

Study with real-world scenario. Our study was conducted in the lab, in a controlled environment, where participants could ask questions if they had problems. The instructor also corrected them, if they performed the modality wrong. But it would also be interesting to investigate, how the different authentication systems perform in a real-world scenario. An interesting research questions would be: How do participants get along with the modalities, if no or little instructions are available? The instructions, for example, could be integrated into the user interface of the modalities. When authenticating in public like on ATMs, user would already know their PIN code. In our study by announcing the code verbally through a speech API, we noticed that participants would forget the PIN code or sometimes switch two digits, resulting in higher mental demand. Some reported in the interview, that they felt frustrated after several mistakes. A proposal for a solution would be to replace the verbal announcement by showing the PIN code on the screen before an authentication attempt or showing it constantly in a corner of the screen. For the field study participants would try out all input modalities. The preferred one would be examined by letting the user choose the authentication method.

Security Study. For this thesis we only conducted a usability studies. But to compare authentication methods it is also important to evaluate aspects like security in detail by another study, that focuses on the security aspects. During the study, we also recorded the sessions with two cameras, one filmed the participants' execution of the modalities sideways from behind the participants, and the other camera filmed in front of the participants. For gaze-based authentication we could record the eye movements of the user and also the display at the same time, and for mid-air gestures we could record the screen and the execution from behind at the side in directly in front of him. With these video materials, we could construct a threat scenario for the security study. In this study we would simulate a shoulder surfing attack. Participants would have to observe the user by watching the video and try to guess what PIN had been entered. We could then evaluate how secure the authentication schemes are by evaluating, for instance, the success rates of the guesses.

8 Conclusion

In this work, authentication schemes via different input modalities for public displays were evaluated. We have compared cue-based authentication methods via touch, mid-air gestures and gaze, classical authentication methods via aforementioned input modalities and also the two authentication schemes with each other.

The results indicate that cue-based touch input is significantly slower than traditional PIN-entry (3.7s vs 1.3s) and has lower success rates (93% vs 99%), but is still better than the other presented modalities, which use mid-air gestures or gaze. Participants reported that using SwiPIN is fast, easy and secure. We showed that the cue-based authentication scheme SwiPIN, which was originally developed for mobile devices, can also be ported to public situated displays with little to no difference in performance. It is conceivably that other authentication schemes for mobile devices can be migrated for public situated displays with similar performance.

Cue-based mid-air gestures are more secure than classical mid-air gestures. Both have similar performance. But the results of both study also show that there is a low acceptance for authentication by mid-air gestures in public due to the fact that it is embarrassing. Expansive and big hand gesture movements will attract attention from people in the vicinity, thus resulting in less privacy.

Authentication via eye-based input using Pursuit is not viable yet due to high entry times. We noticed that selection times vary by the use of trajectories. First observation show that linear trajectories perform best, followed by circular and then sinusoidal trajectories. An alternative to Pursuit can be fixation with a dwell time. It performed better with lower entry times and lower error rates. However, current implementation is affected by Midas Touch problem. But we are confident that by changing the thus resulting in a better user experience.

During both studies we face several problems and challenges, which can be subject to further research.

Every modality has its advantages and disadvantages. This thesis gives an overview over the aforementioned modalities and can hopefully help in choosing the optimal modalities.

Content of enclosed CD

- *bachelor−thesis/* PDF and LaTeX version of the thesis
- *bibliography/* PDF files of the literature
- *evaluation/* documents created for evaluation
- *presentation/* presentation slides
- *source−code/* source code of cue-based authentication and classical authentication
- *study/* files used and gathered during the study like logfiles, questionnaires etc.

Bibliography

References

[1] Yomna Abdelrahman, Mohamed Khamis, Stefan Schneegass, and Florian Alt. Stay cool! understanding thermal attacks on mobile-based user authentication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3751–3763, New York, NY, USA, 2017. ACM.

[2] Florian Alt, Stefan Schneegaß, Albrecht Schmidt, Jörg Müller, and Nemanja Memarovic. How to evaluate public displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays*, PerDis '12, pages 17:1–17:6, New York, NY, USA, 2012. ACM.

[3] C. Ardito, M. F. Costabile, R. Lanzilotti, A. De Angeli, and G. Desolda. A field study of a multi-touch display at a conference. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pages 580–587, New York, NY, USA, 2012. ACM.

[4] Carmelo Ardito, Paolo Buono, Maria Francesca Costabile, and Giuseppe Desolda. Interaction with large displays: A survey. *ACM Comput. Surv.*, 47(3):46:1–46:38, February 2015.

[5] Ilhan Aslan, Andreas Uhl, Alexander Meschtscherjakov, and Manfred Tscheligi. Design and exploration of mid-air authentication gestures. *ACM Trans. Interact. Intell. Syst.*, 6(3):23:1–23:22, September 2016.

[6] John Brainard, Ari Juels, Ronald L Rivest, Michael Szydlo, and Moti Yung. Fourth-factor authentication: Somebody you know. 2006.

[7] Dominic Canare, Barbara Chaparro, and Jibo He. A comparison of gaze-based and gesture-based input for a point-and-click task. In *International Conference on Universal Access in Human-Computer Interaction*, pages 15–24. Springer, 2015.

[8] Ishan Chatterjee, Robert Xiao, and Chris Harrison. Gaze+gesture: Expressive, precise and targeted free-space interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, pages 131–138, New York, NY, USA, 2015. ACM.

[9] Ashley Colley, Lasse Virtanen, Jani Väyrynen, and Jonna Häkkilä. Physically guiding touch screen interaction with public displays. In *Proceedings of the 4th International Symposium on Pervasive Displays*, PerDis '15, pages 255–256, New York, NY, USA, 2015. ACM.

[10] Dietlind Helene Cymek, Antje Christine Venjakob, Stefan Ruff, Otto Hans-Martin Lutz, Simon Hofmann, and Matthias Roetting. Entering pin codes by smooth pursuit eye movements. *Journal of Eye Movement Research*, 7(4), 2014.

[11] Nigel Davies, Sarah Clinch, and Florian Alt. Pervasive displays: understanding the future of digital signage. *Synthesis Lectures on Mobile and Pervasive Computing*, 8(1):1–128, 2014.

[12] Alexander De Luca, Marc Langheinrich, and Heinrich Hussmann. Towards understanding atm security: A field study of real world atm use. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 16:1–16:10, New York, NY, USA, 2010. ACM.

[13] Alexander De Luca, Roman Weiss, and Heiko Drewes. 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, pages 199–202, New York, NY, USA, 2007. ACM.

[14] Malin Eiband, Mohamed Khamis, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. Understanding shoulder surfing in the wild: Stories from users and observers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 4254–4265, New York, NY, USA, 2017. ACM.

[15] Celeste Groenewald, Craig Anslow, Junayed Islam, Chris Rooney, Peter Passmore, and William Wong. Understanding 3d mid-air hand gestures with interactive surfaces and displays: a systematic literature review. In *Proceedings of the 30th International BCS Human Computer Interaction Conference: Fusion!*, page 43. BCS Learning & Development Ltd., 2016.

[16] Jens Grubert. Nasa-tlx short (non weighted) version in html javascript, Sep 2014. Accessed on February 19, 2018.

[17] Jonna Häkkilä, Olli Koskenranta, Maaret Posti, Leena Ventä-Olkkonen, and Ashley Colley. Clearing the virtual window: Connecting two locations with interactive public displays. In *Proceedings of the 2Nd ACM International Symposium on Pervasive Displays*, PerDis '13, pages 85–90, New York, NY, USA, 2013. ACM.

[18] Uta Hinrichs, Sheelagh Carpendale, Nina Valkanova, Kai Kuikkaniemi, Giulio Jacucci, and Andrew Vande Moere. Interactive public displays. *IEEE Computer Graphics and Applications*, 33(2):25–27, 2013.

[19] INTERNATIONAL ATOMIC ENERGY AGENCY. *IAEA Safeguards Glossary*. Number 3 in International Nuclear Verification Series. INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna, 2002.

[20] Mikkel R. Jakobsen, Yvonne Jansen, Sebastian Boring, and Kasper Hornbæk. Should i stay or should i go? selecting between touch and mid-air gestures for large-display interaction. In Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler, editors, *Human-Computer Interaction – INTERACT 2015*, pages 455–473, Cham, 2015. Springer International Publishing.

[21] Angelique Kessels, Evert van Loenen, and Tatiana Lashina. Evaluating gaze and touch interaction and two feedback techniques on a large display in a shopping environment. In Tom Gross, Jan Gulliksen, Paula Kotzé, Lars Oestreicher, Philippe Palanque, Raquel Oliveira Prates, and Marco Winckler, editors, *Human-Computer Interaction – INTERACT 2009*, pages 595–607, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[22] Mohamed Khamis, Florian Alt, and Andreas Bulling. A field study on spontaneous gaze-based interaction with a public display using pursuits. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, UbiComp/ISWC'15 Adjunct, pages 863–872, New York, NY, USA, 2015. ACM.

[23] Mohamed Khamis, Florian Alt, Mariam Hassib, Emanuel von Zezschwitz, Regina Hasholzner, and Andreas Bulling. Gazetouchpass: Multimodal authentication using gaze and touch on mobile devices. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 2156–2164, New York, NY, USA, 2016. ACM.

[24] Mohamed Khamis, Ludwig Trotter, Ville Mäkelä, Jens Le, Andreas Bulling, and Florian Alt. Comparing touch, mid-air gestures, and gaze for user authentication on situated displays. Technical Report, 2017.

[25] Mohamed Khamis, Ludwig Trotter, Markus Tessmann, Christina Dannhart, Andreas Bulling, and Florian Alt. Eyevote in the wild: Do users bother correcting system errors on public displays? In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, MUM '16, pages 57–62, New York, NY, USA, 2016. ACM.

[26] Anna Lekova, D Ryan, and Reggie Davidrajuh. Fingers and gesture recognition with kinect v2 sensor. *Information Technologies and Control*, 14(3):24–30, 2016.

[27] Eva Lösch, Florian Alt, and Michael Koch. Mirror, mirror on the wall: Attracting passers-by to public touch displays with user representations. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, ISS '17, pages 22–31, New York, NY, USA, 2017. ACM.

[28] Ville Mäkelä, Mohamed Khamis, Lukas Mecke, Jobin James, Markku Turunen, and Florian Alt. Pocket transfers: Interaction techniques for transferring content from situated displays to mobile devices. 2018.

[29] Ville Mäkelä, Sumita Sharma, Jaakko Hakulinen, Tomi Heimonen, and Markku Turunen. Challenges in public display deployments: A taxonomy of external factors. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3426–3475, New York, NY, USA, 2017. ACM.

[30] Emilie Møllenbach, John Paulin Hansen, and Martin Lillholm. Eye movements in gaze interaction. *Journal of Eye Movement Research*, 6(2), 2013.

[31] Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. Requirements and design space for interactive public displays. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, pages 1285–1294, New York, NY, USA, 2010. ACM.

[32] Jörg Müller, Gilles Bailly, Thor Bossuyt, and Niklas Hillgren. Mirrortouch: Combining touch and mid-air gestures for public displays. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services*, MobileHCI '14, pages 319–328, New York, NY, USA, 2014. ACM.

[33] Donald A. Norman and Jakob Nielsen. Gestural interfaces: A step backward in usability. *interactions*, 17(5):46–49, September 2010.

[34] Peter Peltonen, Esko Kurvinen, Antti Salovaara, Giulio Jacucci, Tommi Ilmonen, John Evans, Antti Oulasvirta, and Petri Saarikko. It's mine, don't touch!: Interactions at a large multi-touch display in a city centre. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1285–1294, New York, NY, USA, 2008. ACM.

[35] Abdul Moiz Penkar, Christof Lutteroth, and Gerald Weber. Designing for the eye: Design parameters for dwell in gaze interaction. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, OzCHI '12, pages 479–488, New York, NY, USA, 2012. ACM.

[36] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. Gaze-touch: Combining gaze with multi-touch for interaction on the same surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 509–518, New York, NY, USA, 2014. ACM.

[37] Linda E. Sibert and Robert J. K. Jacob. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pages 281–288, New York, NY, USA, 2000. ACM.

[38] Eduardo Velloso, Carl Oechsner, Katharina Sachmann, Markus Wirth, and Hans Gellersen. Arcade+: A platform for public deployment and evaluation of multi-modal games. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '15, pages 271–275, New York, NY, USA, 2015. ACM.

[39] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 439–448, New York, NY, USA, 2013. ACM.

[40] Emanuel von Zezschwitz, Alexander De Luca, Bruno Brunkow, and Heinrich Hussmann. Swipin: Fast and secure pin-entry on smartphones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1403–1406, New York, NY, USA, 2015. ACM.

[41] Robert Walter, Gilles Bailly, and Jörg Müller. Strikeapose: Revealing mid-air gestures on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 841–850, New York, NY, USA, 2013. ACM.

[42] Robert Walter, Gilles Bailly, Nina Valkanova, and Jörg Müller. Cuenesics: Using mid-air gestures to select items on interactive public displays. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services*, MobileHCI '14, pages 299–308, New York, NY, USA, 2014. ACM.

[43] Soojeong Yoo, Callum Parker, Judy Kay, and Martin Tomitsch. To dwell or not to dwell: An evaluation of mid-air gestures for large information displays. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, OzCHI '15, pages 187–191, New York, NY, USA, 2015. ACM.

[44] Yan Zhai, Guoying Zhao, Toni Alatalo, Janne Heikkilä, Timo Ojala, and Xinyuan Huang. Gesture interaction for wall-sized touchscreen display. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, UbiComp '13 Adjunct, pages 175–178, New York, NY, USA, 2013. ACM.

[45] Yang Zhang, Peng Xia, Junzhou Luo, Zhen Ling, Benyuan Liu, and Xinwen Fu. Fingerprint attack against touch-enabled devices. In *Proceedings of the Second ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*, SPSM '12, pages 57–68, New York, NY, USA, 2012. ACM.

[46] Z. Zhang. Microsoft kinect sensor and its effect. *IEEE MultiMedia*, 19(2):4–10, Feb 2012.