Shoulder-Surfing Resistant Authentication for Augmented Reality

Reyhan Düzgün
reyhan.duezguen@kit.edu
Karlsruhe Institute of Technology
Germany

Peter Mayer
peter.mayer@kit.edu
Karlsruhe Institute of Technology
Germany

Melanie Volkamer
melanie.volkamer@kit.edu
Karlsruhe Institute of Technology
Germany

ABSTRACT

Augmented Reality (AR) Head-Mounted Displays (HMD) are increasingly used in industry to digitize processes and enhance user experience by enabling real-time interaction with both physical and virtual objects. In this context, HMD provide access to sensitive data and applications which demand authenticating users before granting access. Furthermore, these devices are often used in shared spaces. Thus, shoulder-surfing attacks need to be addressed. As users can remember pictures more easily than text, we applied the recognition-based graphical password scheme "Things" from previous work on an AR HMD while placing the pictures for each authentication attempt in a random order. We implemented this scheme for the HMD Microsoft HoloLens and conducted a user study evaluating Things's usability. All participants could be successfully authenticated and the System Usability Scale (SUS) score is with 74 categorized as above average. We discuss as future work how to improve the SUS scores, e.g., by using different grid designs and input methods.

CCS CONCEPTS

• **Security and privacy → Usability in security and privacy**.

KEYWORDS

Augmented Reality, Head-Mounted Displays, authentication, graphical passwords, usability evaluation

ACM Reference Format:
Reyhan Düzgün, Peter Mayer, and Melanie Volkamer. 2022. Shoulder-Surfing Resistant Authentication for Augmented Reality. In *Nordic Human-Computer Interaction Conference (NordiCHI '22), October 8–12, 2022, Aarhus, Denmark.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3546155.3546663

1 INTRODUCTION

Augmented Reality (AR) provides breakthrough innovations and empowers the industry in many ways. AR Head-Mounted Displays (HMD) digitize many activities by seamlessly integrating virtual elements into the real world and allowing real-time interaction between physical and virtual objects. Over the past years, HMD have been used in various areas like civil engineering, industrial engineering, architecture, medical aids and education [29]. In most of these contexts, these devices store and provide access to sensitive data. Thus, it needs to be ensured that only authorized people have access to the HMD.

In most of these contexts, HMD are used in shared or even public places: For example in the industry, smart factories use AR HMD in areas like production, maintenance or logistics to guide the worker by displaying virtual information whenever it is needed [25, 30]. In medicine, doctors use AR to provide medical data during surgeries and consultation with patients [28]. Thus, the authentication in place at HMD needs to be shoulder-surfing resistant. Note, HMD limit users' peripheral awareness by reducing the field of view and overlaying the real world with virtual elements. This makes it more difficult to notice potential observers when entering the secret (password or PIN).

There are various types of HMD. Our focus is on the Microsoft HoloLens which is one of the most popular AR HMD in industry. To authenticate to the HoloLens, one needs to either enter a (at least 6 digit) PIN or a password on a PIN pad or keyboard respectively. While using voice as input channel is obvious not shoulder-surfing resistant, Kreider showed in [17] that using gestures is also not shoulder-surfing resistant as one can deduce the secret from observing the gesture input (even when the observer does not see the exact position of the virtual PIN pad or the keyboard).

The goal of our research is to propose and evaluate the usability of a shoulder-surfing resistant authentication scheme for the HoloLens. Shoulder-surfing resistant means that someone observing and recording the user while entering his/her secret cannot deduce the actual secret – even if the user is observed during multiple authentication attempts. We discuss the related work on authentication schemes for AR HMD as well as on shoulder-surfing resistant authentication schemes. Based on this discussion, we consider the recognition-based graphical authentication scheme *Things* – proposed by [23] for the non-AR context – as most appropriate for the HoloLens. In the desktop setting, previous work states a higher effectiveness and memorability of graphical passwords compared to alphanumeric passwords [4, 24] . Furthermore, the *Things* scheme was shown to have the highest effectiveness compared to other recognition-based and recall-based graphical passwords [22]. In this scheme, the secret consists of images. During authentication, grids with shuffled images are displayed and for each grid the user has to select the one image which is part of his/her secret. Due to the private display of AR HMD the shuffling of images provides already full protection against shoulder-surfing attacks. Based on our literature review in Sec. 2.2, we could not find any work so far that investigated the usability of recognition-based graphical passwords in the AR context.

NordiCHI '22, October 8–12, 2022, Aarhus, Denmark
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9699-8/22/10.
https://doi.org/10.1145/3546155.3546663

We implemented the *Things* scheme for the Microsoft HoloLens and evaluated its usability in a lab study. The usability of the scheme including both enrollment and authentication process was measured by the effectiveness (all participants could successfully authenticate), efficiency (authentication took on average 32.2 seconds), System Usability Scale (SUS) (score of 74, i.e. above average), perceived usability (e.g., ease of use was rated on average with 4 out of 5) and perceived security (e.g., perceived protection against shoulder-surfing was rated in average with about 4 out of 5). We discuss our results with respect to existing literature on comparable schemes and discover areas of future research.

The paper is structured in the following sections: Sec. 2 details the literature on AR authentication as well as recognition-based graphical authentication schemes. Sec. 3 describes the selection process of the proposed scheme, details the design decisions of *Things* and outlines the enrollment as well as authentication process of the scheme. Thereafter, in Sec. 4 we introduce our research questions and give details on the user study. Sec. 5 presents the results of the user study. To discuss the results, Sec. 6 compares the study results with existing work and indicates the limitations of our work as well as areas for future research. Finally, we conclude our work in Sec. 7.

2 RELATED WORK

Various knowledge-based authentication schemes for AR HMD have been published in the literature: while some are shoulder-surfing resistant and others are not. Some have been evaluated with respect to their usability others not. Furthermore, shoulder-surfing resistant schemes have been proposed and evaluated with respect to their usability in the non-AR-HMD context.

2.1 Knowledge-Based Authentication Schemes for AR HMD Not Being Shoulder-Surfing Resistant

Schemes with graphical passwords – including recall-based as well as recognition-based passwords – have been proposed. The recall-based scheme proposed by Friström et al. [10] is based on users entering a free form pattern by gaze gestures. Hadjidemetriou et al. [13] propose for the HoloLens a cued recall-based scheme which requests users to enter a pattern on specific positions of an image through hand gestures. Other schemes use haptic patterns as the password by tapping a pattern on the touch pad of the Google Glass or other HMD equipped with a touch pad. In [16], the tapping gestures are combinations of flat finger and finger tip taps, and in [34] combinations of various swipes and taps. In [15], Hutchins et al. propose to enter a beat of a song or jingle. Funk et al. [11] propose for the HoloLens to use spatial passwords by choosing coordinates from the real environment through looking at real targets via head-gaze. While all these authors conducted user studies with the proposed authentication schemes, these schemes are not shoulder-surfing resistant.

Two-factor authentication (2FA) schemes have also been proposed: Those which combine the knowledge of the user with a second factor, e.g. biometrics [9, 35] or tokens [1]. Thereby, only the combination with a second factor makes the schemes resistant to shoulder-surfing while the knowledge alone does not protect the user from shoulder-surfing.

2.2 Shoulder-Surfing Resistant Knowledge-Based Authentication Schemes for AR HMD

There are also proposals for shoulder-surfing resistant schemes: e.g., Funk et al. [11] present for the HoloLens a recognition-based scheme which requires users to select virtual 3D objects by head-gaze. Thereby, various objects are randomly scattered in the view field of the user and once the user has selected the first object, the positions of the objects are changed before selecting the next one. Thus, the objects are at random positions for each input. Similarly, Bailey et al. [2] propose for the Google Glass to use 2D images which are selected from a grid of decoy pictures. Selection of pictures can be done via speaking out randomly assigned labels or using head movement in combination of a touch pad integrated in the HMD. Duezguen et al. [7] propose another shoulder-surfing resistant scheme: The secret of the user consists of semantic connections of concepts which is entered in a challenge-response fashion via innate human-based computation without disclosing the secret itself. In each authentication session, a different set of challenges are answered with yes and no responses that observers only watch a random series of yes and nos. All three papers *did not evaluate the usability* of their proposals.

A series of authentication schemes based on entering a PIN with a randomized PIN pad shown on the private display of the AR HMD have been proposed for various AR HMD as well as for different input channels. Some only proposed the scheme without conducting user studies: The authors of [12] proposed tapping on any physical surface that faces the virtual PIN pad and capturing the thermal residue. In [20], the authors proposed to speak out the randomly assigned character associated with the PIN digit. Others conducted user studies to investigate the usability of their proposed schemes: In [19, 34], the secret digits are selected by tapping on a touch pad of the Google Glass. The authors of [36] proposed tapping in the air that faces the virtual PIN pad and using a gesture tracker. Furthermore, the authors of [19] proposed for the Google Glass to navigate via head movements. Several papers proposed for the Google Glass to speak out the randomly assigned character or number associated with the PIN digit [2, 19, 34]. Seo et al. [31] is displaying random numbers in the length of the PIN on the private display. The user calculates the difference between the random number and PIN and speaks it out or scrolls the PIN digit on the touch pad of the Google Glass.

Shuffling the entire keyboard to enable users to enter a textual password using the voice channel was proposed by [2] for the Google Glass and by [20] for an unspecified HMD. No user studies were conducted for these schemes.

2.3 Recognition-Based Graphical Passwords in Other Contexts Than AR HMD

Previous work investigated recognition-based graphical passwords in other contexts than AR HMD. Brostoff and Sasse [4] compared textual passwords with Passfaces which is a recognition-based scheme using a set of face images as the password. The effectiveness with Passfaces was higher than with textual passwords, but the efficiency was lower. Moncur and Leplatre [24] compared the memorability of PIN and graphical recognition-based passwords and figured out that the graphical password is substantially more effective than PIN. Additionally, using mnemonics further improved the memorability of the graphical password.

Dunphy et al. [8] investigate the usability of recognition-based schemes with photographs on mobile phones. They measured a similar effectiveness compared to other graphical schemes, but mention the limitation of long login duration times which was unacceptable to about a third of the participants. Note, the duration decreased significantly over two weeks. Thus, it would be interesting to study the duration over time. As AR HMD are currently considered to be used for specific tasks only, we assume that the user needs to authenticate only a few times per day. Thus, we focus on providing a high effectiveness while we hope that using the scheme over time decreases the login duration.

Hlywa et al. [14] studied the usability of different types of images in recognition-based graphical passwords, comparing object images with face images. In a within subject study, participants tested the schemes on a web service at home. The login duration time with object images was much shorter than with face images. The effectiveness of the two different types of images was not significantly different.

Mayer et al. [22] compared different recognition-based and recall-based graphical passwords to each other. They found out that the recognition-based scheme with object images grouped according their semantics showed the highest effectiveness, i.e, success rate during authentication.

All these schemes randomize the position of the images in each authentication session. But as these schemes are proposed for devices like PC or smartphone which do not own a private display, the randomization of the grids only provide limited protection against shoulder-surfing attacks. As the display of AR HMD is only visible to the user, these schemes can provide full shoulder-surfing protection when applied on an AR HMD.

3 PROPOSED AUTHENTICATION SCHEME

This work aims to propose a knowledge-based authentication scheme for the Microsoft HoloLens 2 which is resistant to shoulder-surfing attacks. Currently, the HoloLens provides a PIN based scheme with a length of 6 digits. Thus, the proposed scheme needs to have at least the password space of a 6-digit PIN. Additionally, from the three usability criteria, effectiveness is the most important one as resetting forgotten passwords is more complicated on the HMD, i.e., the user needs to take off the HMD and switch to another device.

3.1 Evaluation of Relevant Schemes from the Literature

We summarized the proposals from the literature for shoulder-surfing resistant schemes in Sec. 2. In this section, we discuss those with respect to their applicability on the HoloLens in a context in which a password space of at least 10^6 (because of the 6-digit PIN) is required and which is shoulder-surfing resistant.

The scheme based on semantic connections proposed by Duezguen et al. [7] would require to answer 25 challenges while allowing one error to achieve 10^6. While the usability of this scheme is not investigated, we expect here a much longer authentication time compared to other proposed schemes. For example, a recognition-based graphical scheme which requires to select, e.g., one image out of 16 for five times for the same security level. Furthermore, a knowledge-base describing the semantic relationships between the concepts need to be developed and evaluated before the scheme can be applied. Thus, we decide to not further consider this scheme for the HoloLens.

Several schemes using PIN as secret while the PIN pad shows the numbers in a random order have been proposed and evaluated; e.g. in [2, 12, 19, 31, 34, 36]. Similarly, the shuffling of the characters of a keyboard was proposed to enter the secret (most likely a password) even when being observed [2, 20]. There are also proposals to use recognition-based graphical-password schemes, e.g. [2, 11], i.e. objects or pictures are the secret while again their position changes either each time users enter their secret or even before entering the next object/picture. Due to the private display of HMD, for all these schemes it holds that the positions are only visible to the user.

Past research shows, that graphical authentication schemes outperform those with numbers/characters with respect to effectiveness – see e.g. [24]. As for our context effectiveness is the most important usability criteria, we decided to not consider those schemes based on PIN pads / keyboards. Note, in some studies graphical passwords were less efficient than PIN [25, 30]. However as effectiveness is considered more important than the other usability criteria and as the industry context does usually not require entering the passwords several times a day, we acknowledge their findings, but keep the decision to consider a graphical-authentication scheme for the HoloLens.

The recognition-based graphical schemes – proposed e.g. in [2, 11, 14, 23] – are randomizing the position of various types of images and 3D objects respectively before users can select their secrets from the set of displayed options. According to Hlywa et al. [14] object images have a higher efficiency than face images. Furthermore, the *Things* scheme – proposed by [23] – which uses object images and groups the images according their semantics was identified as the scheme with the highest effectiveness among other graphical schemes. Thus, we decided to go with objects rather than faces. The authors in [11] propose to randomly distribute the objects in the room while *Things* places images of objects in a grid. We decided to go with the *Things* approach, i.e. displaying images in a grid, in order to reduce the mental effort for searching the secret image.

3.2 *Things* Related Design Decisions

3.2.1 Image Collection. The images chosen for the scheme are images of objects and each image of the password belongs to a semantic group, e.g., fruits, flowers or animals. During the authentication session, each grid shows random images of a single semantic group and every grid is assigned to a different semantic group. The data base of the object images was composed according the findings of Weinshall and Kirkpatrick [32] as well as Hlywa et al. [14]. Weinshall and Kirkpatrick report that images with a clear central subject or theme are easier to recognize. Hlywa et al. advise to choose images with white backgrounds and bright colored objects. The images in the grid should also vary in color, shape and semantic. Thus, we focus on these criteria when selecting the images for each semantic group. For example, when composing the data base for the group of fruits, we make sure that every image shows a different fruit type which differs from shape and color and is depicted on a white background (see Fig. 1c). The images were collected from royalty free images on the web.

3.2.2 Password Space. The length of the password (i.e., number of images which are part of the password) and the size of the grids depend on the requirements towards the strength of the password. Microsoft HoloLens requires the user to enter a 6-digit PIN for getting access to the HMD. The password space of a 6-digit PIN is $10^6 = 19.93$ bits. The arrangement of the grid for recognition-based schemes is quadratic [4, 8, 23]. To define the grid size for *Things* on the HoloLens, we checked the visibility of virtual images in the field of view of the user for various quadratic grid sizes. Note, the optical system displays the holographic images in a field of view of 53 degrees. Our pre-studies found out that a grid of up to 16 images was clearly visible to the user, whereas a grid 25 images and above made it difficult to clearly identify the objects and see all images at a glance due to the restricted size of the view field of the HMD. A grid of 9 images would require a password length bigger than 6 for a security level not lower than of the 6-digit PIN. To have the password space as close as to the 6-digit PIN, we chose a grid size of 16 images and a password length of 5 which leads to a password space of $16^5 = 1.048.576 = 20$ bits. Note, the fact that users only select 5 images rather than 6 numbers may also address the fact that in some studies entering secrets on shuffled PIN pads was faster than on shuffled images.

3.2.3 Password Choice. We decided to assign the user a random password due to predictability issues of recognition-based passwords [5] and increased memorability of pictures [26, 27]. Due to randomly chosen passwords the theoretical password space is aligned with the effective password space, i.e. we achieve at least the same security level as with a 6-digit PIN.

3.2.4 Interaction Method. The Microsoft HoloLens 2 provides various interaction methods to navigate in the virtual space and 'manipulate' virtual objects. The following input options are available: (1) gaze by looking at the target, (2) air-tap gestures by touching the holographic image with the finger, and (3) voice by speaking out assigned labels. We decided to use air-tap gestures for selecting the corresponding image. The gesture method is the most similar one to those methods people are familiar with when using wearables or laptops, i.e., touch display or touch pad. Furthermore, it is very similar to the way people interact with PIN pads, e.g., at ATMs. There are also further disadvantages of the other two: The first one may come with privacy issues as some aspects of biometric authentication. The third one would require to put shuffled symbols or numbers on the images which users would read out loud to select an image.

3.3 Enrollment and Authentication

Based on our design decisions, all interactions during both enrollment and authentication are performed via hand gestures. Furthermore, the information provided to users is displayed on the private display of the HMD and is only visible to the user.

The steps of the enrollment process are as follows: First, the user enters the user name by holding a QR code in front of the camera of the HMD and scanning the QR code (see Fig. 1a). Note, we decided to enter the user name by scanning a QR code instead of using the virtual keyboard as entering the name with the virtual keyboard is very cumbersome. Additionally, in the industry context they may have a worker ID which they could scan both for the enrolment as well as for the actual authentication. Once the user confirms the entered name, 5 images are randomly selected by the scheme from its database of 5 times 16 images. Those images are displayed one after the other (see Fig. 1b). The time before switching to the next one is five seconds. The time was chosen according the approach by De Angeli et al. [6]. Additionally, the pictures can also be manually navigated via back and forth buttons.

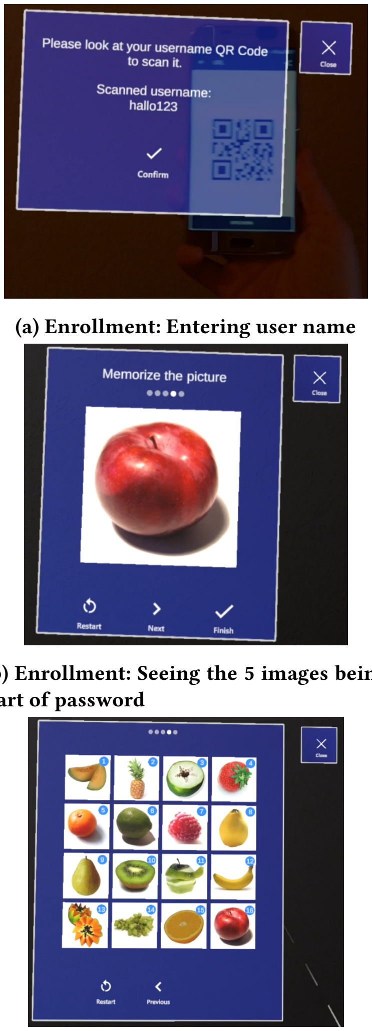
For authenticating on the system, the user needs to enter the images (one per grid) which were assigned during enrollment. Once the user requests to authenticate, first the user ID needs to be provided, e.g., again with the QR Code. Afterwards, the first grid of 16 images is displayed while the images are displayed in a random order. The user selects one image. Then, the scheme displays the next grid of 16 images. If wished, users can correct their selection by going back to the previous grid (or previous grids). After selecting the fifth image, the user is asked to confirm that he/she is done with the authentication. Afterwards the scheme informs him/her whether the password was correct, i.e. he/she can use the HMD or not. If the user made one or more mistakes (while the scheme is not informing about the number of mistakes), he/she can try two more times.

3.4 Shoulder-Surfing Resistance

The threat of shoulder-surfing emerges by the opportunity of an attacker to capture the password of the user by observing the authentication session. When using AR HMD, users' peripheral awareness is limited due to the restricted field of view of the HMD and superimposing virtual objects on the real world. Thus, shoulder-surfing is a feasible threat in the AR context and need to be combated by shoulder-surfing resistant schemes.

Shoulder-Surfing attacks are conducted in public or shared spaces with or without technical equipment [21]. Wiese and Roth [33] differentiate between the following four types of shoulder-surfing attacks: Attackers can directly observe the authentication session without any equipment. Thereby, they are called *opportunistic observers* when they were able to observe only a few authentication sessions and *insider observers* when they were able to observe a high number of authentication sessions. Attacks conducted by recording the authentication session are called *single recording* when only a small amount of recordings were possible and *multiple recording* when a high amount of recordings were done.

In the Things scheme, the position of the images in the grid are randomized for each password entry. Due to the private display of AR HMD, it is not possible for an outsider to observe which of the



Please look at your username QR Code to scan it.

Scanned username: hallo123

(a) Enrollment: Entering user name

Memorize the picture

(b) Enrollment: Seeing the 5 images being part of password

(c) Authentication

Figure 1: Screenshots of the authentication scheme *Things*.

images were chosen. The attacker can only figure out the length of the password. Thus, the scheme protects the user from all four types of shoulder-surfing attacks described in [33].

4 EVALUATION

This section outlines our research questions and corresponding evaluation criteria which are measured in the user study. Next, the user study including the hygiene measures we take when conducting the study, study procedure as well as recruiting and ethics are presented.

4.1 Research Questions and Evaluation Criteria

Our goal is to answer three research questions wrt. the usability (in terms of effectiveness, efficiency and satisfaction) of *Things*. We also added a research question on the perceived usability and the perceived security.

RQ-1: How high is the effectiveness of the Things authentication scheme on the HoloLens when entering the password with air-tap gestures? The effectiveness of the scheme is measured by the success rate to enter the entire secret correctly. The system logs the selected images after confirming the password entry and compares them with the images assigned during enrollment. The user is considered as successful if all selected images match with the password.

RQ-2: How high is the efficiency of the Things authentication scheme on the HoloLens when entering the password with air-tap gestures? The efficiency of the scheme is measured by the time needed to provide the secret. The system logs the time from displaying the first grid till confirming the last entered image.

RQ-3: How high is the satisfaction of users when authenticating with Things on the HoloLens by using air-tap gestures ? Participants degree of satisfaction with *Things* is measured with the System Usability Scale (SUS) [3].

RQ-4: How do the users perceive the usability of the authentication scheme Things when interacting with air-tap gestures on the HoloLens? Perceived usability is measured by asking participants questions on ease of use, ease of remembering the password, login duration and future use of the scheme on the HoloLens. The questions are answered on a Likert scale from 1-5 (1=strongly disagree, 5=strongly agree). The exact questions are available in appendix Sec. B in Fig. 6. Furthermore, to validate the display time of the images during enrollment, we asked questions regarding the appropriateness of the time for memorizing the images.

RQ-5: How do the users perceive the security of the authentication scheme Things when interacting with air-tap gestures on the HoloLens? Perceived security is measured by asking participants how they perceived the security of the scheme in general as well as related to shoulder-surfing attacks by observations from surrounded people. The questions are also answered on a Likert scale from 1-5 (1=very insecure, 5=very secure). The exact questions are available in appendix Sec. B in Fig. 7.

4.2 Study Design

4.2.1 *Hygiene Measures.*

To evaluate the scheme, we conducted the study as a lab experiment in Germany. Due to the current pandemic, we apply the following hygiene measures to minimize the risk for an infection. Once the participant arrived, we checked the immunity status and only allow participation when a proof for immunity by vaccination or previous infection was available. Then, the contact details of the participant were collected for contract tracing in case of a virus outbreak. These precautions were requested by our university from everyone entering the university building at that time due to the pandemic. The participant received disposable gloves and take seat in a room in front of a laptop. The participant was the only person in the room. The instructor communicated with the participant via video call to minimize the time being together. The HoloLens was covered with a plastic bag which was refreshed after every use (see Fig. 3). Sanitizers and disinfecting tissues were made available for the participants. Throughout the study, the participants were required to wear a mask and only take the mask off when using the HoloLens as the glasses would steamed up with the mask and impaired the view of the user. After each experiment the participant room was ventilated for 10 minutes.

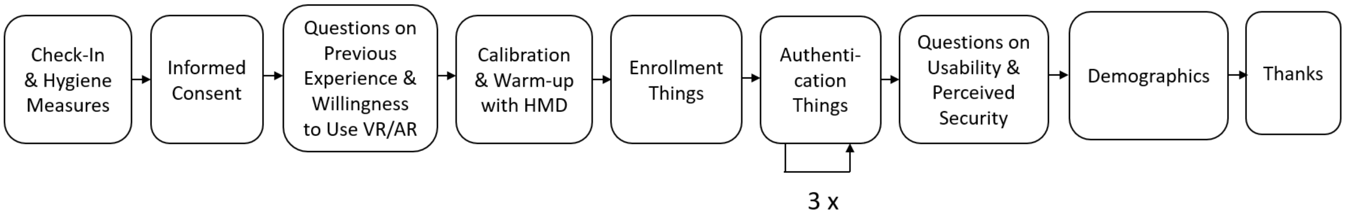

Check-In & Hygiene Measures
Informed Consent
Questions on Previous Experience & Willingness to Use VR/AR
Calibration & Warm-up with HMD
Enrollment Things
Authenti-cation Things
3 x
Questions on Usability & Perceived Security
Demographics
Thanks

Figure 2: Study Procedure.

Figure 3: The HoloLens was covered with a plastic bag and refreshed after every use to keep the HMD clean.

4.2.2 Study Procedure. An overview of the study procedure is depicted in Fig. 2. The nine steps of the user study are described as follows: After the check-in of the participant and applying the hygiene measures described in 4.2.1 (step 'Check-In & Hygiene Measures'), the experiment began with an online survey to be filled out on the laptop provided in the lab room. The survey was implemented in SoSci Survey[1]. First, the participant received some information on the study and consented to participate and processing of their data (step 'Informed Consent'). Afterwards, the survey asked questions on participant's previous experience with Augmented or Virtual Reality HMD, willingness to use in future and experience with authentication on HMD as part of the demographics (step 'Questions on Prev. Experience & Willingness to Use VR/AR'). The exact questions are listed in appendix Sec. A in Fig. 5. After having answered the questions, the participant was prompted to pause with the survey and inform the instructor about it to continue with the next tasks on the HoloLens.

The HoloLens was brought to the participant and positioned properly on participant's head. The user interface on the HMD was shared on the monitor of the instructor to enable a proper navigation of the participant. During the usage of the HoloLens (incl. steps from 'Calibration & Warm-up with HMD' to 'Authentication Things'), the instructions on the tasks of the participant were read out via video call by the instructor. After putting on the HMD, a calibration of the HMD to participant's eyes was conducted. To get used to the interaction methods of the HMD, the participant went through a short training called HoloLens Tips[2] which is preinstalled on the HoloLens. As the scheme requires only gesture interaction, the training covered only gesture control (step 'Calibration & Warm-up with HMD') . After the training, the participant continue with testing the *Things* scheme.

During the enrollment, the participant entered the assigned user ID by scanning the QR code which was available on the participant's table. Before displaying the images, the instructor pointed out to the participant to make sure to memorize the images as it would not be possible to see the password again once the enrollment process is done. Furthermore, it was noted that it will be enough to memorize the object depicted on the image as during authentication only the object needs to be recognized out of different other objects in the grid. Then, a password consisting of five images was displayed sequentially to memorize the password as detailed in 3.3 (step 'Enrollment Things'). Each participant received a different set of random images to avoid bias by the selection of images. The participant completed the enrollment process by going to finish or closing the screen.

Subsequently, the participant started with the authentication session (step 'Authentication Things'). A series of five grids with each 16 images were shown one after the other. The images in the grid were chosen from the data base according the assigned password. For each password the exact same images per grid were chosen, but displayed in a randomized order in each authentication session. The task of the participant was to choose the image belonging to the password for each grid. After selecting the last image, the participant confirmed the input and received a message that the authentication was successful or not successful. Then, the message was closed and the participant received the instruction to repeat the same authentication process two more times. Conducting the authentication session several times will show the participant that the grids are shuffled in each authentication attempt and thus shoulder-surfing attacks due to observations are not possible. After three iterations of authentication, the participant received the instructions to put down the HMD and continue with the survey to evaluate the tested scheme. The survey included the System Usability Scale (SUS) and questions on perceived usability and security as described in 4.1 (step 'Questions on Usability & Perceived Security').

The survey concludes with the questions on demographics, i.e., age, gender, education and employment status (step 'Demographics'). In the end, we thanked participants for attending the study (step 'Thanks'). The study took place in Germany. Thus, all instructions and question were provided in German.

4.3 Recruiting & Ethics

Participants were recruited by advertising the experiment on social networks and distributing flyers. While our institution did not mandate formal ethical approval for this study (this is only

[1] https://www.soscisurvey.de/
[2] https://www.microsoft.com/en-us/p/hololens-tips/9pd4cxkklc47#activetab=pivot:overviewtab

required for studies with especially sensitive participant samples such as children or individuals with disabilities), our methodology conforms to all requirements of our university regarding studies with human participants. In particular, on the first page of the survey, participants received an informed consent by revealing the study's purpose and data processing. For any doubts or questions regarding the study, the instructor was available throughout the study to answer questions. Furthermore, contact information of the researcher were given to the participants for questions after the experiment. Participants had the option to withdraw from the study at any point without providing any reason by closing the tab of their browser with the survey and contacting the instructor. They were also instructed that by cancelling the experiment, all data collected so far would be deleted. Participants were assured that their responses are evaluated in an aggregated and anonymized form, so that no conclusions can be done about their person. Furthermore, the user data which is also handled by SoSci Survey is stored in Germany and thus meet the General Data Protection Regulation (GDPR). Additional biometric data, e.g., iris data which is collected automatically by the HoloLens was deleted after the experiment and was not evaluated. In addition, the glasses were used without internet connection, so that no data exchange can take place between the glasses and Microsoft or third-party providers during use. After the study, the HoloLens was reset to make sure that all captured data is erased from the HMD. Each participant received a compensation of 10 Euros. The compensation was calculated by considering duration of the experiment and the minimum wage in Germany which was at 9.60 Euro at the time. The duration of the experiment was set to 40 minutes based on the average duration of pre-studies. Documents with contact details of the participants which we collected for contact tracing due to Covid were destroyed after 4 weeks, as required for this type of information.

5 RESULTS

This section describes the demographics of the sample and presents the results of the usability and perceived security evaluation of the scheme.

5.1 Participants

After conducting the experiments, none of the participants or instructors reported a Covid infection or any other issue. 16 participants attended the experiment in total. The average age was 24 (SD: 5.91). 56% were male and 44% were female. 75% had already experience with using HMD and all of them had used VR HMD, but had no experience with AR HMD. Of those having experience with HMD, only one participant is using an HMD regularly, the rest is using HMD very rarely or not anymore. But, almost all of them can imagine to use HMD in the future – also including those not having experience with HMD before. Only one participant mentioned to own a VR HMD. All other participants do not own any HMD. None of the participants had ever entered a password on an HMD.

5.2 Research Questions 1-3: Usability

The scheme achieved an overall SUS score of 74 which is categorized as a B-level system standing for "good" usability and above average (> 68)[18].

The average duration for enrollment amounted to 62.21 seconds (SD=24.76). The appropriateness of the time to show pictures of the password in a 5 seconds interval was rated with 3.81 out of 5. But, 40% of the participants would rather increase the time of the interval. The duration for the authentication session for all three iterations in average was 32.2 seconds (SD=9.39). Thereby, the first, second and third iteration took 40.5 seconds (SD=22.9), 30.7 seconds (SD=7.8) and 25.2 seconds (SD=7.5) respectively. The allocation of the duration time in each session and in average is presented in Fig. 4.

The success rate for all three iterations in average was at 90% (SD=19). Overall the first, second and third iteration showed success rates of 81% (SD=39), 100% (SD=0), 88% (SD=33) respectively. Thereby, 12 (75%) succeeded in all three iterations, 3 (19%) succeeded in two iterations, and 1 (6%) succeeded in only one iteration.

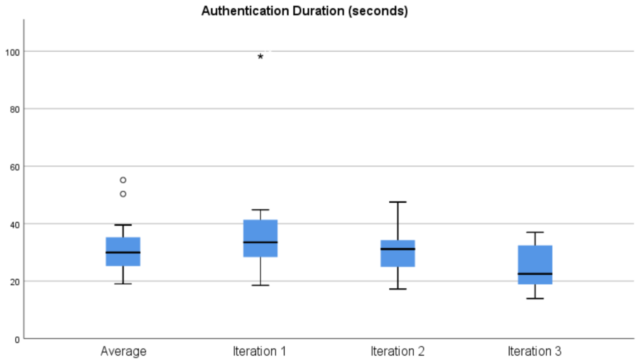


Figure 4: Authentication duration of iterations 1-3 and the average duration of all three iterations. ○=Outlier (farther than 1.5 interquartile ranges, but closer than 3 interquartile ranges). ⋆=Extreme Outlier (farther than 3 interquartile ranges).

5.3 Research Questions 4-5: Perceived Security and Perceived Usability

Participants rated the security of the scheme as follows:

- Perceived security in general: 3.19 (SD=1.01)
- perceived security related to shoulder-surfing: 3.94 (SD=1.09)

The general security of the scheme was perceived as medium which shows that there is some need to increase the security perception. But, the difficulty of guessing the password from observations was rated high. Thus, we assume that the protection mechanism of the scheme by shuffling the images was well understood.

The usability of the scheme was perceived as follows:

- Easy to use: 4.00 (SD=1.12)
- Easy to remember password: 4.31 (SD=0.77)
- Fast login: 3.19 (SD=1.38)
- Future use on HoloLens: 3.25 (SD=1.09)

The ease of use and memorability of the scheme were rated high. This confirms our assumptions in Sec. 3 on the high effectiveness and memorability of recognition-based graphical schemes. The perceived speed during the authentication session as well as willingness to use *Things* on the HoloLens in the future was rated as medium.

6 DISCUSSION

This section discusses the results of the user study by comparing them with results of other studies from the literature. Furthermore, we outline the limitations of our work and propose future research.

6.1 Comparing *Things* with Existing Work

We investigated the usability and perceived security of the recognition-based graphical scheme *Things* applied on the HMD HoloLens. Comparing the scheme's usability performance with other schemes from the literature is not easily possible due to the different security levels, use of other devices, application of other metrics or completely different user study settings. Furthermore, we could not find any literature that investigates the usability of the scheme's authentication process on the HoloLens. Hadjidemtriou et al. conducted a user study for a cued recall-based graphical scheme on the HoloLens, but evaluated only the enrollment process of the scheme [13]. Yet, there are a few studies which follow a similar approach and can help us when assessing our results.

Li et al. [19] investigated the usability of a 6-digit randomized PIN scheme on the HMD Google Glass which has about the same security level as *Things*. Unlike *Things* which uses gesture input, the PIN is entered by tapping on the touch pad positioned on the temple of the glasses. As like in our study, a random secret was assigned to the user. Entering the PIN on the Google Glass took in average 11.2 seconds. The *Things* scheme on the HoloLens was about three times slower. However, the study on the PIN scheme only measured the time for successful attempts out of 6 attempts in total, while our *Things* study also considers the time of failed attempts (all three attempts). The success rate of the PIN scheme was at 98.4% and is about 8% better than *Things*. For perceived usability the following results were obtained for the PIN scheme: "Willingness to use" was agreed by 100%, "easy to learn" was strongly agreed by 50% and "fast to login" strongly agreed by 70%. While the willingness for future use of *Things* was at 3.25 out of 5.0, the PIN scheme received a higher score in this regard. Note, that the PIN study had a within-subject design, where participants tested different input methods, while in our *Things* study the participants tested only a single input method.

Mayer et al. [23] evaluated the *Things* scheme on the PC, but with a password space of 28 bits which is 8 bits more than the *Things* scheme we implemented in our study. The perceived usability of *Things* by Mayer et al. was rated as follows using the same metrics: Easy to use: 4.55; easy to remember: 4.17; and fast to login: 3.77. Regarding to ease of use and perceived authentication time, the scheme on the HoloLens was rated about 0.5 points worse, whereas the memorability of the password was rated similar. The unfamiliar gesture input which is also not as mature as, e.g., a mouse or touch pad might be the cause of the slightly worse evaluation of *Things* on the HoloLens.

Hlywa et al. [14] studied the usability of recognition-based graphical passwords, but without grouping the object images according their semantics. Additionally, they also evaluated pictures of faces. In a within subject study, participants tested the schemes on a web service at home. The grid size and the number of pictures of the password in their study was the same as we used for the *Things* scheme. In Hlywa et al.'s study, the participants needed on average 22.55 seconds (SD=10.02) with object pictures and 35.96 seconds (SD=18.1) with face pictures. The scheme on the HoloLens took about 10 seconds longer to enter than the objects-based scheme on the PC in the study by Hlywa et al. Note, the authors did not mention if the passwords were assigned or chosen by the participants and the number of sessions they have considered for calculating the average duration.

6.2 Limitations and Future Work

The participants of the user study were chosen among young adults which led to an average sample age of 24 years. AR technology is still very new to most people. Thus, participants need to learn first how to interact with AR HMD like the HoloLens. This was confirmed also by our study results, that none of the participants had experience with AR HMD, but had used VR HMD before. As young adults usually have used VR HMD and can adopt their knowledge in VR when using an AR HMD, we decided to keep the sample young. But, the industry employs also elderly people who would need to work with HMD. Thus, it would be interesting to investigate the experience of elderly people with the scheme in future studies. Furthermore, participants were not asked if they have visual impairments. The usability results might have been influenced if participants had difficulties to see the images due to their poor eyesight. But, as it was allowed to wear corrective glasses during the usage of the HMD, we assume this being a minor problem.

We conducted an exploratory study to get some first insights on *Things* applied in the AR context. Thus, we conducted the study with a small sample size of 16 people and did not analyze the results with any inferential statistical tests which would require a larger sample size. For future work, we plan to conduct a more in-depth analysis of *Things* including a comparison with other schemes.

The memorability and effectiveness of the scheme is influenced by the assigned password and the time the participant took to memorize the images. A randomly assigned password is more difficult to memorize than a chosen password. Yet, due to the vulnerability of graphical passwords to guessing attacks [5], we believe a random password needs to be assigned to the user. During enrollment, the instructor gave clear instructions on how to memorize the images. However, it must be acknowledged that the natural behaviour of memorizing a graphical password might be different than the behaviour in a lab environment. Additionally, the images were displayed in a loop (each for 5 seconds) following the approach of [6]. However, an appropriate time for displaying the images needs to be determined with an empirical study. We measured the time participants needed in total for memorizing the pictures during registration. But, we did not measure the time participants took to memorize each image. It can be interesting in future work to compare memorization time for each image with the corresponding authentication performance. Additionally, the number of corrections when entering the password would be interesting to measure in future work as in the current study we only consider if the authentication was successful.

The limitations of the HMD might have influenced the usability performance of the scheme which needs to be investigated separately. For example, participants with corrective glasses wore the HMD on top of their glasses. This might have influenced their experience with the HMD. Also, the input method might have had an effect. The learnability of the gesture input after the training varies among participants and the influence of this factor might impair the usability on HMD, but further research is needed to provide conclusive evidence for this.

Different variants of the scheme would be required to make the scheme usable for individuals with disabilities, e.g., adapting the images for color-blind people or providing other interaction methods for people with reduced mobility.

Comparing the usability results of *Things* with existing literature was only possible to a limited extent due to different security levels of the schemes and different settings in the user studies. Thus, we plan to implement various scheme types, including also a PIN based scheme and analyze their usability in a comparative study.

For future work, design iterations of the *Things* scheme can attempt to optimize the scheme's usability. Thereby, the grid design can be changed regarding the number of displayed images per gird and the arrangement of the images. Furthermore, other interaction methods like speech and gaze input can be tested as gesture input led to long authentication times in the user study.

7 CONCLUSION

This work applies the authentication scheme *Things* from previous work on the AR HMD Microsoft HoloLens to protect the user from shoulder-surfing attacks during authentication. *Things* is a recognition-based graphical scheme that lets the user select an image out of decoy images which are displayed in a randomized order for each authentication session. Due to the private display of the HMD the images are not visible to outside observers. We conducted a lab study to investigate the scheme's usability and perceived security. The measured usability aspects include effectiveness, efficiency, satisfaction, perceived usability, and perceived security. Due to the good memorability of graphical passwords, the scheme shows a high effectiveness. According to the System Usability Scale (SUS), the scheme can be categorized as good with a score of 74. Yet, the long authentication duration in the *Things* scheme needs to be improved in future work. This might be done by switching to an alternative interaction method, i.e., replacing the gesture input with speech or eye-gaze.

ACKNOWLEDGMENTS

This research was supported by funding from the topic Engineering Secure Systems, subtopic 46.23.01 Methods for Engineering Secure Systems, of the Helmholtz Association (HGF) and by KASTEL Security Research Labs.

REFERENCES

[1] Mozhgan Azimpourkivi, Umut Topkara, and Bogdan Carbunar. 2017. Camera based two factor authentication through mobile and wearable devices. *Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–37.

[2] Daniel V Bailey, Markus Dürmuth, and Christof Paar. 2014. "Typing" passwords with voice recognition: How to authenticate to Google Glass. In *Proc. of the Symposium on Usable Privacy and Security.* Citeseer, Citeseer, CA, USA, 1–2.

[3] John Brooke. 1996. Sus: a "quick and dirty'usability. *Usability evaluation in industry* 189, 3 (1996), 189–194.

[4] Sacha Brostoff and M Angela Sasse. 2000. Are Passfaces more usable than passwords? A field trial investigation. In *People and computers XIV—usability or else!* Springer, London, 405–424.

[5] Darren Davis, Fabian Monrose, and Michael K Reiter. 2004. On user choice in graphical password schemes.. In *USENIX security symposium*, Vol. 13. USENIX Association, USA, 11–11.

[6] Antonella De Angeli, Lynne Coventry, Graham Johnson, and Karen Renaud. 2005. Is a picture really worth a thousand words? Exploring the feasibility of graphical authentication systems. *IJHCS* 63, 1-2 (2005), 128–152.

[7] Reyhan Duezguen, Peter Mayer, Sanchari Das, and Melanie Volkamer. 2020. Towards secure and usable authentication for augmented and virtual reality head-mounted displays. In *WAY Workshop.* WAY, Virtual, 1–6.

[8] Paul Dunphy, Andreas P Heiner, and N Asokan. 2010. A closer look at recognition-based graphical passwords on mobile devices. In *Proceedings of the Sixth Symposium on Usable Privacy and Security.* ACM, New York, NY, USA, 1–12.

[9] Rainhard Dieter Findling, Tahmid Quddus, and Stephan Sigg. 2019. Hide my gaze with EOG! towards closed-eye gaze gesture passwords that resist observation-attacks with electrooculography in smart glasses. In *International Conference on Advances in Mobile Computing & Multimedia.* ACM, NY, USA, 107–116.

[10] Eira Friström, Elias Lius, Niki Ulmanen, Paavo Hietala, Pauliina Kärkkäinen, Tommi Mäkinen, Stephan Sigg, and Rainhard Dieter Findling. 2019. Free-Form Gaze Passwords from Cameras Embedded in Smart Glasses. In *International Conference on Advances in Mobile Computing & Multimedia.* ACM, USA, 136–144.

[11] Markus Funk, Karola Marky, Iori Mizutani, Mareike Kritzler, Simon Mayer, and Florian Michahelles. 2019. Lookunlock: Using spatial-targets for user-authentication on hmds. In *Extended Abstracts CHI.* ACM, NY, USA, 1–6.

[12] Gabriela Gheorghe, Nicolas Louveton, Benoît Martin, Benjamin Viraize, Louis Mougin, Sébastien Faye, and Thomas Engel. 2016. Heat is in the eye of the beholder: Towards better authenticating on smartglasses. In *2016 9th International Conference on Human System Interactions (HSI).* IEEE, Portsmouth, UK, 490–496.

[13] George Hadjidemetriou, Marios Belk, Christos Fidas, and Andreas Pitsillides. 2019. Picture passwords in mixed reality: Implementation and evaluation. In *Extended Abstracts of the 2019 CHI.* ACM, NY, USA, 1–6.

[14] Max Hlywa, Robert Biddle, and Andrew S Patrick. 2011. Facing the facts about image type in recognition-based graphical passwords. In *Proceedings of the 27th Annual Computer Security Applications Conference.* ACM, New York, NY, USA, 149–158.

[15] Ben Hutchins, Anudeep Reddy, Wenqiang Jin, Michael Zhou, Ming Li, and Lei Yang. 2018. Beat-pin: A user authentication mechanism for wearable devices through secret beats. In *ASIACCS.* ACM, NY, USA, 101–115.

[16] MD Rasel Islam, Doyoung Lee, Liza Suraiya Jahan, and Ian Oakley. 2018. Glass-pass: Tapping gestures to unlock smart glasses. In *AH'18.* ACM, NY, USA, 1–8.

[17] Christopher Kreider. 2018. The Discoverability of Password Entry Using Virtual Keyboards in an Augmented Reality Wearable: An Initial Proof of Concept. In *Southern Association for Information Systems.* AISeL, UK, 1–6.

[18] James R Lewis and Jeff Sauro. 2018. Item benchmarks for the system usability scale. *Journal of Usability Studies* 13, 3 (2018), 158–167.

[19] Yan Li, Yao Cheng, Weizhi Meng, Yingjiu Li, and Robert H Deng. 2021. Designing leakage-resilient password entry on head-mounted smart wearable glass devices. *IEEE Transactions on Information Forensics and Security* 16 (2021), 307–321.

[20] Yingjiu Li, Qiang Yan, and Robert H Deng. 2015. ShadowKey: A Practical Leakage Resilient Password System. In *Leakage Resilient Password Systems.* Springer, Cham, 53–64.

[21] Federico Maggi, Alberto Volpatto, Simone Gasparini, Giacomo Boracchi, and Stefano Zanero. 2011. A fast eavesdropping attack against touchscreens. In *2011 7th International Conference on Information Assurance and Security (IAS).* IEEE, Melacca, Malaysia, 320–325.

[22] Peter Mayer and Melanie Volkamer. 2018. Addressing misconceptions about password security effectively. In *STAST.* ACM, NY, USA, 16–27.

[23] Peter Mayer, Melanie Volkamer, and Michaela Kauer. 2014. Authentication schemes-comparison and effective password spaces. In *International Conference on Information Systems Security.* Springer, Cham, 204–225.

[24] Wendy Moncur and Grégory Leplâtre. 2007. Pictures at the ATM: exploring the usability of multiple graphical passwords. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM, New York, NY, USA, 887–894.

[25] Volker Paelke. 2014. Augmented reality in the smart factory: Supporting workers in an industry 4.0. environment. In *Proceedings of the 2014 IEEE emerging technology and factory automation (ETFA).* IEEE, Barcelona, Spain, 1–4.

[26] Allan Paivio and Kalman Csapo. 1973. Picture superiority in free recall: Imagery or dual coding? *Cognitive psychology* 5, 2 (1973), 176–206.

[27] Allan Paivio, Timothy B Rogers, and Padric C Smythe. 1968. Why are pictures easier to recall than words? *Psychonomic Science* 11, 4 (1968), 137–138.

[28] Pranav Parekh, Shireen Patel, Nivedita Patel, and Manan Shah. 2020. Systematic review and meta-analysis of augmented reality in medicine, retail, and games. *Visual computing for industry, biomedicine, and art* 3, 1 (2020), 1–20.

[29] Sebeom Park, Shokhrukh Bokijonov, and Yosoon Choi. 2021. Review of Microsoft HoloLens Applications over the Past Five Years. *Applied Sciences* 11, 16 (Jan. 2021), 7259. Number: 16 Publisher: Multidisciplinary Digital Publishing Institute.

[30] Roberto Pierdicca, Emanuele Frontoni, Rama Pollini, Matteo Trani, and Lorenzo Verdini. 2017. The use of augmented reality glasses for the application in industry 4.0. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, Cham, 389–401.

[31] Hwajeong Seo, Jiye Kim, Howon Kim, and Zhe Liu. 2017. Personal identification number entry for Google glass. *C&EE* 63 (2017), 160–167.

[32] Daphna Weinshall and Scott Kirkpatrick. 2004. Passwords you'll never forget, but can't recall. In *CHI'04 extended abstracts on Human factors in computing systems*. ACM, New York, NY, USA, 1399–1402.

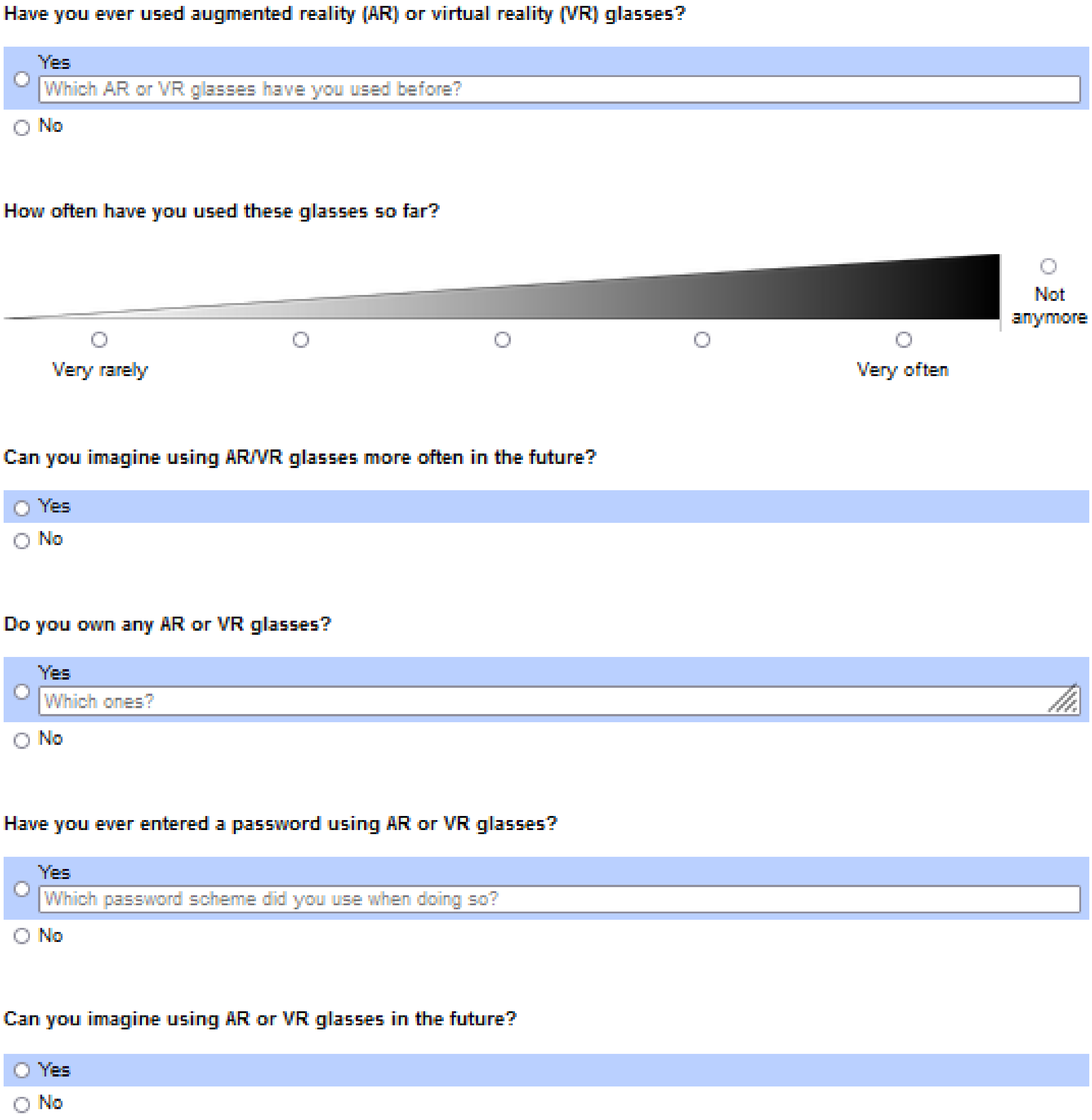
[33] Oliver Wiese and Volker Roth. 2016. See you next time: A model for modern shoulder surfers. In *MobileHCI*. ACM, Florence, Italy, 453–464.

[34] Dhruv Kumar Yadav, Beatrice Ionascu, Sai Vamsi Krishna Ongole, Aditi Roy, and Nasir Memon. 2015. Design and analysis of shoulder surfing resistant pin based authentication mechanisms on google glass. In *FC*. Springer, Germany, 281–297.

[35] Shanhe Yi, Zhengrui Qin, Ed Novak, Yafeng Yin, and Qun Li. 2016. Glassgesture: Exploring head gesture interface of smart glasses. In *International Conference on Computer Communications*. IEEE, CA, USA, 1–9.

[36] Ruide Zhang, Ning Zhang, Changlai Du, Wenjing Lou, Y Thomas Hou, and Yuichi Kawamoto. 2017. AugAuth: Shoulder-surfing resistant authentication for augmented reality. In *ICC*. IEEE, Paris, France, 1–6.

A SURVEY QUESTIONS: PREVIOUS EXPERIENCE



Have you ever used augmented reality (AR) or virtual reality (VR) glasses?

○ Yes
Which AR or VR glasses have you used before?

○ No

How often have you used these glasses so far?

○ Very rarely ○ ○ ○ ○ Very often ○ Not anymore

Can you imagine using AR/VR glasses more often in the future?

○ Yes

○ No

Do you own any AR or VR glasses?

○ Yes
Which ones?

○ No

Have you ever entered a password using AR or VR glasses?

○ Yes
Which password scheme did you use when doing so?

○ No

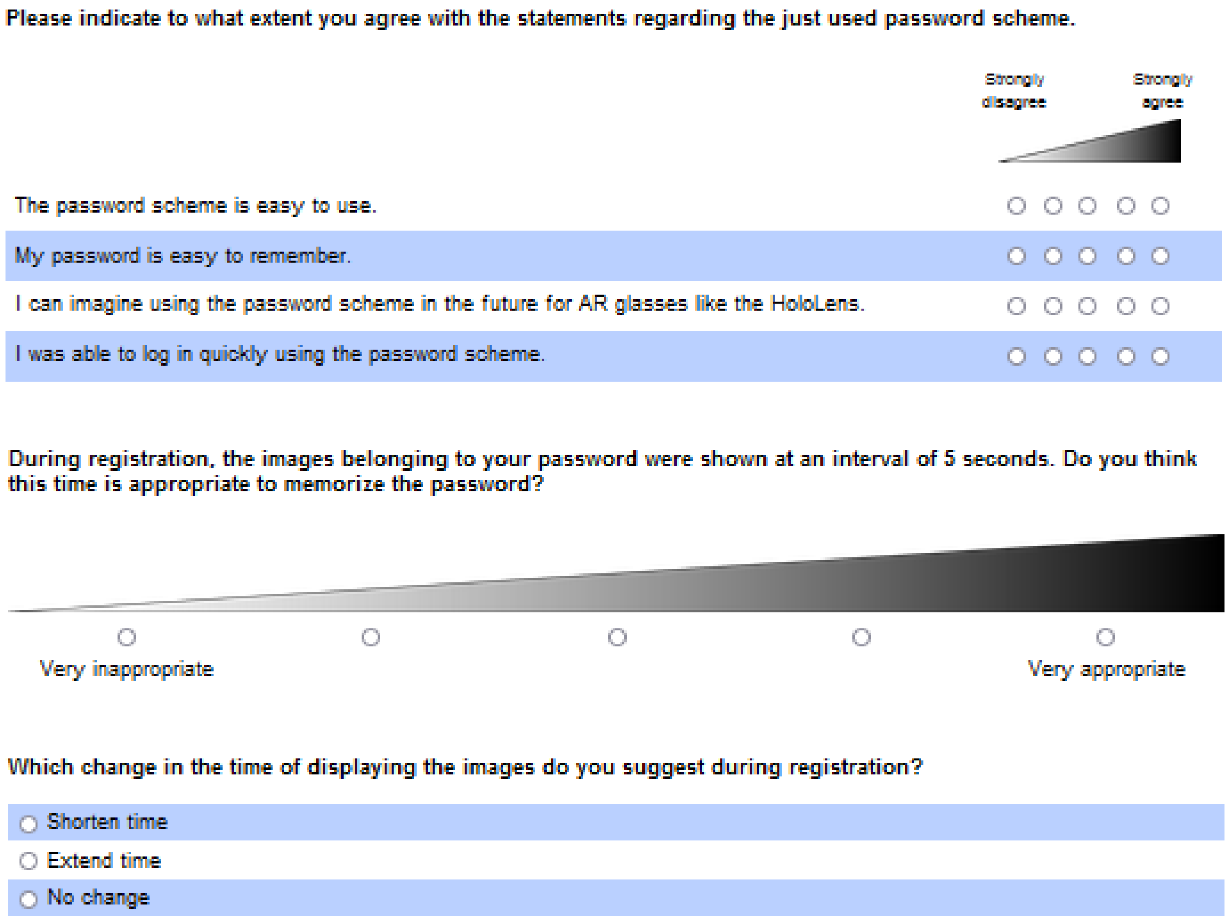
Can you imagine using AR or VR glasses in the future?

○ Yes

○ No

Figure 5: Survey questions on previous experience with AR and VR HMD.

B SURVEY QUESTIONS: PERCEIVED USABILITY AND SECURITY



Please indicate to what extent you agree with the statements regarding the just used password scheme.

	Strongly disagree				Strongly agree
The password scheme is easy to use.	○	○	○	○	○
My password is easy to remember.	○	○	○	○	○
I can imagine using the password scheme in the future for AR glasses like the HoloLens.	○	○	○	○	○
I was able to log in quickly using the password scheme.	○	○	○	○	○

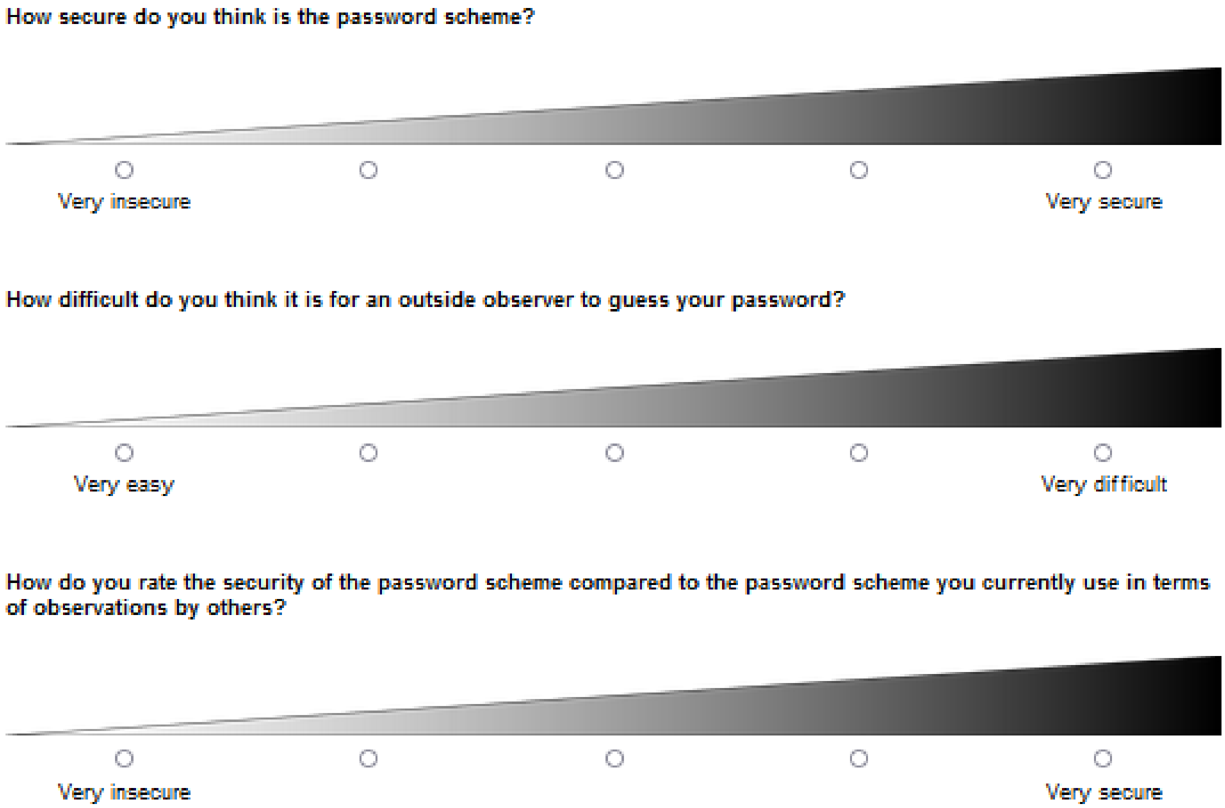
During registration, the images belonging to your password were shown at an interval of 5 seconds. Do you think this time is appropriate to memorize the password?

○ Very inappropriate ○ ○ ○ ○ Very appropriate

Which change in the time of displaying the images do you suggest during registration?

○ Shorten time
○ Extend time
○ No change

Figure 6: Survey questions on perceived usability.



How secure do you think is the password scheme?

○	○	○	○	○
Very insecure				Very secure

How difficult do you think it is for an outside observer to guess your password?

○	○	○	○	○
Very easy				Very difficult

How do you rate the security of the password scheme compared to the password scheme you currently use in terms of observations by others?

○	○	○	○	○
Very insecure				Very secure

Figure 7: Survey questions on perceived security.