Understanding Bystanders' Tendency to Shoulder Surf Smartphones Using 360-degree Videos in Virtual Reality

ALIA SAAD, University of Duisburg-Essen, Germany
JONATHAN LIEBERS, University of Duisburg-Essen, Germany
UWE GRUENEFELD, University of Duisburg-Essen, Germany
FLORIAN ALT, Bundeswehr University Munich, Germany
STEFAN SCHNEEGASS, University of Duisburg-Essen, Germany

(a) Standing-Standing (b) Standing-Sitting (c) Sitting-Standing (d) Sitting-Sitting

Fig. 1. Staged videos: (top) panoramic view of a 360-degree video frame, and (bottom) corresponding participant's point of view, in the four different observer-user postures.

Shoulder surfing is an omnipresent risk for smartphone users. However, investigating these attacks in the wild is difficult because of either privacy concerns, lack of consent, or the fact that asking for consent would influence people's behavior (e.g., they could try to avoid looking at smartphones). Thus, we propose utilizing 360-degree videos in Virtual Reality (VR), recorded in staged real-life situations on public transport. Despite differences between perceiving videos in VR and experiencing real-world situations, we believe this approach to allow novel insights on observers' tendency to shoulder surf another person's phone authentication and interaction to be gained. By conducting a study (N=16), we demonstrate that a better understanding of shoulder surfers' behavior can be obtained by analyzing gaze data during video watching and comparing it to post-hoc interview responses. On average, participants looked at the phone for about 11% of the time it was visible and could remember half of the applications used.

CCS Concepts: • **Human-centered computing** → **Virtual reality**; *User studies*; • **Security and privacy** → *Usability in security and privacy*.

Manuscript submitted to ACM

Additional Key Words and Phrases: Virtual Reality; Shoulder Surfing; Eye Tracking; Omnidirectional Videos

ACM Reference Format:

Alia Saad, Jonathan Liebers, Uwe Gruenefeld, Florian Alt, and Stefan Schneegass. 2021. Understanding Bystanders' Tendency to Shoulder Surf Smartphones Using 360-degree Videos in Virtual Reality. In *23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '21), September 27-October 1, 2021, Toulouse & Virtual, France.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3447526.3472058

1 INTRODUCTION

Shoulder surfing is one of the most prominent privacy threats smartphone users are facing. According to a survey by Eiband et al. [6], shoulder surfing is most likely to happen during users' daily commutes in public transportation. However, they also noted that attacks happen in a much wider range of situations and contexts and that there are many different factors that contribute to the success of shoulder surfing attacks – such as the spatial relationship between observer and victim. Participants reported their feelings on being observed as violated, harassed, vulnerable, and uncomfortable. Moreover, while most prior work focused on shoulder surfing during authentication events, it is today well understood, that also other content on the smartphone is subject to shoulder surfing.

At the same time, it remains challenging for researchers to conduct studies that allow shoulder surfing incidents to be more holistically understood. Such research would yet be valuable since an in-depth understanding of shoulder surfing situations could allow researchers to design novel concepts with the objective to better protect users and their data.

Studies involving someone observing another individual during authentication and interaction have mainly been conducted either in controlled environments or via surveys [6, 13]. In-the-wild studies are much less popular, primarily due to privacy and ethical concerns [34]. These limitations exist because recording shoulder surfing events would disclose observers' identities without their approval, and looking at others' phones without prior consent would invade their privacy. Also, asking for consent might reveal the study's purpose and likely result in unnatural behavior.

In this work we investigate the use of omnidirectional or 360° videosin Virtual Reality for investigating shoulder surfing. This has been shown to be a promising technology when it comes to investigating immersive experiences and novel user interfaces [35]. Due to the high resolution provided by both cameras and headsets, spherical videos have opened the door to more studies and real-life simulations. A particular strength we see with this approach is the fact that it allows different aspects of shoulder surfing incidents, such as the content involved or the spatial relationship between observers and victims, to be investigated in a controlled manner. In addition, the approach also allows sensing technology to be used that is difficult to deploy in the real world, such as eye-tracking. In this way, it becomes possible to obtain a more nuanced understanding of, for example, where and for how long observers direct their attention [8].

To demonstrate both the feasibility as well as understand the challenges and opportunities of this approach, We focused on an investigation of shoulder surfing incidents in public transportation. In particular, we replayed situations which were reported in prior work [6] to likely elicit shoulder surfing. The camera capturing the observer's view was placed at two different heights to imitate the average human eye level height in both standing and sitting positions. At the same time, the phone user was either sitting or standing, resulting in overall four different investigated situations. We later showed these recordings to participants on a VR headset equipped with an eye tracker to understand when shoulder surfing behavior takes place in near real-life situations without compromising the observer or smartphone user's privacy. Observing users' eye movements and analyzing where they directed their attention to allowed us to obtain an in-depth understanding of which situations mostly put users at risk.

Our findings show that all participants glanced at the phone at least once through the four different videos, with an average gaze duration of 5.3 seconds, and 87% of them could identify at least one application in use. Considering the presence experience in 360° videos, participants reported a general presence of 4.13, on the IGROUP presence questionnaire [28], which suggests that we managed to create a realistic experience for participants.

Contribution Statement. We contribute a study on shoulder surfing, using an approach neither compromising users' privacy nor strongly influencing adversaries' natural behavior, that is 360° videos. We recorded 360° videos, displaying them on a head-mounted display, and evaluated attention and gaze behavior towards the phone using eye-tracking.

2 BACKGROUND AND RELATED WORK

Our work is based on several strands of prior research, most notably observative user-centered attacks, known as shoulder surfing attacks, Virtual Reality applications with a focus on 360° videos, and eye tracking in virtual environments.

2.1 Shoulder Surfing

Earlier studies investigating observation attacks focused on three main aspects: understanding the attacks and their likelihood [6, 7, 20], communicating the incident to users [3, 25], and designing efficient countermeasures [9, 11, 14, 15]. In the following, we discuss these aspects in more detail.

Due to several constraints, typical studies focusing on understanding shoulder surfer's behavior and phone unlocking patterns are conducted in the form of collective surveys [6, 7]. Their main findings show an opportunistic rather than a deliberate attitude of shoulder surfers and that curiosity and boredom accounts for the majority of shoulder surfing incidents. However, defining the risk as authentication methods disclosure might not be sufficient, as observing the overall interaction with the smartphone reveals personal and private information that should also be kept concealed. Concerning communicating the attack event to users, prior work used face recognition to count the faces from the scene captured by the phone front camera [25]. Vibro-tactile feedback was found to be a convenient approach to alert users in comparison to other visual feedback methods, such as LED flashlights and camera preview. Lastly, multiple studies investigated different approaches to overcome and prevent shoulder surfing attacks. Some focused on developing different lock pattern interfaces [14, 22, 33], while others considered combining gaze tracking methods along with traditional authentication methods such as PINs or passcodes to hinder observers from deriving the unlocking patterns [9–11]. These approaches focus entirely on authentication, such as PIN, patterns, passwords/codes, but do not provide solutions for attackers glancing at personal and private information. When previous work addressed different standing-sitting postures between the user and observer, it was either in the form of an assumption to the threat model [4], or a shaking action to be performed by the user as an authentication method [38]. Additionally, prior research focuses on surveys and user reports gaining information on the context of potential shoulder surfing incidents whereas we investigate user's gaze behavior gaining different aspects.

2.2 Measuring Visual Attention

Based on the hypothesis by Just and Carpenter [8] that "the locus of the eye fixations reflects what is being internally processed," several studies adopted eye-tracking to understand human attention. Lutz et al. discussed the potential of adding eye-tracking to achieve accurate attention measurements in real-time [16]. Previous studies proposed developing Visual attention models estimating saliency maps to predict video regions drawing a viewer's attention in 360° videosin

VR [21, 31]. Eye-tracking systems were also used by Schrom-Feiertag et al. to evaluate guidance systems and navigation solutions in VR [26, 27]. Their findings showed similar user behavior, specifically in attention and decision-making.

2.3 Virtual Reality as a Research Platform

Employing Virtual Reality (VR) to simulate real-life situations is a well-studied research field. Such studies are of particular value in situations, where in-situ studies would impose a risk on participants. Moussaïd et al. used VR to examine crowd behavior during high-stress evacuations [19]. Deb et al. developed a virtual signalized road intersection to examine pedestrian behavior [5]. Other researchers investigated approaches to enhance realistic responses in immersive virtual environments [30, 36]. Previous work evaluated presence and usability in VR, comparing it to real life situations [1, 2]. Although both studies aimed at building a virtual world looking similar to the real world, they reported differences in terms of presence and usability. Yet, both virtual environments provided a realistic impression so that these can be an alternative to real environments. Recently, researchers are investigating alternative approaches to evaluate their studies, mainly in virtual environments. Mäkelä et al. compared users behavior between virtual and real environments to evaluate performance of public displays [17]. They observed similar behavior in terms of perception and engagement, confirming that virtual field studies can indeed yield meaningful result comparable to real-world investigations. On a larger scale, Voit et al. compared five different empirical study methods: online, virtual reality, augmented reality, lab setup and in-situ [32]. Interestingly, their results showed similar responses from participants for in-situ and Virtual Reality. Consequently, they believe that researchers could benefit from the advantages VR offers, such as complete environment control. Continuing this strain of work, researchers validated the use of lab-based VR setup for evaluating an authentication system, by comparing it with a real-world study [12, 18]. They found no significant differences between both setups, especially in perceived workload, and qualitative feedback. These findings highlight the feasibility of Virtual Reality as a research platform to investigate real-world systems and scenarios. Hence, we believe VR is promising for investigating users' tendency to should surf by replicating real-world situations.

3 STUDY

Understanding bystander behavior and what triggers them to observe other people's phones were mainly studied in collective surveys. Due to privacy restrictions, conducting in-the-wild shoulder surfing studies is a difficult task. Based on previous work, we use 360° videosin VR to obtain first insights into peoples' approaches to shoulder surf. Our driving research question are: **(R1)** *What triggers shoulder surfing attacks?* **(R2)** *What is the influence of observers' and victims spatial relationship?* We evaluate bystander behavior by analyzing the collected eye-tracking data and comparing it to the responses gathered from the semi-structured interviews. We recorded videos in real-world settings and later conducted the study with participants in a controlled lab environment. In this study, we use the term *user* to define the person being shoulder surfed and the term *observer* to define the bystander/shoulder surfer.

3.1 Study Design

We conducted a within-subject controlled lab study in Virtual Reality using the eye-tracking Pico Neo 2 VR headset[1]. To investigate users' tendency to shoulder surf, the videos presented to the participants are all recorded from the point of view of the observer. To examine possible posture and content variations, our two independent variables are observer-user postures (4 levels – standing-standing, standing-sitting, sitting-standing, and sitting-sitting) and commonly used phone applications (4 levels – WhatsApp, Gallery, a Flag Game, and Facebook), with a total number

[1]Pico Neo 2 headset. https://www.pico-interactive.com/us/neo2.html, last accessed on February 1, 2021.

of sixteen different combinations. The postures and applications sequences are counterbalanced using a Latin square design. We did not provide a specific task to the participants as this would not reflect what participants intrinsically wanted to do but rather what we ask them to do. Furthermore, following the analysis of public transport commuters behavior by Zhang and Timmermans [37], it is reported that the majority of public transport users do either window gazing / people watching (31.3%), or nothing / unknown activities (34.8%). We examine participants gaze in terms of both glance count (i.e., number of individual fixations on the mobile phone) and duration (i.e., the time of each fixation on the phone display), in addition to subjective insights and responses collected via post-study semi-structured interviews.

3.2 Procedure

To conceal the research question, the experimenter explained that the purpose of the study was to assess the integration between the 360° camera and the eye-tracking features embedded in the headset. Hence, participants were asked to behave as they naturally would in a similar real environment. After signing the consent form and filling out questionnaires about demographics and familiarity with VR applications, participants put on the headset and use the built-in Tobii calibration app to ensure accurate reporting of eye-tracking data. Our application consists of four different sets, where each set contains one nature video and one underground video, in that order (see video acquisition subsection for details). Before each sequence, a message was displayed to instruct the user to stand up or sit down. After each set, the experimenter would ask general questions about the experience. Next, participants selected the application. A brief experience questionnaire followed each set of two videos. Once done with all four sets, the participant answered an IGROUP Presence Questionnaire (IPQ). The last phase of the study was a semi-structured interview, in which participants were asked questions about shoulder surfing, and a self-assessment and comparison of their behavior in real and virtual environments. We finally revealed the purpose of the study. The average duration of the study was 50 minutes.

3.3 Apparatus

We used the Unity3D[2] game development platform to enable the participants to view the recorded videos in a VR headset in 360°, to control over the procedure of the study (the order of videos and interview questions), to seamlessly transition between scenes, and to record data. Recorded videos were added to the inside surface of a spherical object in the center of which the camera was placed. Afterward, we used the PicoVR Unity SDK[3] to integrate the eye-tracking features of the Pico Neo 2 headset. The HMD offers a 4k resolution at a 75 Hz refresh rate and a field of view of 101°. In our implementation, a glance is counted when the gaze points are matched to the phone area on the sphere, we did not use any data from the VR headset.

3.4 Video Acquisition

One of the main findings of the survey conducted by Eiband et al. was that *"shoulder surfing was most common among strangers, in public transport, during commuting times"* [6]. Therefore, we chose a public transport setting for our research as well. In order to make the experience as realistic and immersive as possible, we used a 360° 3D camera, namely the Vuze+ plus 3D 360° camera[4]. Upon approval from the official authorities, the recording took place in an underground commuter train at 5 am on four different days. We chose the underground and early time settings for the

[2]Unity3D. https://unity.com, last accessed on February 1, 2021.
[3]PicoVR Unity SDK. https://developer.pico-interactive.com/sdk, last accessed on February 1, 2021.
[4]Vuze+ plus 3D camera. https://store.vuze.camera/buy/vuze-plus-camera, last accessed on February 1, 2021.

controllable lighting conditions. For privacy reasons, two students volunteered to act as regular commuters, and we recorded in parts of the subway where other passengers could not be identified. To distract participants from the main purpose of the study, we recorded another set of shorter videos in outdoor settings. The outdoor videos were recorded in two locations: a woodland (i.e. small forest area), no people perceived, and a stranded area by a river, in some videos, people could be seen from a distance but not identified. The average duration of underground videos is 164.18 seconds ($SD = 17.65s$), whereas the average nature video duration is 52.5 seconds ($SD = 23.19s$).

We placed the camera, as well as the user, at four different spots, each representing a different setting. Thus, we replicated four typical observer-user positioning options: standing-standing, standing-sitting, sitting-standing, and sitting-sitting (cf., Figure 1). The camera height was set to mimic an average human eye height, both in standing and sitting positions. As a result, the camera was set at a height of 175 cm and 115 cm, respectively. These numbers go in accordance withe findings by Rothe et al., where differences of 10 cm between camera height and participant's eye height are acceptable and do not trigger disturbance or sickness while wearing the headset [24]. Since the camera captures scenes omnidirectionally, i.e. from all four sides, we placed the user in the middle of a camera view at one of the sides to avoid obvious stitch lines. Additionally, refined stitching was made during the video rendering process. For auditory effects, each video's background noise was recorded in a stereo format, which provides depth perception to the headset user. In each recorded session, the video started with three people sitting or standing in the underground train. The user then took the phone out of his pocket, unlocked the device with an Android unlock pattern, and started to use four different applications. We chose four applications that each require distinctive hand gestures: social media (scrolling), messaging (typing), gallery viewing (swiping), and a game (tapping). These applications were also reported by prior surveys [6, 20]. In each recording day, we changed the order in which the applications were used. The video sequences between days and postures were counterbalanced using a Latin square design.

3.5 Participants

We invited 16 participants (13 males, 3 females), age between 20 and 42 (M= 27.1, SD= 5.7) to our lab. Height average is 177.93cm, SD = 8.29cm. None of them suffered from Color Vision Deficiency (CVD). When conducting the study, all the participants vision was either normal or corrected by eyeglasses or contact lenses. Recruitment was conducted via mailing lists and social networks.

4 RESULTS

In our results, we report mean (M), median (Md), and inter-quartile range (IQR). We do not assume normal distribution of our data, and hence, perform non-parametric tests and report IQR instead of standard deviation.

4.1 Analyzes of Gaze Behavior

Due to technical issues with the eye-tracking, we had to remove one participant (P2) from our gaze behavior analysis. For the gaze behavior, we consider each glance at the phone for longer than one second to be a valid shoulder surfing incident. Moreover, we count gaze behavior as two distinct shoulder surfing attacks when at least three seconds passed since the last time the participant was looking at the smartphone. For the analysis of the gaze behavior, we consider one target smartphone used by the person closest to the participants in all videos.

4.1.1 Number of Eye Contacts with Phone. All participants gazed at the phone at least once in the four presented videos. On average over all four videos, each participant glanced at least 6.73 times at the smartphone in the video

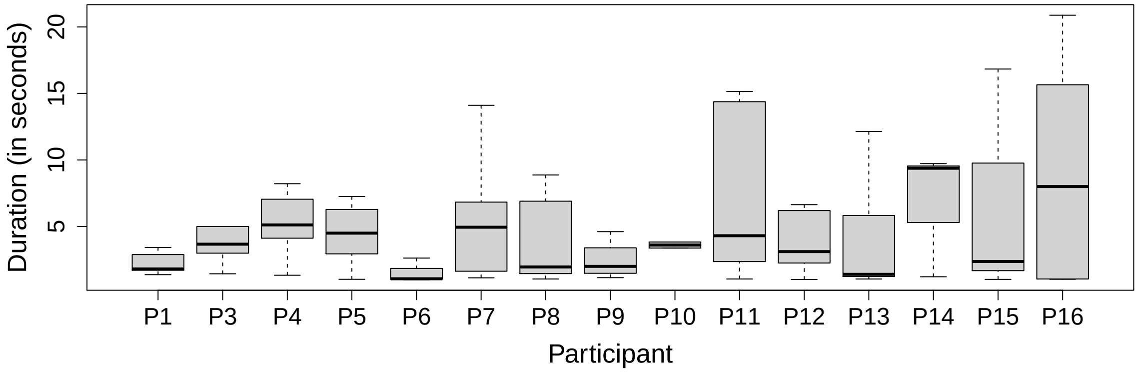
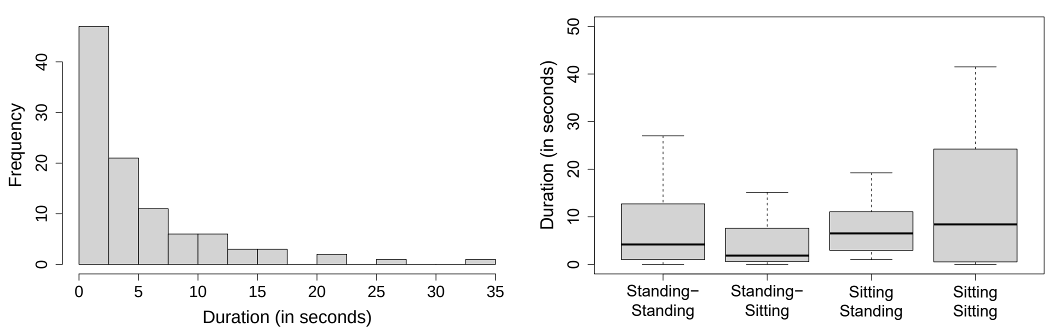


Fig. 2. Boxplots illustrating the variance of gaze duration among the participants.

(M = 1.68 times per video). The maximum number of eye contacts was 11 times (P7, P9), while the minimum count was 2 times (P10). To further analyze our data, we looked at the average number of phone gazes of each participant per condition in descending order (mean for each video): *Sitting-Standing* = 2.33 (IQR = 1.5), *Standing-Standing* = 1.73 (IQR = 1.5), *Sitting-Sitting* = 1.67 (IQR = 1.5), and *Standing-Sitting* = 1.00 (IQR = 0.5). We performed a Friedman test that revealed a significant effect of condition on the number of phone gazes ($\chi^2(3) = 8.60$, $p = .035$, $N = 16$). A post-hoc test using Wilcoxon Signed-rank with Bonferroni correction showed a significant difference between *Standing-Sitting* and *Sitting-Standing* ($W = 6$, $Z = -2.59$, $p = .046$, $r = .47$) (r: > 0.1 small, > 0.3 medium, and > 0.5 large effect). We can therefore say that sitting participants gazed at a standing person's smartphone significantly more often than a standing participant looked at the phone of a sitting smartphone user.

4.1.2 Duration of Eye Contact with Phone. On average, participants continued to look at the smartphone in the video for a duration of 5.30 s (IQR = 5.25s). Each of the videos showed a smartphone for 100 seconds, while each video was 150 seconds in total (so there were 50 s in which a smartphone was not visible). During each video, participants looked at the smartphone for 11.16 s on average. The durations for each participant are shown in Figure 2.

The maximum duration of a participant looking at the smartphone was 33.96s. While most of the time participants looked for less than five seconds. All durations and their frequencies can be seen in Figure 3(a).



(a) Frequency of the durations for gazing at the phone.

(b) Durations of phone gazes for each condition.

Fig. 3. Histogram and Boxplot showing the durations of phone gazes and their frequencies.

To understand the influence of the conditions, we further analyzed the gaze duration of the participants in each condition in descending order (mean for each video): *Sitting-Sitting* = 13.08s (IQR = 23.71s), *Standing-Standing* = 8.27s (IQR = 13.75s), *SittingStanding* = 9.31s (IQR = 8.11s), and *Standing-Sitting* = 5.06s (IQR = 7.03s). We depict the duration of each condition in Figure 3(b). To analyze the duration for each condition, we performed a Friedman test that revealed no significant effect of condition on gaze duration ($\chi^2(3) = 2.39$, $p = .495$, $N = 16$).

4.2 Questionnaire and Interviews

Besides the eye-tracking data, we conducted a presence questionnaire after all conditions were finished to understand the quality of the experience, and interviews to get detailed insights into participants' behavior and shoulder surfing.

4.2.1 Presence Questionnaire. To get insights into the experienced presence of the participants throughout the 360-degree videos, we conducted the IGROUP Presence Questionnaire (IPQ) after all videos were done [28]. The IPQ rates the presence in four subscales, and was rated by the participants as follows: general presence (4.13), spatial presence (2.75), involvement (3.63), and experienced realism (2.84). The results indicate a good general presence and involvement, indicating participants were devoting their attention to the virtual environment [23, 28, 29]. However, spatial presence was rated average mainly during the underground videos, due to the shaking triggered by train movements while recording, whereas participants were standing on a steady floor.

4.2.2 Interviews. After the Presence Questionnaire, we conducted a closing semi-structured interview with each participant. Here, we revealed the purpose of the study. We started by defining smartphone shoulder surfing and thereafter, asked questions regarding their behavior during the study and previous experiences in real-life situations.

Study Behavior. When asked if they were shoulder surfing the user in the videos, two of the 15 participants claimed that they did not shoulder surf. However, all the participants confirmed noticing at least one smartphone in each of the public transportation videos. In the following, some examples of statements from participants during the study.

> *P1: "I could see what he was doing on his mobile". (M, 32y)*
> *P3: "I can see the display of the guy next to me, not readable but I have a very good idea about what happens". (M, 32y)*
> *P9: "I noticed that the guy [..] always looking at pictures, and writing". (M, 26y)*
> *P15: "I could sneak on the phone of the person next to me". (M, 31y)*
> *P16: "He was playing a flag game, was really easy! [..], also looking at photos [of a girl] I think she is in the woods [..] sure he unlocked his phone and texted via Whatsapp." (F, 24y)*

Authentication Pattern. Participants were asked if they have observed the phone while unlocking and would be able to redraw the authentication pattern, five participants were able to correctly redraw the pattern, four of which with 100% accuracy. One participant guessed the pattern with 87.5% accuracy, missing the last dot connection. Participants who observed the pattern looked at the smartphone on average for 6.52s (IQR = 4.59s), while participants who did not observe the pattern looked on average for 4.84s (IQR = 5.27s). A pairwise comparison using Wilcoxon Signed-rank did not reveal a significant difference between the groups ($W = 1$, $Z = 1.46$, $p = .250$).

Applications. During the videos presented to the participant, the smartphone user used four different applications (WhatsApp, Facebook, Gallery, and Game). During the post-study interview, we asked participants if they observed specific apps. Overall, 14 out of 16 participants (87.5%) could name at least one application. One participant could name all applications. On average, each participant observed 2.06 apps ($IQR = 1.00$): WhatsApp=12 (75.0%), Facebook=3 (18.8%), Gallery=11 (75.0%), and Game=6 (37.5%). Three participants (p8, p11, p16) provided more details, describing the pictures visible in the Gallery (p11, p16) and details of the Game (p11) or WhatsApp (p8).

For the applications, we were interested if there is a correlation between observed applications and count or duration of phone gazes. Since we do not assume normal distribution of our data, we applied the Spearman's rank correlation coefficient. We neither found a statistically significant correlation for count of phone gazes ($r(14) = 0.115$, $p = 0.670$) nor did we find a statistically significant correlation for duration of phone gazes ($r(14) = 0.359$, $p = 0.172$). The duration of phones gazes in relation to the recognized application is shown in Figure 4.

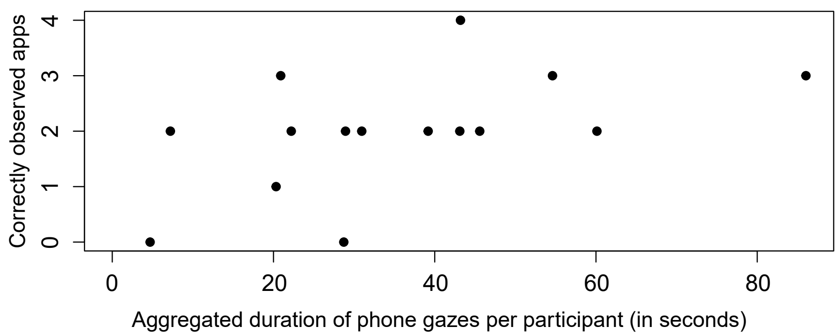


Fig. 4. Duration of phone gazes per participant in relation to the correctly observed apps.

Real-life Behavior. The last questions in the interview were about the actual behavior in real-life situations, all participants admitted being an observer at least once. However, and complying with the survey [6], they were mostly opportunistic and out of boredom. Four subjects said they are not aware of being observed before.

5 DISCUSSION

Gaze Behavior. Overall, we found that all participants glanced at the mobile phone at least once per video. On average, they spent about 11.16% of their time looking at the user's phone. While this number is high, it can be explained by the fact that shoulder surfing is often justified with boredom or curiosity [6]. Given the fact that participants could only passively perceive the video, we assume that boredom might be an influencing factor.

We found that the user's gaze behavior changes based on the position of the user and observer. The sitting observer and the standing user results in most shoulder surfing. This is not in line with the results of Eiband et al. who found that standing observers are more common [6]. In the interviews, participants mentioned that in the sitting position, they were less distracted by the motion of the train and, thus, more observing.

Shoulder Surfing vs. Shoulder Surfing Attack. In this work, we recorded the gaze of the observer. We identified every fixation on the phone and, thus, identify shoulder surfing incidents. The main challenge, however, is to understand the motivation of the user and whether private information of the user are perceived by the observer. We, therefore, classify these incidents as shoulder surfing and not shoulder surfing attacks. On the other hand, four of the observers knew the login pattern of the user afterwards, indicating that this was indeed a shoulder surfing attack. The results indicate that the time they looked at the phone might be an indicator of an attack, since they on average looked longer on the phone.

Observed Content. In the interviews, we found that participants could remember the applications used on the mobile phone. Besides the lock pattern, this underlines that bystanders could also be interested in pictures and other content shown on the phone. In fact, in the interviews, participants could actually recall details of the pictures shown in the gallery application (e.g., girl in the woods). This shows that besides the fear of being shoulder surfed while authentication, the regularly used content is shoulder surfed as well, posing a potential threat to the users privacy. Additionally, we could not find a significant correlation between the gaze duration logged in the user study and the correctly observed apps. This means that also short glances might be enough to disclose content and, subsequently, violate one's privacy.

Feasibility of Methodology. In this work, we used a new methodology to investigate shoulder surfing in the lab. In contrast to prior work that surveyed users about their prior incidents [6], we aim at understanding shoulder surfing in-situ. We used 360° videos of a public transport situation in which we analyzed the user's gaze behavior. In the study, this methodology achieved a high presence score, indicating that participants were immersed in the video, despite the content readability limitations. In contrast, real-world observations might not provide gaze data or violate the privacy of users. While the artificial study environment - despite deceiving users - might have changed the gaze behavior of the user, this methodology is still a step forward to understanding the gaze behavior of the shoulder surfer. However, throughout designing the user study, we learned that designing a realistic environment is key to this methodology. For instance, using a fixed camera stand is important to create non-shaking videos. In pre-tests, we found that slight shaking already induces simulator sickness to the participants. Although we improved the setup through a tripod, the movement of the underground still induced mild simulator sickness. Similarly, the height of the camera needs to be adjusted to be within a range of 10cm [24] to create a realistic impression and not induce further sickness. Last, since the 360° video consists of multiple concurrent videos stitched together, the quality of stitching of the different videos highly affects the experience of the participants.

Limitations. We acknowledge the following limitations to our work. Firstly, the readability of the smartphone textual content is limited by the resolution of the used camera and head-mounted display. Secondly, our detailed simulation of a real-life settings in VR cannot represent all aspects of a real environment. However, previous work confirmed comparable behavior of users in both environments. As a result, findings need to be treated carefully, yet we believe they can still uncover interesting aspects that are promising for future research that would not have been available otherwise. Third, We acknowledge that we used four distinct applications as previously identified in related work that require four different hand gestures as input. We investigated both together since specific applications typically require dedicated input. However, we did not investigate if the application or hand gesture is eliciting more attention of the shoulder surfer. For example, users might expect more interesting content when they see the user swiping (e.g., image gallery) than typing.

Future Work. For future work, we extend our work to gain further understanding on the phone applications impact upon shoulder surfing behavior. We are considering other applications, different visual and audio content, other authentication methods, and other user-phone interactions to closely simulate realistic scenarios. Moreover, we extend the research to examine the effect of multiple smartphones upon the adversary's gaze behavior. We also consider the impact of readable content upon the observation and attention.

6 CONCLUSION

In this work, we present our methodology of combining 360° videos with eye-tracking to investigate observative attacks. In the recorded videos, we considered actual scenarios with reported high likelihood of attacks, in addition to possible postures combinations. Our investigations mainly focused on participants' gaze count and duration, using eye tracking technology embedded in the Head-Mounted Display.

With regards to **R1**, our findings show that all participants have looked at the smartphone of interest at least once in each of the videos, independent of the shown content. They could also perceive the nature of the interaction (e.g., typing vs. swiping). Four participants were in addition able to recreate the authentication pattern and name the apps used on the phone, showing that authentication schemata also spark interest of shoulder surfers. With regards to **R2**, we found that the sitting observer and the sitting user results in most shoulder surfing. Our work contributes to the body of knowledge on shoulder surfing, gaining additional insights using a new study methodology. We shed light on how users behave in a virtual scenario, which brings us one step closer to the goal of understanding shoulder surfing.

ACKNOWLEDGMENTS

This project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 425869382 and 426052422 and is part of Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments.

REFERENCES

[1] Jennifer Brade, Mario Lorenz, Marc Busch, Niels Hammer, Manfred Tscheligi, and Philipp Klimant. 2017. Being there again - Presence in real and virtual environments and its relation to usability and user experience using a mobile navigation task. *International Journal of Human-Computer Studies* 101 (2017), 76–87. https://doi.org/10.1016/j.ijhcs.2017.01.004

[2] Marc Busch, Mario Lorenz, Manfred Tscheligi, Christina Hochleitner, and Trenton Schulz. 2014. Being There for Real: Presence in Real and Virtual Environments and Its Relation to Usability. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (Helsinki, Finland) *(NordiCHI '14)*. Association for Computing Machinery, New York, NY, USA, 117–126. https://doi.org/10.1145/2639189.2639224

[3] Shudi Chen and Youwen Zhu. 2019. A Textual Password Entry Method Resistant to Human Shoulder-Surfing Attack. In *International Symposium on Cyberspace Safety and Security*. Springer, 409–420.

[4] Alexander De Luca, Emanuel Von Zezschwitz, Ngo Dieu Huong Nguyen, Max-Emanuel Maurer, Elisa Rubegni, Marcello Paolo Scipioni, and Marc Langheinrich. 2013. Back-of-device authentication on smartphones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2389–2398.

[5] Shuchisnigdha Deb, Daniel W Carruth, Richard Sween, Lesley Strawderman, and Teena M Garrison. 2017. Efficacy of virtual reality in pedestrian safety research. *Applied ergonomics* 65 (2017), 449–460.

[6] Malin Eiband, Mohamed Khamis, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. 2017. Understanding Shoulder Surfing in the Wild: Stories from Users and Observers. In *Proceedings of the 35th Annual ACM Conference on Human Factors in Computing Systems* (Denver, CO, USA) *(CHI '17)*. ACM, New York, NY, USA. https://doi.org/10.1145/3025453.3025636

[7] Marian Harbach, Emanuel Von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. 2014. It's a hard lock life: A field study of smartphone (un)locking behavior and risk perception. In *10th Symposium On Usable Privacy and Security ({SOUPS} 2014)*. 213–230.

[8] Marcel Adam Just and Patricia A Carpenter. 1976. Eye fixations and cognitive processes. *Cognitive psychology* 8, 4 (1976), 441–480.

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

[10] Mohamed Khamis, Regina Hasholzner, Andreas Bulling, and Florian Alt. 2017. GTmoPass: two-factor authentication on public displays using gaze-touch passwords and personal mobile devices. In *Proceedings of the 6th ACM International Symposium on Pervasive Displays*. 1–9.

[11] Mohamed Khamis, Mariam Hassib, Emanuel von Zezschwitz, Andreas Bulling, and Florian Alt. 2017. GazeTouchPIN: protecting sensitive data on mobile devices using secure multimodal authentication. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. 446–450.

[12] Mohamed Khamis, Ludwig Trotter, Ville Mäkelä, Emanuel von Zezschwitz, Jens Le, Andreas Bulling, and Florian Alt. 2018. Cueauth: Comparing touch, mid-air gestures, and gaze for cue-based authentication on situated displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–22.

[13] Hassan Khan, Urs Hengartner, and Daniel Vogel. 2018. Evaluating attack and defense strategies for smartphone pin shoulder surfing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–10.

[14] Katharina Krombholz, Thomas Hupperich, and Thorsten Holz. 2016. Use the force: Evaluating force-sensitive authentication for mobile devices. In *Twelfth Symposium on Usable Privacy and Security ({SOUPS} 2016)*. 207–219.

[15] K. Krombholz, T. Hupperich, and T. Holz. 2017. May the Force Be with You: The Future of Force-Sensitive Authentication. *IEEE Internet Computing* 21, 3 (May 2017), 64–69. https://doi.org/10.1109/MIC.2017.78

[16] Otto Hans-Martin Lutz, Charlotte Burmeister, Luara Ferreira dos Santos, Nadine Morkisch, Christian Dohle, and Jörg Krüger. 2017. Application of head-mounted devices with eye-tracking in virtual reality therapy. *Current Directions in Biomedical Engineering* 3, 1 (2017), 53–56.

[17] Ville Mäkelä, Rivu Radiah, Saleh Alsherif, Mohamed Khamis, Chong Xiao, Lisa Borchert, Albrecht Schmidt, and Florian Alt. 2020. Virtual Field Studies: Conducting Studies on Public Displays in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[18] Florian Mathis, Kami Vaniea, and Mohamed Khamis. 2021. RepliCueAuth: Validating the Use of a lab-based Virtual Reality Setup for Evaluating Authentication Systems. In *Proceedings of the 39th Annual ACM Conference on Human Factors in Computing Systems (Yokohama, Japan)(CHI'21). ACM, New York, NY, USA*.

[19] Mehdi Moussaïd, Mubbasir Kapadia, Tyler Thrash, Robert W Sumner, Markus Gross, Dirk Helbing, and Christoph Hölscher. 2016. Crowd behaviour during high-stress evacuations in an immersive virtual environment. *Journal of The Royal Society Interface* 13, 122 (2016), 20160414.

[20] Ildar Muslukhov, Yazan Boshmaf, Cynthia Kuo, Jonathan Lester, and Konstantin Beznosov. 2013. Know your enemy: the risk of unauthorized access in smartphones by insiders. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*. 271–280.

[21] Cagri Ozcinar and Aljosa Smolic. 2018. Visual attention in omnidirectional video for virtual reality applications. In *2018 Tenth international conference on quality of multimedia experience (QoMEX)*. IEEE, 1–6.

[22] Athanasios Papadopoulos, Toan Nguyen, Emre Durmus, and Nasir Memon. 2017. Illusionpin: Shoulder-surfing resistant authentication using hybrid images. *IEEE Transactions on Information Forensics and Security* 12, 12 (2017), 2875–2889.

[23] Holger Regenbrecht and Thomas Schubert. 2002. Real and illusory interactions enhance presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 11, 4 (2002), 425–434.

[24] Sylvia Rothe, Boris Kegeles, and Heinrich Hussmann. 2019. Camera heights in cinematic virtual reality: How viewers perceive mismatches between camera and eye height. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*. 25–34.

[25] Alia Saad, Michael Chukwu, and Stefan Schneegass. 2018. Communicating Shoulder Surfing Attacks to Users. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*. 147–152.

[26] Helmut Schrom-Feiertag, Christoph Schinko, Volker Settgast, and Stefan Seer. 2014. Evaluation of Guidance Systems in Public Infrastructures Using Eye Tracking in an Immersive Virtual Environment. In *ET4S@GIScience (CEUR Workshop Proceedings)*, Vol. 1241. CEUR-WS.org, 62–66.

[27] Helmut Schrom-Feiertag, Volker Settgast, and Stefan Seer. 2017. Evaluation of indoor guidance systems using eye tracking in an immersive virtual environment. *Spatial Cognition & Computation* 17, 1-2 (2017), 163–183.

[28] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments* 10, 3 (2001), 266–281.

[29] Thomas W Schubert. 2003. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Z. für Medienpsychologie* 15, 2 (2003), 69–71.

[30] Mel Slater, Pankaj Khanna, Jesper Mortensen, and Insu Yu. 2009. Visual realism enhances realistic response in an immersive virtual environment. *IEEE computer graphics and applications* 29, 3 (2009), 76–84.

[31] Evgeniy Upenik and Touradj Ebrahimi. 2017. A simple method to obtain visual attention data in head mounted virtual reality. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 73–78.

[32] Alexandra Voit, Sven Mayer, Valentin Schwind, and Niels Henze. 2019. Online, VR, AR, Lab, and In-Situ: Comparison of Research Methods to Evaluate Smart Artifacts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[33] Emanuel Von Zezschwitz, Alexander De Luca, Bruno Brunkow, and Heinrich Hussmann. 2015. Swipin: Fast and secure pin-entry on smartphones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1403–1406.

[34] Oliver Wiese and Volker Roth. 2015. Pitfalls of shoulder surfing studies. In *NDSS Workshop on Usable Security*. 1–6.

[35] Abid Yaqoob, Ting Bi, and Gabriel-Miro Muntean. 2020. A Survey on Adaptive 360° Video Streaming: Solutions, Challenges and Opportunities. *IEEE Communications Surveys & Tutorials* 22, 4 (2020), 2801–2838.

[36] Insu Yu, Jesper Mortensen, Pankaj Khanna, Bernhard Spanlang, and Mel Slater. 2012. Visual realism enhances realistic response in an immersive virtual environment-part 2. *IEEE Computer Graphics and Applications* 32, 6 (2012), 36–45.

[37] Junyi Zhang and Harry Timmermans. 2010. Scobit-based panel analysis of multitasking behavior of public transport users. *Transportation research record* 2157, 1 (2010), 46–53.

[38] Hongzi Zhu, Jingmei Hu, Shan Chang, and Li Lu. 2017. ShakeIn: secure user authentication of smartphones with single-handed shakes. *IEEE transactions on mobile computing* 16, 10 (2017), 2901–2912.