Minimal mobile human computer interaction
el Ali, A.A.

Citation for published version (APA):
el Ali, A. (2013). Minimal mobile human computer interaction
Usability and Security Tradeoff in 3D Gestural Authentication

Using the magnet-based ADI paradigm introduced in Chapter 6, we investigate the usability and security trade-off in 3D gestural authentication (Gestural Authentication Study). We replicated a controlled security study to assess the vulnerability of this authentication method against video-based shoulder surfing attacks. Additionally, in a separate controlled study, we measured the usability and user experience issues associated with performing air gestural signatures for smartphone security access. For this, we measured user experience using experience, subjective workload, and trust questionnaires as well as analyzed performed signature durations. Our security analysis provided further validation of the security of this authentication method, and with our user experience research, we arrived at a set of design recommendations for optimizing the user experience of using this magnetic gestural authentication method. The work presented in this chapter is currently under peer-review as “Investigating the Usability and Security Trade-off in Magnetic Gestural Authentication” in the International Journal of Human Computer Studies (El Ali et al., 2013a).

7.1 Introduction

As presented in Chapter 6, magnet-based ADI is a subset of ADI, which also allows gestural interaction in the whole 3D space around the device by deforming an embedded compass sensor’s magnetic field. Previous work (Sieger et al., 2010) has shown a need for additional layers of security for different settings, including entering a pin code to access ATMs or unlock smartphones. Indeed, in a survey with 465 participants asking about security methods on mobile phones, Ben-Asher et al. (2011) found that only 26.7% of respondents perceived PIN-based entry methods to be a secure method of user authentication.

ADI opens up new forms of user authentication, where users can perform mid-air 3D gestures (e.g., their signature) to gain access to a system (e.g., Patel et al. (2004)). This promises a fast, secure and natural method for user authentication. User authentication is achieved here by allowing a user to gesture in mid-air and verifying whether this signature matches that user’s recorded template signature (Guse et al., 2011). Despite that 3D gesture authentication is not generally perceived (as assessed by a web survey) as providing a high
level of security by users (Sieger et al., 2010), recent work has shown that 3D gestures are in fact quite secure against video-based shoulder surfing attacks (Sahami Shirazi et al., 2012). However, to fully understand whether this kind of authentication method would be culturally and commercially adopted, the trade-off between security and usability issues (Weir et al., 2010; Cranor and Garfinkel, 2005; Yee, 2004; Schultz et al., 2001) behind this form of interaction needs to be further investigated.

In this chapter, we look closely at the security and usability issues associated with magnet-based ADI and its applied use for user authentication purposes. While this work focuses on smartphone security access, this method of authentication is applicable to security access of any device (e.g., laptops, doors or ATMs) if embedded with a magnetometer. To investigate the usability issues associated with this method of authentication, a controlled laboratory experiment was set up where users had to define magnetic 3D signatures. Usability and user experience data was collected using a a mixed-methods approach, including measuring signature performance duration, usability and user experience Likert-scale questionnaires, and interviews. To assess the vulnerability of this authentication method, a separate video-based shoulder surfing attack scenario (cf., Sahami Shirazi et al. (2012)) was set up in a controlled setting where users had to try to forge some of the signatures defined in the usability study.

The rest of the chapter is structured as follows: first we provide our research questions followed by a review of related work and a description of our magnet-based ADI framework. We then introduce the usability study design and methods, followed by the security study design and methods. Afterward, we present our results and discuss them, provide design recommendations and conclude.

7.2 Research Questions

In this chapter, our main research question is:

**RQ 6:** How does 3D gestural interaction affect the usability and security of mobile gestural user authentication?

Specifically, what is the security of this method (as measured by recognition accuracy under a video-based shoulder surfing attack scenario), and what are the usability issues (as measured by overall system acceptance, perceived security, gesture recall, and gesture duration) associated with using this kind of user authentication method?

This chapter builds directly on previous work, where under a video-based shoulder surfing attack scenario, Sahami Shirazi et al. (2012) showed that this authentication method is indeed quite secure against visual-based forgeries. The present study attempts to replicate the findings by Sahami Shirazi et al. (2012), and additionally investigate the usability and user experience issues associated with using this method. Guse et al. (2011) have shown that accelerometer and gyroscope-based gestural authentication for predefined simple gestures (e.g., circle gesture, infinity gesture) under different forgery conditions (Naive Forgery, Semi-naive Forgery, and Visual Forgery) and for different recognition algorithms (Dynamic Time Warping (DTW) (Sakoe and Chiba, 1978), Hidden Markov Models (HMMs) (Rabiner, 1989)) is a secure method for user authentication. These previous
findings led to our first hypothesis, that magnetic gestural authentication is a secure method of user authentication under a video-based shoulder surfing attack scenario.

Previous work has shown that users perceive performing 3D motion gestures for HCI-related tasks (e.g., smart-home control (Khnel et al., 2011)) as a natural mode of interaction (Ruiz et al., 2011; Grandhi et al., 2011), which depending on the gesture performed can also be enjoyable (Rico and Brewster, 2010). Given that magnetic gestural signatures allow a natural mode of interaction, and given the prevalence and acceptance of handwritten paper-based (or wet ink) signatures, our second hypothesis was that magnetic gestural authentication would be perceived to be a usable method for authentication, and provide a positive user experience (UX) amongst participants.

Assessing the usability and user experience issues associated with this method includes measuring responses on whether users are able to define their own unique signature (characteristic-based component), recall their own signature (knowledge-based component), and are willing to carry an external accessory such as a magnet for authentication purposes (token-based component). Additionally, investigating the usability and user experience issues associated with this method also requires investigating how long on average it takes participants to perform and recall a signature (i.e., signature duration), the perceived level of trust by users towards this type of authentication system, as well as how users perceive the difficulty in forging a signature with full video evidence.

Investigating the security, usability and user experience afforded by this authentication method here yields two main research contributions. First, it provides further evidence on the vulnerability of the magnetic gestural authentication method towards video-based shoulder surfing attacks. Second, it provides insight into whether this authentication method (even if shown to be secure) is a usable method for user authentication, and in cases where it falls short, how the usability and UX issues can be addressed.

7.3 Related Work

In addition to the related work listed below, this chapter builds on the related work on using 3D gestures in HCI discussed in Chapter 5 in Section 5.3.1 as well as the related work on Around Device Interaction discussed in Chapter 6 in Section 6.3.1.

7.3.1 Protecting Against Shoulder Surfing Attacks

The degree of a shoulder surfing attack generally depends on the situation in which it occurs (e.g., on the street or at the cashier in a supermarket). Keypads or touch screens in alphanumeric or graphical passwords (Biddle et al., 2012; Suo et al., 2005; Renaud and Angeli, 2004) are particularly vulnerable, since an adversary can easily obtain a direct view of the interaction with the authentication interface. Examples of graphical passwords that use graphics or pictorial representations include PassFaces (Tari et al., 2006), Jiminy (Renaud and Smith, 2001), VIP (De Angeli et al., 2002), Déjà Vu (Dhamija and Perrig, 2000), and Passpoints (Wiedenbeck et al., 2005).

UX here is based on ISO 9241-210 (1994) definition: A person’s perceptions and responses that result from the use or anticipated use of a product, system or service.”

117
Magnet-based gesture authentication intersects with 3 authentication schemes available (Shanmugapriya and Padmavathi, 2009; Farella et al., 2006): 1) knowledge based, where users have to recall their own defined gestural signatures (cf., password recall) 2) token-based, where users have to possess a magnet (cf., an RFID tag) and 3) Characteristic or biometric-based, which can either be physiological (e.g., fingerprints, face or iris recognition) or behavioral (e.g., keystroke dynamics (Shanmugapriya and Padmavathi, 2009), gait, or hand motion). Since users’ signatures are (largely) unique and characteristic of their hand and arm motion (cf., paper-based or 2D touchscreen signatures), magnetic gestural authentication falls under a biometric behavioral scheme.

While physiological biometric methods provide a high level of security given that they are based on an individual’s unique biological markers, they typically involve costly hardware, and even then are not completely risk free (cf., the review by Faundez-Zanuy (2004) on the vulnerability of biometric security systems, or a specific instance applied to attacking fingerprint-based authentication (Uludag and Jain, 2004)). Likewise, in a typical password authentication scheme that involves alphanumeric passwords, dictionary attacks (Morris and Thompson, 1979) can easily succeed, especially since people are used to choosing easy to recall passwords. Indeed, in a case study of 14,000 UNIX passwords, searching from a dictionary of merely 3x10^6 words revealed almost 25% of the passwords (Klein, 1990). Preventing dictionary attacks for many techniques leads to either heavy computational load (Lin et al., 2001) or user requirements that reduce overall system acceptance by users. Other techniques, such as adopting strong password policies (Computer Security Center U.S., 1986), graphical passwords (Biddle et al., 2012), or designing cognitive games (Roth et al., 2004) again illustrate the classic trade-off between usability and security of user authentication methods (Tari et al., 2006).

A suitable alternative for dealing with shoulder surfing attacks include gaze-based authentication schemes (Kumar et al., 2007) and iris recognition (Daugman, 2004), however these methods pose their own challenges. In the case of gaze-based authentication, it is difficult for users to select a secure password. While users’ recall strategies can be augmented by different cognitive mechanisms such as cued-recall (Bulling et al., 2012), this provides further processing costs on the user. Iris recognition, on the other hand, can be vulnerable to attacks that mimic the vitality of the iris or live tissue verification (Prabhakar et al., 2003). Likewise with fingerprint-based authentication schemes, where research has shown that such systems are vulnerable to attacks using artificial ‘gummy’ fingers made of molds where the live finger was pressed (Tari et al., 2006). The foregoing studies highlight the high tension between different user authentication methods that are both secure and usable.

### 7.4 Magnetic Gestural Authentication Framework

The idea behind our magnetic gestural authentication framework is to use the embedded magnetic sensor (or magnetometer) of a smartphone as a means of authenticating users. A piece of magnet when moved close enough to a smartphone can influence the compass sensor. A typical magnetic sensor contains a 3D Hall effect sensor that registers the strength of the magnetic field along different dimensions. A Hall effect sensor produces a voltage (Hall potential VH) proportional to the magnetic flux density (B in Tesla), due in part to this so-called Hall effect. The output from the sensor is provided in the x, y, and z coordinates.
of the smartphone. This output can have different ranges depending on the device (e.g., the iPhone 4® has a value range between 128 µT). Sliding a magnet around the device changes the original magnetic field around the device. The temporal pattern of such an influence is registered by the compass sensor, and can be interpreted as a gestural command using appropriate machine learning algorithms.

A time derivative function is applied to the sensor readings in order to highlight changes in the pattern of magnetic field, and remove effects of earth’s magnetic field (which is almost constant). The sequence of vectors is divided into overlapping windows for gesture recognition. In order to match templates, we adapt a template matching algorithm called multi-dimensional Dynamic Time Warping (DTW) (ten Holt et al., 2007) to analyze different 3D magnetic signatures. DTW is suitable for measuring similarities between two signal sequences that may vary in time or speed, and can operate with a limited number of templates and still achieve very accurate results. Getting useful information from the magnetic sensor here is not only algorithmically simpler than implementing computer vision techniques, but this approach also does not suffer from illumination variation and occlusion problems. In other words, it does not require direct line of sight into the camera, which enables covering the whole 360° space around the device for interaction.

Whenever a user performs a new signature around the device (illustrated in Fig. E.10 on a iPhone 3GS®), the compass sensor registers the temporal patterns of magnetic field along its three axes. Then, DTW is used to compare this multi-dimensional time series signal with pre-recorded templates of the user’s signature for authentication. If the distance of a new input gesture with respect to the prerecorded signature is less than some threshold, the person is considered as a legitimate user and granted access to a smartphone or protected device. In the tested prototype (described in Section 7.5.1), in order to define an authentication gesture or magnetic signature, the user arbitrarily moves an appropriate permanent magnet (e.g., a magnetic token/stylus or a magnet in a finger ring) around the device along 3D trajectories.

7.5 Usability Study

7.5.1 Study Design

To investigate the potential of this magnetic gestural authentication system, we designed a controlled study to collect unique signatures from participants and to test the usability
and user experience of this authentication method. A controlled laboratory study was suitable in this case as it parallels real-world situations of user authentication (e.g., unlocking a mobile device or laptop in one’s own room, or verification at an ATM in an enclosed space). Additionally, it allows drawing rich user insights and concept validation without the unpredictability of field testing. We adopted a mixed-methods approach that was largely qualitative and based on user insights, in order to test how practical this method would be if it were to be adopted and integrated in the lives of users (Dourish et al., 2004). The collected signatures here were used as material for the shoulder surfing attack scenario (described in the Security Study). Given these design decisions of collecting participant signatures and testing the usability of the method, we set up a single factor experimental design where participants were required to make a unique gestural signature, and later recall this signature.

**Apparatus**

For recording signatures, we used our magnetic gestural authentication prototype. To obtain precise magnetometer signal information, the SHAKE SK6 (Hughes, 2006) sensor was used instead of an iPhone 3GS/4®. The SHAKE sensor is able to sense magnetic fields from proximate magnetic material and transmit the data to a PC over a Bluetooth connection. We used an authentication software on the PC to capture the magnetometer signals. The SHAKE sensor however has a small form factor (almost the size of a matchbox), which might influence how participants viewed this interaction method. To ensure that there was no such influence, the SHAKE sensor was embedded in a foam model (shown in Fig. 7.2(a)) that replicated the height (115 mm) and width (59 mm) dimensions of an iPhone 4®, however kept the thickness of the SHAKE sensor box (~15 mm). To perform magnetic gestural signatures, participants were provided with a bar-shaped magnet (~5cm height, ~0.5cm width) with one magnetic pole marked (shown in Fig. 7.2(b)).
7.5. Usability Study

Defining & Recalling Gestural Signatures

For defining a signature, participants held the foam device model in one hand, and the magnet in the other hand with the marked end always facing upward (to ensure consistency across recorded templates). All users used the same magnet to define the signatures. To ensure that the recorded signature was unique, they were allowed to practice performing this signature as long as they wanted. There was no limitation on the shape or length of the signature, nor on whether they used one or two hands to perform the signature, so long as the magnetometer was able to detect the interaction. Once they were done practicing, they could then record this signature using the authentication software. To ensure that the recorded signatures for participants were consistent, they had to record their signature five times, resulting in five signature templates. The average distance between the five templates ($\text{Sign}_d$) was used as the target template for attacks in the security study. To collect video evidence of the performed signatures for later use in the shoulder surfing attack scenario, participants were recorded from four different angles (front, left, right, rear) while they performed their initial signature.

Once participants recorded their signature, they had to fill in questionnaires prompting them about the usability and user experience of this security method. In order to assess whether they could correctly recall their own signature, they were asked to perform the same signature again after filling in the questionnaires. For the recall signature, participants had three chances to perform the correct signature, resulting in three signature templates. They were however not given feedback (discussed in Section 7.8) as to whether or not they performed the recalled signature correctly, as signal comparisons were done offline. To ensure that the recall scenario was natural, participants were not instructed that they would have to recall the signature they performed, as it would have risked excessively deliberate efforts to memorize the exact signature gesture movement. The lag period between the original signature and the recall signature lasted between 7-10 min (depending on how fast a participant filled out the forms).

Measuring Usability & User Experience

Aside from collecting signal data, to measure the usability and user experience (our dependent variables) of interaction using the magnetic gestural authentication system, seven data sources were collected: a) signing duration of original and recall signature templates b) System Usability Scale (SUS) (Brooke, 1996) responses c) NASA-TLX questionnaire (Hart and Wickens, 1990) responses d) System Trust Scale (STS) (Jian et al., 2000) responses e) Likert-scale questions about participants’ attitudes toward using magnetic gestural signatures as an authentication scheme f) video recordings of participants’ signatures from four angles g) post-experiment interviews, to get direct user feedback on this authentication method.

Durations of signatures and recalled signatures were measured as they provide knowledge into how long a given signature takes in comparison to existing security methods (e.g., PIN or graphical passwords), as well as provide validation whether original signatures differ from recalled signatures, and from attempted forgery attacks. We administered the SUS (10-item questionnaire on a 5-point Likert scale) to gain insight into the ease of use, efficiency, and satisfaction of this authentication system. The SUS has been shown to be a
robust and reliable tool for measuring perceived usability of interactive systems, where a score of 70 and above indicates an acceptable score (Bangor et al., 2008). We gave participants the NASA-TLX questionnaire (Hart and Wickens, 1990) to assess their perceived subjective workload quantitatively ([0,20] response range) through the index’s constituent categories: Mental Workload, Physical Workload, Time Pressure, Effort Expended, Performance Level Achieved and Frustration Experienced. Subjective Workload was calculated as follows: (Mental + Physical + Time + Frustration + (20 - Performance Level)) / 5. The STS was given to gain insight into whether participants trusted or distrusted this authentication system. Finally, Likert-scale questions (Cronbach’s $\alpha$=.78) and post-session interviews were given to gain additional insight into how participants perceived the magnetic gestural authentication method.

Participants

20 participants (14 male, 6 female) aged between 20-38 ($M_{age}= 29.7; \ SD_{age}= 5$) were recruited. Our participant sample spanned 13 different nationalities, where most were right-handed (18/20). Half (10/20) had a technical background, and more than half (13/20) were familiar with gaming consoles that use some form of gesture recognition technology (e.g., Nintendo Wii© or Microsoft Kinect©). Nearly half (8/20) use some form of security scheme to secure their mobile device or laptop, where 4-digit PIN passwords were the most common. However, all were familiar with password and PIN-based security schemes.

Setup & Procedure

The study was carried out at the usability lab at Telekom Innovation Laboratories (Berlin, Germany). Each session took between 45-60 min. Participants were tested in pairs, where each was provided with a foam model with embedded SHAKE sensor and a pole labeled magnet (shown in Fig. 7.2). Participants were guided by two experimenters. They were seated at opposite ends of a table. Webcams captured the front and side angles of their gesture interaction space, and a tripod-mounted camera was mounted behind them to capture the rear angle (see illustration in Fig. 7.3). Recorded video streams were all synchronized using EvoCam 4©. They were allowed to define their own interaction space (within the cameras’ angle views) to ensure their comfort during the session. At the start of the session, each participant filled a background information form, signed an informed consent form, and read through instructions for performing the task. Before starting, they were given a quick tutorial and demo on how to hold the magnet and foam-based sensor, and how to record a signature. To record a signature, they had to press and hold the button on the SHAKE sensor (see Fig. 7.2(a)), where the LED turns blue when the button is held. To stop recording, the button had to be released.

After the tutorial, participants would record five signature templates. After recording their signature, they were asked to fill in the SUS, NASA-TLX, STS, and the Likert-scale questionnaire. Afterwards, they were asked to reproduce their original signatures, and record those recalled signatures three times. Once they finished recording their recalled signature templates, they were briefly interviewed about their experiences (\~10 min.) of the experimental session and the magnetic gestural authentication system. Finally, they were thanked for participating, signed a receipt form, and offered a monetary reward for...
Figure 7.3: An illustration of the standard laboratory setup for collecting signature data from participants. Note: in the actual setup, two participants were tested per session.

7.6 Security Study

7.6.1 Study Design

Video-based Shoulder Surfing Attacks

To investigate the security of the magnetic gestural authentication method, we built on previous work (Sahami Shirazi et al., 2012) and designed a follow-up controlled experiment to assess the vulnerability of this method against video-based shoulder surfing attacks. Under this scenario, we assume the worst case scenario where the adversary has full access to HD video evidence of the different angles of performed signatures. Since 2D cameras are widely available to attackers, these were used instead of depth cameras (e.g., Kinect®). To ensure that the adversary has sufficient information on the performed gestural signatures, the video recordings of signatures from the usability study were provided to attackers to try and forge the targeted signatures. These videos captured defined signatures from four different angles: front, left, right, rear. To make this scenario realistic, we put a restriction on the number of security attacks, where an adversary was allowed only a total of three attempts. It should be noted that our study is based on adversarial attacks of average, everyday persons, and not skilled forgers.
Measuring Security

As in ink-based 2D signatures, some signatures are easier to forge than others (e.g., few strokes or clear letters of a person’s name). If our gestural authentication method is to provide security across all signature types, then it has to be secure enough against both easy and difficult signatures, and against a variation in signature styles (e.g., using one or two hands for performing a signature). Therefore, we decided to additionally test the vulnerability of this method when videos for attack showed both easy and difficult signatures, and signatures that used one or two hands to perform. To determine which signatures from the usability study were easy or difficult, two independent coders were recruited and asked to make a checklist amongst the resulting (post data cleaning) 15 signatures (see Section 7.7.1) made and to determine which are easy and difficult. The easy and difficult signatures were varied amongst those with a variety of styles (such as using one or two hands to perform). Their lists were subsequently cross-checked, and matching candidates for easy and difficult signatures were nominated. The experimenter then selected two easy signatures and two difficult signatures for forgers to target in this study.

Figure 7.4: Participant using paper aids for targeting a signature.

The foregoing design decisions led to a within-subject factorial (2 x 1) design, where all participants had to forge gestural signatures, and signature difficulty (2 levels: easy vs. difficult) was a within-subjects factor. Participant assignment was randomized, and order of presented videos was counterbalanced. As in the usability study, participants were given a short training and all relevant hints for forging, such as grasping the foam SHAKE device with the correct position and orientation and how the magnet was held (marker on magnet always up). The four videos (corresponding to different view angles of each signature) recorded in the usability study were shown to each participant, who were then asked to forge the targeted signatures. Participants used the same foam SHAKE device and magnet as the one used in the video. There was no restriction to the study duration, where participants could watch the videos as many times as desired. Additionally, they could speed/slow down the videos, as well as step through each frame individually. Aside from being allowed to practice the signature motion as long as desired, they were also given a notepad and pen to draw the gestured signature motion if they wanted (e.g., Fig. 7.4).

As in the usability study, duration of forged signatures were measured. This was done to make duration comparisons between the original signature and mean forged signature duration. After forging each signature, participants were given a short Likert-scale questionnaire (Cronbach’s $\alpha=.64$) that asked about their experiences in forging that particular
signature. Responses for each signature were recorded on the same questionnaire, taking into account the counterbalancing of the presented videos for forgery. After all forgery attempts were made, participants were given another short Likert-scale questionnaire (Cronbach’s $\alpha = .7$) and an exit semi-structured interview to inquire about the difficulty in forging signatures in general, the clarity of the videos and their video angle preferences, and what their general attitudes are to this gestural authentication scheme.

Participants

20 participants (11 male, 9 female) aged between 20-34 ($M_{age} = 27.1; SD_{age} = 3.6$) were recruited. Participant sample spanned 10 different nationalities, where all were right-handed. Nearly half (19/20) had a technical background, and half (10/20) were familiar with gaming consoles that use some form of gesture recognition technology. More than half (13/20) used some form of security scheme to secure their mobile device or laptop, where again 4-digit PIN passwords were the most common, and all were familiar with password and PIN-based security schemes.

Setup & Procedure

The study was also carried out at the usability lab at Telekom Innovation Laboratories (Berlin, Germany). Each session took around 60 min. Participants were again tested in pairs, where each was provided with a foam model with embedded SHAKE sensor and the same pole labeled magnet used in the usability study. Participants were guided by two experimenters. They were seated at opposite ends of a table. Each participant was seated in front of a PC, where they could inspect the video footage of the gestural signatures to be forged. At start of the session, each participant filled a background information form, signed an informed consent form, and read through instructions for performing the task.

Before each condition (easy or difficult signature), they were given a tutorial and instructed how to hold the magnet and foam-based sensor as was done in the usability study. They had as long as they wanted to practice the signature. Once they were ready, they were allowed to record three signature attacks. After recording the signatures, they were asked to fill in a Likert-scale questionnaire, and were briefly interviewed about their experiences (~10 min.) of the experimental session and their forgery attempts. Finally, they were thanked for participating, signed a receipt form, and offered a monetary reward for participating.

7.7 Results

7.7.1 Security

Data Cleaning

To validate signature templates recorded by participants in the usability study for consistency, a data cleaning procedure using 5-fold cross-validation was used. A cutoff point on similarity between signature templates ($\text{Sign}_{t}, t = \{1,...,5\}$) was set. If the ratio was greater than 1.5, then the signature data for that participant was considered inconsistent, and had
to be rejected. If it was lower than 1.5, then the signature was shown to be consistent. This cleaning process resulted in removal of data of five participants, leaving consistent signature template data for 15 participants. For the security study, one participant did not correctly record the forgery templates, and had to be excluded. This left data from 19 forgers to be considered.

**Recall**

As in 2D ink-based signatures, a person’s signature varies each time it is performed to some degree. In order to define a 3D magnetic signature and check the repeatability, the user is required to enter a signature template five times ($\text{Sign}_t, \ t = \{1, \ldots, 5\}$). Once the user successfully registers his own personalized 3D signature, the system can be used. The average DTW distance of all templates is then calculated and used as the main signature ($\text{Sign}_d$). For each signature ($\text{Sign}_t$), a ratio was calculated by dividing by the main signature ($\text{Ratio}_t = \frac{\text{Sign}_t}{\text{Sign}_d}$).

To show whether a participant was able to recall his own signature, the recall distance score ($\text{Recall}_t, \ t = \{1,2,3\}$) for each signature was compared with the ratio of the main signature ($\text{Ratio}_t$). If the ratio value for a recalled signature was above this value, then it can be shown that the participant was unable to recall his signature. To find the acceptable threshold, we found the minimum and maximum thresholds for accepting a signature as a participant’s own, in addition to the ratio across all participants. The lower the threshold ($\theta$) value, the higher the acceptance rate. For original signatures made, the thresholds across 15 participants for each recall attempt are shown in Fig. 7.5, with threshold values summarized in Table 7.1. At $\bar{\theta} = 2$, the percentage of successful login attempts is 84.4%. However, as will be shown below (Section 7.7.1), this threshold value need not be this low to protect against security attacks, while still allowing eligible logins.

![Signature Login](image)

Figure 7.5: For login attempts across users ($N = 15$), with $\theta = 6.5$, the percentage of successful logins is 100% successful.
7.7. Results

<table>
<thead>
<tr>
<th>Recalled Signature</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>Range</th>
<th>$\bar{\theta}$</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.8</td>
<td>1.5</td>
<td>1.6</td>
<td>.7-3.6</td>
<td>1.6</td>
<td>.2</td>
</tr>
</tbody>
</table>

Table 7.1: Recall attempt thresholds ($\theta$) and mean threshold $\bar{\theta}$ for recalled signatures ($N = 15$).

Forgery of Original Signatures

For forgery attempts, a similar procedure was applied to identify the corresponding ratio of threshold values across trials. In order to forge a login, an attacker has three attempts ($\text{Forge}_t, t = \{1,2,3\}$). For each attack attempt, $\text{Ratio}_t = \text{Forge}_t / \text{Sign}_d$ is calculated. If $\text{Ratio}_t$ is smaller than a given threshold ($\theta$), the forgery is successful. Despite that the first successful attempt is enough to authenticate the forger, we take the average forgery threshold value in our analysis. For attack of easy and difficult signatures, the thresholds across 19 participants for easy and difficult signature attacks are shown in Fig. 7.6. Threshold values are summarized in Table 7.2.

![Shoulder Surfing Attack](image)

Figure 7.6: For easy and difficult signature attacks across forgers ($N = 19$), with $\theta = 2.4$, the percentage of successful attacks is less than 10%.

Based on these results, with $\theta = 2.4$, the percentage of successful attacks on both easy and difficult signatures is less than 10%. To evaluate the accuracy of authenticating eligible users for these signatures, we compared the original main signature made by that participant ($\text{Sign}_d$) with each recall attempt made ($\text{Recall}_t, t = \{1,2,3\}$). For recall attempts across 15 participants (Section 7.7.1), with $\bar{\theta} = 2.4$, the percentage of successful logins is 86.7%. This shows that tuning the threshold value for signature recognition around 2.4 provides a good balance between false acceptance (allowing attackers to login) and true acceptance (permitting eligible logins) in protecting against video-based shoulder-surfing
Table 7.2: Attack and recall attempt thresholds ($\theta$) and mean thresholds ($\bar{\theta}$) for recalled ($N = 4$) and forged signatures ($N = 19$).

<table>
<thead>
<tr>
<th>Condition</th>
<th>$t1$</th>
<th>$t2$</th>
<th>$t3$</th>
<th>Range</th>
<th>$\bar{\theta}$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy Signature_1</td>
<td>5.1</td>
<td>4.7</td>
<td>4.5</td>
<td>2.3-14.3</td>
<td>4.8</td>
<td>.3</td>
</tr>
<tr>
<td>Forgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy Signature_1</td>
<td>3.4</td>
<td>1.2</td>
<td>2.5</td>
<td></td>
<td>2.4</td>
<td>1.1</td>
</tr>
<tr>
<td>Recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy Signature_2</td>
<td>4.4</td>
<td>4.5</td>
<td>4.5</td>
<td>2.2-9</td>
<td>4.5</td>
<td>.06</td>
</tr>
<tr>
<td>Forgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy Signature_2</td>
<td>2</td>
<td>1.7</td>
<td>1.6</td>
<td></td>
<td>1.8</td>
<td>.21</td>
</tr>
<tr>
<td>Recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficult Signature_1</td>
<td>5</td>
<td>4.9</td>
<td>4.8</td>
<td>2.3-11.5</td>
<td>4.9</td>
<td>.1</td>
</tr>
<tr>
<td>Forgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficult Signature_1</td>
<td>1.5</td>
<td>1.4</td>
<td>1.4</td>
<td></td>
<td>1.4</td>
<td>.1</td>
</tr>
<tr>
<td>Recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficult Signature_2</td>
<td>3.6</td>
<td>3.9</td>
<td>3.9</td>
<td>1.4-12.5</td>
<td>3.8</td>
<td>.2</td>
</tr>
<tr>
<td>Forgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficult Signature_2</td>
<td>1.5</td>
<td>.9</td>
<td>1</td>
<td></td>
<td>1.1</td>
<td>.3</td>
</tr>
<tr>
<td>Recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ROC Analysis of Magnetic Signature Data

To gain further insight into the security of the gestural authentication method, we used Equal Error Rate (EER) to measure accuracy. This is the rate at which False Acceptance Rate (FAR) and False Rejection Rate (FRR) are equal. FAR is the probability that a non-authorized person is identified and FRR is the probability that an authorized person is not identified. We use this measure because typically the number of genuine cases in a verification system are much smaller than the number of forgery cases. In this case, we made use of all recalled signatures defined in the usability study, as well as forgeries made in the security study. We have 15 (signers) x 3 (recall samples) = 45 cases for genuine recall signatures, and 19 (forgers) x 3 (attack samples) x 4 (signatures) = 228 attacks. FAR and FRR were calculated as follows:

\[
FAR = \frac{\text{# of verified forgery cases}}{\text{# of forgery cases}}
\]

\[
FRR = \frac{\text{# of rejected genuine cases}}{\text{# of genuine cases}}
\]

To calculate EER, for each threshold value (0-14.3) the corresponding FAR and FRR were derived. The value of the point at which FAR and FRR are equal is the EER. As shown in Table 7.3, the The EER is 13%, at threshold value of 2.4, which shows that the magnetic gestural authentication system provides both security (even in high-risk situations), as well as usable access to users of the system. To plot the Receiver Operating Characteristic (ROC) curve, the True Acceptance Rate (TAR) was calculated (100 - FRR). All the (FAR,TAR) pairs were used to plot the ROC curve (shown in Fig. 7.7).

To further investigate whether the EER rate was an artifact of the 2-handed signatures, which might have tipped the balance between having a strong signature and being able to recall it easily, we removed those signatures from our analysis. In this case, we had 15 (signers) x 3 (recall samples) = 45 cases for genuine recall signatures, and 19 (forgers) x 3 (attack samples) x 2 (signature difficulty) = 114 attacks. The EER rate for these two signatures drops to 4% at threshold value of 2.4 (Table 7.3), with the corresponding ROC
7.7. Results

curve (shown in Fig. 7.8). This highlights the need for users to consider making a strong signature one the one hand, but one that is also easy to recall (discussed in Section 7.9).

<table>
<thead>
<tr>
<th></th>
<th>Threshold</th>
<th>EER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All signatures</td>
<td>2.4</td>
<td>13</td>
</tr>
<tr>
<td>1-handed signatures</td>
<td>2.4</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 7.3: Threshold and EER (%) for detection error rates.

Figure 7.7: ROC curve for magnetic gestural authentication system.

Original & Forged Signature Duration

To further establish the difference between original signatures defined and forged signatures, the mean duration of the attacks on easy and difficult signatures were compared with the mean duration of the corresponding original signature. Means, standard deviations, and one-sample t test statistics of easy and difficult signatures for both original and forged signatures are summarized in Table 7.4. The one-sample t test (where the T value was set as the original mean signature duration for each signature respectively) revealed no significance differences for all but the second easy signature. This shows that attackers were on average able to closely reproduce the duration of the attacked signatures, except for a single signature. However as the forgery results (Section 7.7.1) showed, attackers were largely unable to successfully break into the system.

7.7.2 Usability & User Experience

Gestural Signature Duration

To assess whether the durations of recalled signatures were similar to the original signatures made, the mean duration of signatures were analyzed for the 20 participants in the usability
7. Usability and Security Tradeoff in 3D Gestural Authentication

Figure 7.8: ROC curve for magnetic gestural authentication system based on two (1-handed) signatures only.

<table>
<thead>
<tr>
<th>Signature</th>
<th>Mean Original</th>
<th>SD</th>
<th>Mean Forgery</th>
<th>SD</th>
<th>One-sample T-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy_1</td>
<td>3412</td>
<td>211</td>
<td>3253</td>
<td>557</td>
<td>t(18)= -1.2, p=.23</td>
</tr>
<tr>
<td>Easy_2</td>
<td>2884</td>
<td>277</td>
<td>3782</td>
<td>708</td>
<td>t(18)= 5.5, p=.00</td>
</tr>
<tr>
<td>Difficult_1</td>
<td>4417</td>
<td>497</td>
<td>4593</td>
<td>1574</td>
<td>t(18)= .5, p=.63</td>
</tr>
<tr>
<td>Difficult_2</td>
<td>2475</td>
<td>527</td>
<td>3382</td>
<td>1235</td>
<td>t(18)= 3.2, p=.01</td>
</tr>
</tbody>
</table>

Table 7.4: Mean duration (m/s), standard deviations, and one-sample t test statistics of original signatures based on 5 templates each and mean durations and standard deviations of forged signatures based on 3 attacks across forgers (N=19).

study. Despite that for security analysis 5 participants were removed due to inconsistent signature templates, here we chose to analyze the mean durations of defined and recalled signatures for all participants. This was done to ensure that the duration analysis paralleled real-world situations, where a signature may differ in shape and movement, but still comply with the overall duration across different recorded templates. A paired-samples t-test was run, however revealed no significant differences between mean durations of original and recalled signatures. These findings are summarized in Table 7.5. This is in line with our expectations, where signatures that belong to participants are not easily forgotten, at least with respect to duration of the performed signature. To further assess the user experience of this authentication method, participants in the exit questionnaire for the usability study where asked whether they perceived magnetic gestural authentication to be a fast method for user authentication. Participants were quite positive, stating that this is indeed a fast method for authentication (Md=5.5, IQR=3-6).
7.7. Results

<table>
<thead>
<tr>
<th>Signature</th>
<th>M</th>
<th>SD</th>
<th>Paired-samples T-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>3329</td>
<td>1379</td>
<td>t(39) = 2.1, p = .04</td>
</tr>
<tr>
<td>Recall</td>
<td>3143</td>
<td>1232</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.5: Mean duration (m/s) of original 5 signature templates made by users (N=20) and later recall duration (based on 3 templates) with paired-samples t test statistic.

System Usability Scale Responses

Measured SUS responses across participants (N=20) were calculated according to Brooke (1996), and analyzed in terms of average score frequency distributions. Results are shown in Fig. 7.9. Only few participants (6/20) gave a score greater than 70, which indicates that the tested magnetic gestural authentication prototype is not yet ready for use in the consumer market. This is not surprising, given that the system was still a prototype (involving bulky and light foam models with embedded SHAKE sensors and a complicated toolkit interface for recording gestures on a PC). Nevertheless, it does reveal that if this kind of authentication system is to be perceived as usable, it will have to provide (as discussed in Section 7.9) for smoother interaction without burdening the user with the details of template recording and allow for actual mobile devices to be used for recording magnetic signals instead of foam models.

![System Usability Scale Scores](image)

Figure 7.9: Frequency distribution of mean System Usability Scale responses across participants (N = 20) to the magnetic gestural authentication system.

NASA-TLX Responses

To investigate the overall subjective workload incurred on participants using the magnetic gestural authentication system, the NASA-TLX questionnaire (Hart and Stavenland, 1988) was administered after participants defined their original signature. Mean responses and confidence intervals are shown in Fig. 7.10. From these results, it can be seen that the
mean Subjective Workload is 6.7, which provides evidence that this authentication method does not impose high subjective workload on participants. This provides additional support to our hypothesis that magnetic gestural authentication provides a natural and fast method of user authentication.

![NASA-TLX Scores](image)

Figure 7.10: Mean NASA-TLX workload responses [range 0-20] across participants \((N = 20)\) to the magnetic gestural authentication system. Capped error bars represent 95% confidence intervals.

**System Trust Scale Responses**

To investigate whether participants trusted the magnetic gestural authentication system they interacted with, we administered the System Trust Scale (Jian et al., 2000) after participants defined their signature. Participant responses were split into separate categories for trust and distrust of the system, where responses on both categories followed a normal distribution (Shapiro-Wilk test for Trust \((t(20) = .97; p = .80)\) and Distrust \((t(20) = .97, p = .82)\)). A paired-samples t-test was conducted to assess the differences in mean scores between perceived trust and distrust, however no significant differences were found. Means, standard deviations, and paired-samples t-test statistic are shown in Table 7.6.

Analyzing the responses in terms of number of participants who trusted the system versus those who distrusted the system, we found that half of participants \((10/20)\) trusted the system. As is common with novel user authentication schemes (cf., Ben-Asher et al. (2011), it takes time as well evidence for an authentication system to gain the trust of users. In our case, it seems that only half of our participant sample trusted the system. This makes it difficult to state clearly at this early stage of the system whether participants would truly trust the system in the future, however our security results suggest that if participants were informed about the actual security of the system, an increase in system trust over time seems likely.
### 7.7. Results

<table>
<thead>
<tr>
<th>STS Category</th>
<th>M</th>
<th>SD</th>
<th>Paired-samples T-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust</td>
<td>3.9</td>
<td>1</td>
<td>t(19) = .5, p = .15</td>
</tr>
<tr>
<td>Distrust</td>
<td>3.4</td>
<td>.7</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.6: Means, standard deviations, and paired-samples t-test statistics for Trust and Distrust categories of the System trust Scale responses across participants (N=20) to the magnetic gestural authentication system.

#### 7.7.3 Users’ Subjective Feedback

**Overall Gestural Authentication System Acceptance**

To assess the overall acceptance of the proposed magnetic gestural authentication system, users’ subjective feedback was gathered via Likert-scale questionnaire responses and interview responses. This data was collected for both participants in the usability study (N=20) and the security study (N=20).

**Usability Study Participant Responses**

For the usability study, when asked whether they would use this gestural authentication method on a daily basis (e.g., to unlock their smartphone), participants found this method to be quite suitable for daily use (Md=5, IQR=3-6). When asked whether this method was secure enough for their devices, participants responded positively (Md=4, IQR=3-5.25). However, when asked whether this gestural authentication method was better than PINs or passwords, participants did not think so (Md=2.5, IQR=2-5.25). With respect to whether they found the defined air signatures easy to recall, participants were confident that these signatures are easy to recall (Md=6, IQR=4.75-7). When asked whether making air signatures caused them any fatigue, participants did not find these air signatures tiring (Md=5.5, IQR=3-6.25). Together, these questionnaire responses highlight that while the magnetic gestural authentication system is usable, and is perceived as an adequate scheme to secure devices, it is still perceived as less secure than traditional security methods such as PINs and passwords.

After participants completed the usability study, they were asked for their own subjective feedback concerning the presented gestural authentication system. When asked about their overall impression concerning security access using magnet-based air signatures, very few participants were positive (3/20), nearly half were not sure (9/20), with the remaining participants negative about the security such a system provides (8/20). Amongst those who were positive, one participant brought up the point that one needs to recall only one signature, akin to having one handwritten signature associated with one’s bank account (P7: “It’s secure because normally a person uses only one signature, for example in the bank for your bank account. Yes, I find it secure.”).

Amongst those who were not sure about the system, they raised valid concerns, including whether or not the system would be able to recognize their own signature later if they forgot it (P3: “Sometimes I wasn’t sure what I was writing, so if it is to replace passwords, I wonder whether my device can recognize my own signature.”), whether performing their signature in a public place is safe (P14: “My problem is, if someone is looking at you, and your signature is not complex enough, and you do it in public, then someone can steal it.”).
7. Usability and Security Tradeoff in 3D Gestural Authentication

whereas others stated they need study results to determine whether this method is secure enough (P15: “I'd have to wait and see study results to see how secure it was.”). With respect to the first concern, performing air signatures is not very different than performing a handwritten ink-based signature, where with sufficient practice (and recorded signature templates), the system would be able to recognize their signature. Concerning the second main concern raised, the performed signature gesture would have to be small and perhaps performed just slightly above the palm of one’s hand or the smartphone so as to provide cover from others who might be watching. This point will be elaborated on in the discussion (Section 7.8). For the last concern raised, it is not surprising that people need strong evidence about the security of a given authentication scheme before trusting it (which is where the present work fares into), especially given the already widespread adoption of password and PIN-based authentication methods (which have a low bar of acceptance due to their ease of use).

Participants who were negative about this authentication system shared similar concerns to those who were unsure, including the fear that they may forget their own signatures (P8: “I think I wouldn’t be able to get into my own phone again, because I would forget it.”) but also the belief that PINs and passwords are already secure enough (P5: “I would still prefer to use PIN combinations, because I think it is secure enough already.”). For the latter concern, it appears that this notion that PINs and passwords are secure enough stems from the ubiquity of this authentication scheme, however as mentioned before (see Section 7.3.1), this is not only a misconception (Morris and Thompson, 1979), but also is contrary to the findings by (Ben-Asher et al., 2011) who found that participants generally do not perceive PIN-based entry methods to be a secure method of mobile authentication.

Security Study Participant Responses

In the security study, when asked whether gestural magnetic signatures provided strong security for their devices, participants were again positive about this authentication method (Md=5, IQR=4-6). In the post-session interviews, participants were again asked about their overall impression concerning security access using magnet-based air signatures. By contrast to the usability study, here most participants (14/20) found this method to be secure, a few (4/20) who were unsure, and the remaining (2/20) negative about this authentication method. The high discrepancy between the responses in this study likely stems from the fact that in the security study, participants were asked to forge other people’s signatures, which was apparently a difficult task.

Those who had an overall positive impression found this to be a secure method despite knowing that someone may have full video access to the signature under attack (P1: “I think it provides good security, if you do something really difficult, then it is secure. Even if you record it, it is still secure.”). At the same time, the issue of security can also depend on which device one uses this authentication scheme for, and how many other additional layers of security there are besides air signatures (P3: “For a business setting, like my business phone, it is fine. For ATMs, I’m not sure, but if research says it is quite secure, then okay. Probably a good idea to have it besides the PIN.”). From those few that were negative, the main reason was a general distrust of this kind of system and the nature of the interaction method, where the performed signature may take too long to perform (P8: “I don’t like the system. I think it is very useless. Lots of time to do this.”).

The positive impressions given by participants in this study notwithstanding, partici-
7.7. Results

Participant responses were not as positive when asked directly whether they thought this authentication method is more secure than established methods like PINs and (graphical) passwords. Here, less than half of participants (6/20) found magnetic gestural authentication more secure, a few (4/20) stated clearly that it is less secure, where the rest (10/20) stated either that the security level was the same or were unsure (P5: “I’m generally skeptical to new things, probably just as safe or unsafe as the paper signatures.”). Finally, those participants were asked whether they themselves would use this authentication method in their everyday lives (for both smartphone access and ATMs) if this method was shown to be secure. Here, less than half (7/20) stated they would in general (P4: “If it was safe enough, I would use it. I like the idea, I like the concept.”), a few (4/20) said they would not (P10: “For myself, I don’t think so. At this point, no. Maybe if I tried to do my own elaborated signature, then maybe yeah.”), and the rest (9/20) stated they would use this method however it is contingent on which device they are using this security protocol for (P2: “Not for daily use. But for ATMs, I would use it.”; P9: “For a mobile yes. But not for my bank account.”). The latter statements concerning using this method depending on what device the authentication scheme is used for highlights the usability and security tradeoff, where as stated earlier, participants find entering a 4-digit PIN easier and quicker than performing a signature in the air, but at the cost of breakable security.

Perceived Forgery Difficulty

In the security study, participants were given several Likert-scale questions concerning the difficulty of forging air signatures and using this authentication method. In general, participants found forging any air signature to be difficult (Md=5, IQR=2-3.25) and certainly more difficult to forge than handwritten wet ink-based signatures. With respect to the different signatures tested, participants found it difficult to forge the easy signature (Md=4, IQR=3-4.25) and very difficult to forge the difficult signature (Md=6, IQR=6-7). Participants also agreed that both forging easy signatures (Md=4.5, IQR=4-6) and difficult signatures (Md=6, IQR=3-6) were more difficult to forge than forging handwritten signatures.

With respect to the difficulty of following hand movements of users performing a signature in the videos, participants were neutral concerning the easy signatures (Md=3, IQR=2-4) but found it very difficult to follow hand movements of the difficult signatures being performed (Md=6.5, IQR=4.75-7). While these results provide strong evidence concerning the level of perceived forgery difficulty, the vulnerability of a given signature does depend on how simple a given signature is (P6: “Depends if the movements are complicated, then it is very difficult to forge it.”). To gain insight into which video angles were most helpful for participants in their forgery attempts, the back (posterior) view was mentioned to be most helpful (17x), followed by the frontal (anterior) view (9x), the left (3x) and right view (2x).

Gesturing Using Magnets

To investigate how participants in the usability study perceived the gestural interaction afforded by using magnets, they were asked about their attitudes towards gesturing using a magnet, carrying a magnet, and paying for a personal magnet. Participants found the gestural interaction using magnets to be intuitive (Md=4, IQR=3-5), which lends support that
7. Usability and Security Tradeoff in 3D Gestural Authentication

this form of interaction is easy to learn and use (P2: “For me, it’s very intuitive, I think it is very easy to learn and use.”). Nevertheless, one concern raised was fine motor coordination in certain situations such as when a person is intoxicated (P16: “I found the magnets very easy to use. Only thing I worry about is how well it would work if you’re drunk, that could be an issue. Just your coordination, goes out the window when you’re drunk.”). This is a valid concern, and indeed unlike a PIN code or password where the code can be stored away as a backup plan for such situations, authentication using magnetic signatures requires a replication of a predefined level of human motor performance. Nevertheless, with sufficient practice of one’s signature, this problem may be alleviated. Such cases however do point back to the finding raised earlier (see Section 7.7.3) that there should always be a backup plan in case one forgets his own signature (discussed in Section 7.9).

When the usability study participants were asked about whether they would be willing to carry a magnet with them, responses were in general positive (Md=3.5, IQR=2-6), however with clear exceptions expressed by some participants (e.g., P17: “It was alright [the experiment session], but I wouldn’t carry a magnet with me.”). When participants were asked whether they would be willing to pay for a good magnet to ensure that authentication proceeds smoothly (i.e., without any potential weak signals resulting from a poor choice of magnet), participant responses were neutral (Md=3, IQR=2-6). This is not unexpected, as it does bring to question why participants would pay for an additional accessory when existing security methods such as PINs and passwords do not require them. However, this same question was posed to participants in the security study, and their responses were slightly more positive concerning paying for a magnet (Md=3.5, IQR=2.75-5.25). This difference might reflect the fact that participants faced with the forgery situation came to appreciate that they were provided with the very same magnet that the attacked users in the video were using. In such a case, paying for a personalized magnet would indeed provide an additional layer of security. In other words, just as one pays for an extra pair of house keys that one carries around, likewise can be said of magnets for security access. This issue will be elaborated on below (Section 7.7.3).

Form Factors

Magnet Form Factor: Given that the proposed authentication method requires making use of 3D gestures to perform a signature, we expected the form factor of both the magnet and the SHAKE foam model to have an influence on users’ preferences and attitudes towards acceptance of this mode of interaction. To investigate this, participants were asked about whether they faced any difficulties performing the signature gestures, and whether the SHAKE foam model was a well designed probe into the everyday use of this method.

Participants had no problems at all performing the gestures for their defined signatures (Md=6, IQR=4.75-7), however their level of comfort in performing them was borderline acceptable (Md=3.5, IQR=2-6). A few participants mentioned explicitly that they did not like the bar-shaped magnet, because it was too small and can easily slip from the fingers (P10: “Well, I like the experiment. But the magnet is quite uncomfortable to hold, it was quite slippery. If it were more ergonomic, and had a different shape, then it would be good. Definitely better than entering a password.”). Additionally, while others had a problem with the size and thickness of the magnet (P9: “I was not too comfortable with the magnet and device, because magnet was too thick. I prefer paper based, because it is finer, the
shape is finer.”), still others requested a bigger sized magnet (P11: “Also you need a bigger magnet. If it is more like a pen, would feel more natural.”).

These forgoing concerns raise the importance of ensuring an optimal balance between the usability of a given tool (such as a magnet) with a particular form factor and its ergonomic design to ensure comfort in performing air signatures. Additionally, this brings up the issue whether users should in fact be provided with a personalized magnet to ensure greater security (e.g., P5: “If you have a personalized magnet, then it would be quite secure.”). However, if users are provided with a personalized magnet, then the risk of losing this magnet would require additional effort from the user to regain the exact same method, where an additional protocol to allow this has to be implemented (e.g., by banks or smartphone and laptop manufacturers).

**SHAKE Foam Model Form Factor**

To gain insight into how the participants in the usability study found the form factor of the SHAKE foam model, we explicitly asked participants about this in the exit interview. More than half (13/20) of participants found the designed foam model acceptable as a probe into the actual use of this authentication method (P5: “The foam model was realistic. At one point I imagined it would be a smartphone.”), while the rest (7/20) found it too bulky and/or too light (P12: “Maybe it was a little too big. It was also too light.”). Also, one participant mentioned explicitly that he did not like the button on the SHAKE sensor (P11: “I guess it is okay, but I didn’t like the button, too pointy.”).

While the height and width factor of the SHAKE model was designed with the same dimensions as an iPhone 4®, the thickness (depth) of the model was dictated by the thickness of the SHAKE sensor used. Nevertheless, even those participants who found it too thick or too light conceded that they served as believable surrogates for smartphones (P9: “Shape is too big, it is long and thick. But it was fine.”), which meant that we did have external validity, given the prototype stage this authentication system is currently at.

**Social Acceptability**

The aim of an authentication system is to allow users access to the secured system or device, at any place or time. However, for the proposed gestural authentication system, there is a performative aspect where making gestures in public spaces may bring to question social acceptability issues in performing 3D signatures in the air while out in public. To gain an idea of how the users in the usability study perceived such situations, they were asked whether they would feel awkward performing air signatures in public places (e.g., supermarket, mall, etc.). While participants found this mode of interaction to be very socially acceptable (Md=5.5, IQR=3-6), this may boil down to individual differences (e.g., P16: “I think it can be kind of embarrassing if you stand in front of ATMs and write your signature in the air.”).

Moreover, as was shown in previous work (Rico and Brewster, 2010), the willingness to perform a given gesture in public places depends on both the given context and which gesture is being performed. Also, if a given signature gesture is not recognized, and the user has to repeat it, this may also cause reluctance in performing the gesture in a public setting (El Ali et al., 2012). These social acceptability results, which were obtained through a controlled setting (discussed in Section 7.8), provide an early indication that the social
performative aspect of magnetic gestural authentication would have to be further researched longitudinally if a system based on gestural interaction is to be adopted for cultural and commercial use.

7.8 Discussion

There are four potential limitations to the present study. First, in the current magnetic gestural authentication prototype, participants were not given feedback on their performance when both defining their own signature, recalling their own signature, or upon attempted forgery attacks. This was due to the way the data was analyzed, where the magnetometer signal data was processed offline. This limitation was raised by one participant in the security study (P4: “I mean it is hard to tell about my performance, because I don’t know if I made it. No feedback, no percentage of success.”).

Indeed, in a real system, the difference between a performed signature and the recorded templates this signature should match would be visualized to the user in some representation (e.g., a percentage bar indicating a confidence score of how similar the performed signature was to the original). However, if one considers established methods of authentication such as PINs or passwords, then users of these authentication schemes are likewise not given feedback on their accuracy of an entered PIN or password. An exception to this are paper-based handwritten signatures, where for example in some situations, the employees at a person’s bank may choose to disclose that person’s original signature and ask him to prove that he is able to perform that very same signature again in front of them. Since showing the person’s signature is not an option in the case of the proposed authentication system (due to security issues), a visualization of the difference between the performed and original signature would appear to be the most effective method to handle feedback on user performance.

Second, in the usability study, while we investigated the performance of participants to recall their own defined signature, we have only tested short term memory recall (Tulving, 2002, 1993). As has been shown in the cognitive sciences, there are different kinds of episodic memory, including working memory, short-term memory, and long-term memory (Baddeley, 2003; Ericsson and Kintsch, 1995; Baddeley et al., 1974). To assess whether this kind of authentication system can be used for longer periods of time, one would have to investigate whether participants are able to recall their own signature across days, weeks, and possibly months. While this kind of memory test for recall signatures is beyond the scope of the current study, this issue was raised by participants who expressed doubts whether they would be able to access their own device if they were to forget their own signature. In such situations, the recommendation to include a backup authentication scheme (e.g., a security question or PIN) in case one forgets his own signature should be implemented. This will be discussed further in Section 7.9.

Third, in this study, we did not make a direct security comparison between the proposed gestural authentication system and other security methods such as PINs, passwords, graphical passwords (Biddle et al., 2012), keystroke dynamics (Shanmugapriya and Padmavathi, 2009; Monrose and Rubin, 2000), or touch movement dynamics (Sae-Bae et al., 2012). While we did ask participants in the questionnaires and interviews whether they perceived this method to be better or worse than established methods, the comparisons were
limited to what participants were already familiar with (namely, PINs and passwords), and not more advanced (gestural) security methods currently being developed and tested (e.g., behavioral biometric methods based on keystroke pattern dynamics (Shanmugapriya and Padmavathi, 2009)). Nevertheless, in this study we demonstrated the security level that the proposed system provides, as well as the usability issues associated with performing air signatures using a magnet, which we believe provides a sufficient initial assessment of this novel authentication method.

Fourth, while our goal was to investigate the usability and security tradeoff in magnetic gestural authentication across all devices, the current stage of the prototype (in using the SHAKE sensor) was geared towards smartphone security access. In other words, the design of our study was restricted to form factors associated with mobile (smartphone) devices. Nevertheless, we took care in our questionnaire and interview questions to ensure that participants were able to identify that the tested authentication system is applicable to other devices, including laptops, desktops, or ATMs.

7.9 Design Recommendations

In this chapter, our goal was to answer our research question of whether magnetic gestural authentication is both a usable and secure method of authentication for different contexts (e.g., daily use on mobile devices or for security access at bank institutions). In line with our first hypothesis and previous work, our results showed that the gestural authentication system is indeed secure against visual-based forgery attacks. Additionally, our results showed that this system also allows authentication of eligible users, especially when it makes use of a person’s biometric signature (which may be transferred from wet ink-based signatures). With respect to our second hypothesis, our usability results showed that this method of authentication, in allowing for natural 3D gestures, provides a good user experience for participants (namely, low subjective workload, natural gestural interaction, easy to recall biometric signatures). However, given the early stage of the tested authentication prototype, a number of design considerations arose that can aid improving the usability, acceptance and eventual adoption of this kind of security system for personal and institutional use:

1. **Ensuring Transparency of Security Results**: As is common in the introduction of novel authentication systems (Weir et al., 2010), it will take time for users to fully trust the system (cf., Section 7.7.3 & 7.7.2). In order to gain user trust and acceptance, empirical data demonstrating the security of the system should be readily available.

2. **Best Practice Guidelines on Usable and Secure Signatures**: While participants mentioned that making these air signatures was quick, this does depend on the choice of a given signature (cf., Section 7.7.3 & 7.7.3). Currently, there are no best practices for what constitutes a secure signature – however for this system (insofar as the DTW algorithm is used to match signature templates), a long and complicated signature is an indicator for how secure a signature is. In such cases, this may come at the cost of speed of performing a signature (cf., speed accuracy tradeoff (Wickelgren, 1977)), which may negatively impact the usability of the system. Therefore,
arriving at a minimum signature duration could help optimize the balance between security and usability. Furthermore, given the drop from a 13% to 4% EER in our ROC analysis when we factored two handed signatures out provides further evidence on the tradeoff between a usable (specifically easy to recall) signature and a secure signature. Finally, given participants’ concern about public stealing of their signature, guidelines on making a signature in a clandestine manner should be provided (e.g., signing on the palm of one’s hand so as to provide cover from the prying eyes of nearby strangers).

3. **Improving Usability vs. Workload**: While the usability scores on the SUS were low, using the system did not incur high subjective workload for participants (cf., Section 7.7.2 & 7.7.2). As the proposed system is still in the prototype stage, it is not surprising that the general usability of the system is low. Nevertheless, despite this early stage of the system, our NASA-TLX workload results gave clear indications that performing air signatures does not provide high subjective workload, and when additionally considering users’ subjective responses related to gesturing using magnets (Section 7.7.3), this adds to the body of evidence that this authentication scheme is a natural and easy to use authentication method. Therefore, usability of the system, not subjective workload issues, should be addressed.

4. **Complement Standard Authentication Methods**: As was mentioned by some participants, this kind of authentication method would strongly benefit (at least in the earlier stages of use) to be used alongside standard methods (cf., Sections 7.7.3 & 7.7.3). This is to ensure that a backup plan is available in case one forgets his own signature, in addition to providing an additional layer of security.

5. **Designing for Form Factor**: Form factors of the mobile device and magnet appear to play an important role in adoption of this new security method (cf., Section 7.7.3). While some participants found the shape and size of the magnet and foam model to be acceptable, others did not. This raises the issue of whether personalized magnets, in providing more security, should be provided.

6. **Restricting Context of Use Can Facilitate Adoption**: While gesturing using a magnet was perceived to be intuitive, users may not always want to carry a magnet around (cf., Section 7.7.3). This can be solved by limiting the use cases and contexts of use in which magnets are used (e.g., gestural authentication only for security access at bank institutions), or embedding magnets in devices. Relatedly, standardized magnets should perhaps be readily available for purchase, and personalized magnets (if provided) should additionally be provided on demand by the service provider (e.g., mobile device manufacturer or bank institute).

7. **Longitudinal Analysis Requirement For Social Acceptability**: Our early social acceptability results provide initial clues that this method of authentication may indeed be socially acceptable (cf., Section 7.7.3). However, to conclusively state this would require further longitudinal research in users’ natural settings.
7.10 Conclusions

In this chapter, we have presented a user authentication method based on magnetic gestural interaction. In line with our hypothesis and prior work, our results showed that this authentication method is secure against visual-based shoulder surfing attacks. With respect to usability, participants found the authentication method to be quite natural, easy to recall biometric signatures, and providing overall low subjective workload. However, the current stage of the system raised doubts about the general usability (as measured by the SUS) and trust of the system. From these results, design considerations were derived that should serve as a starting point if this kind of authentication method is to be accepted as a standard authentication method on today’s smartphones, but also for ATMs and desktops/laptops. Taken together, we hope that our findings have provided a solid overview of the security and usability tradeoff in magnetic gestural authentication, and can guide future authentication methods that draw on the natural gestural mode of interaction afforded by (magnet-based) ADI.

The three chapters in Part II of this thesis showed that gestural input techniques, like context-aware solutions, can contribute to making user interactions simpler and provide a positive user experience across different domains; namely, task-independent, playfulness, and user authentication. In the following chapter (Chapter 8), we provide a summary of the research carried out in this thesis, discuss how this ties into minimal mobile HCI, and provide future directions.