Exploring Redirection and Shifting Techniques to Mask Hand Movements from Shoulder-Surfing Attacks during PIN Authentication in Virtual Reality

YANNICK WEISS, LMU Munich, Germany
STEEVEN VILLA, LMU Munich, Germany
JESSE W. GROOTJEN, LMU Munich and Munich Center for Machine Learning (MCML), Germany
MATTHIAS HOPPE, Keio University Graduate School of Media Design, Japan
YASIN KALE, LMU Munich, Germany
FLORIAN MÜLLER, LMU Munich, Germany

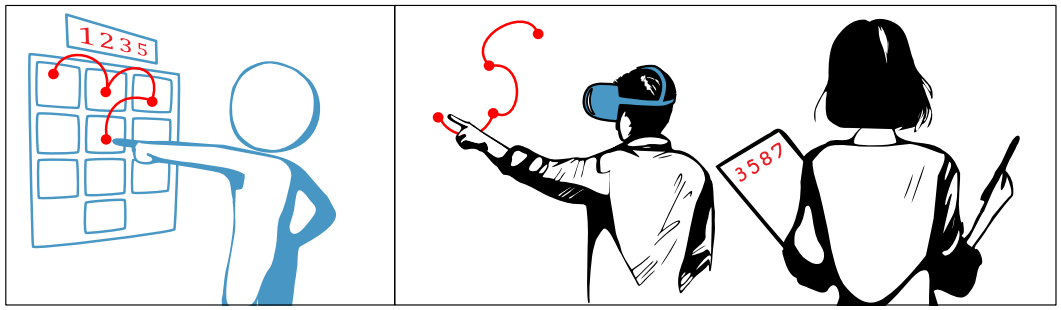


Fig. 1. Mobile headsets prompt the need for robust security measures to safeguard our personal data. Conventional PIN entry methods for Virtual Reality via virtual keypads are vulnerable to shoulder-surfing attacks. To address this, we employ techniques to conceal the hand movements of users during PIN authentication in VR through hand redirection or shifting of the virtual keypad.

The proliferation of mobile Virtual Reality (VR) headsets shifts our interaction with virtual worlds beyond our living rooms into shared spaces. Consequently, we are entrusting more and more personal data to these devices, calling for strong security measures and authentication. However, the standard authentication method of such devices - entering PINs via virtual keyboards - is vulnerable to shoulder-surfing, as movements to enter keys can be monitored by an unnoticed observer. To address this, we evaluated masking techniques to obscure VR users' input during PIN authentication by diverting their hand movements. Through two experimental studies, we demonstrate that these methods increase users' security against shoulder-surfing attacks from observers without excessively impacting their experience and performance. With these discoveries, we aim to

Authors' Contact Information: Yannick Weiss, LMU Munich, Munich, Germany, yannick.weiss@ifi.lmu.de; Steeven Villa, LMU Munich, Munich, Germany, villa@posthci.com; Jesse W. Grootjen, LMU Munich and Munich Center for Machine Learning (MCML), Munich, Germany, jesse.grootjen@ifi.lmu.de; Matthias Hoppe, Keio University Graduate School of Media Design, Tokyo, Japan; Yasin Kale, LMU Munich, Munich, Germany; Florian Müller, LMU Munich, Munich, Germany, florian.mueller@ifi.lmu.de.

ACM 2573-0142/2024/9-ART257
https://doi.org/10.1145/3676502

enhance the security of future VR authentication without disrupting the virtual experience or necessitating additional hardware or training of users.

CCS Concepts: • **Human-centered computing** → **Virtual reality**; **Interaction techniques**.

Additional Key Words and Phrases: hand redirection, shoulder-surfing, virtual reality

ACM Reference Format:
Yannick Weiss, Steeven Villa, Jesse W. Grootjen, Matthias Hoppe, Yasin Kale, and Florian Müller. 2024. Exploring Redirection and Shifting Techniques to Mask Hand Movements from Shoulder-Surfing Attacks during PIN Authentication in Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 8, MHCI, Article 257 (September 2024), 24 pages. https://doi.org/10.1145/3676502

1 Introduction

With the proliferation of mobile Virtual Reality (VR) and Augmented Reality (AR) headsets[1,2], interaction with such devices increasingly moves out of our living rooms and into more public and shared spaces. This shift into the public sphere coincides with a rapid increase in VR content and the push of large social platforms like the Metaverse[3], causing more security-critical interactions, such as sharing personal information or making payments, to occur in the virtual environment. Consequently, the need for secure and usable authentication systems for VR is increasing.

Traditional authentication methods, such as entering passwords via physical keyboards or touchscreens, are not feasible with mobile VR headsets. As a result, various proposed approaches are tailored to secure authentication with such devices. Research shows adapting established methods like PIN and pattern authentication to mid-air interactions in VR to be very effective and easy to use [11], and has thus remained the state-of-the-art approach for authentication in such systems. However, these techniques suffer from shoulder surfing attacks [11, 28, 39], where observers may try to gain access to the systems by guessing the user's input based on their hand movements. With the more prevalent use of VR in public or surrounded by people, the threat of shoulder surfing increases, as users are isolated visually (and oftentimes auditory) and thus cannot perceive the observers who are watching them. To combat this, behavioral biometric techniques instead identify users through their gait [33], throwing technique [19, 25], gaze behavior [44] or other movements of the body [29]. However, these systems demand the collection of even more personal data [20], require users to do specific tasks [19], or introduce novel risks like the possibility of attackers mimicking users to gain access [27]. Alternatively, research proposed button- or gesture-based approaches using tracked VR controllers [1] or novel devices [24]. While these can provide increased protection against shoulder surfing, they are generally slower and more cumbersome than a PIN or pattern. Controllers can also constrain the user and might become obsolete with the proliferation of free-hand interactions in mobile headsets. Consequently, research has brought forth many promising techniques for authentication in VR without the risk of shoulder surfing. However, they either mandate exposing more private data or require more attention and the learning of novel input methods, thus impacting the authentication system's usability and the user's experience.

To address these issues, we propose and evaluate authentication techniques that disguise user input during PIN authentication in VR to protect against shoulder surfing attacks without diminishing the user's experience. We built on the convenient and well-established approach of free hand PIN entry on a 10-digit virtual keypad. We extend this state-of-the-art approach and take advantage of the fact that users cannot see their real hands in VR and have to rely on visual

[1] https://www.meta.com/de/quest/quest-pro/, last accessed: 2023-09-12
[2] https://www.apple.com/apple-vision-pro/, last accessed: 2023-09-12
[3] https://about.meta.com/metaverse/, last accessed: 2023-09-12

representations that we are able to manipulate. Previous work has shown that a deliberate shift in the visual representation of users' hands does not break users' perception of body ownership but adapts their proprioception – the knowledge about the positions and movements of the limbs – to the visual presentation [12, 30, 42]. Based on these findings, we continuously divert the visual trajectory of the hands while reaching out and thus unconsciously redirect their hand movements when entering a PIN. In addition to this approach, which warps the hands of the user continuously during the interaction, we also evaluate a discrete approach, which instead shifts the entire virtual keypad before every key press. Employing these redirection and shifting techniques during PIN authentication in VR causes the hand movements of users to be obscured from outside observers trying to gain access through shoulder surfing attacks.

1.1 Contribution Statement

In this paper, we contribute (1) four masking techniques concealing the users' hand movements during PIN authentication in VR to protect against shoulder surfing attacks (Section 3), (2) two user studies evaluating the effect of these methods on performance and subjective experience in VR (Section 4) and their protection from shoulder surfing attacks of outside observers (Section 5). Lastly, (3) we derive recommendations for future PIN authentication in VR based on the collected insights (Section 6).

2 Related Work

Our work builds on a wealth of related literature. We structured these works based on the topics of shoulder surfing (2.1), the integration of knowledge-based and biometric authentication methods to combat it (2.2), and the use of hand redirection techniques in VR (2.3).

2.1 Shoulder Surfing in VR

The increased use of VR to share private information or make payments requires usable and secure authentication techniques. At the same time, VR poses additional challenges and opportunities to authentication methods. Well-established authentication techniques using 2D PINs or swipe patterns on a keypad can be transferred well to VR with their usability matching their performance known from the physical world [11] and outperforming three-dimensional approaches [39]. While these knowledge-based approaches remain the current state-of-the-art for authentication, the input of PINs or patterns faces the challenge of shoulder-surfing attacks. Shoulder-surfing attacks refer to malicious outsiders gaining access to someone's system by observing their inputs or behavior. This is not a new phenomenon and has been extensively researched in the context of PIN inputs on smartphones [36, 40]. However, in VR, this problem is more pronounced due to the larger movements being made in 3D space as compared to on a smartphone screen and the fact that the VR user is not able to see whether they are being observed. George et al. [11] showed that attackers were able to guess 18% of users' PINs or patterns by immediate observation, and Yu et al. [39] reported that most people were able to guess the inputs of two-dimensional PINs and patterns on a virtual keypad by watching back video recordings. Olade et al. [28] tested the vulnerability of swipe patterns in VR to shoulder-surfing attacks by showing recordings of three viewpoints and reported a success rate of 20% for outside attackers unfamiliar with the system. This highlights the necessity of increasing the security of authentication in VR to protect against common attacks.

2.2 Knowledge-based and Biometric Authentication

To combat shoulder-surfing attacks for VR systems, research proposes numerous novel knowledge-based and behavioral biometrics systems. On the spectrum of knowledge-based authentication methods, we can find techniques using three-dimensional patterns [39], gaze to put in passwords [16,

18], copying randomly changing gestures [1], or pointing at different virtual or real objects in a room [10]. Mathis et al. [24] instead employed a virtual cube with keypads on each side held in one hand, while the other hand, head position, or gaze can be used to select numbers on the different sides. These techniques offer more security to shoulder-surfing due to their increased complexity of inputs [10, 24, 39], randomized aspects [1], or concealed interactions [16, 18]. However, compared to the simple and pervasive gesture of using your fingers to type in a code, the randomization and increase in complexity makes the authentication process more demanding [1] and slower [1, 10, 18, 39] without training.

The use of behavioral biometrics promises more implicit and, therefore, less challenging authentication while remaining secure against shoulder-surfing attacks. Research investigated the use of electroencephalography (EEG) [17, 20] and eye-tracking [3, 17, 21, 43, 44] for authentication or employed deep learning models classifying the identity of users based on their head movement [26, 34] or the behavior of the whole body [19, 23, 25–27, 29, 33] during specific tasks. These include walking [33], throwing a ball [19, 25], or blinking rhythmically [44]. However, behavioral biometric approaches demand the collection and storing of additional personal immutable information and often require the implementation of supplementary hardware like EEGs [20] or eye trackers [44]. Furthermore, they carry novel risks, such as susceptibility to attackers mimicking the victim's behavior, especially if they are of similar build and height [27].

2.3 Hand Redirection Techniques

Previous research already extensively explored the use of hand redirection in the field of human-computer interaction (HCI) to enhance haptic experiences in VR. Hand redirection relies on the dominance of visual perception over proprioception in judging the position of our limbs in space. One of the first reports of this phenomenon was in the Rubber Hand Illusion (RHI) [7]. In this experiment, the investigators hide one of the participant's hands and place a rubber hand in front of them instead. Simultaneous and congruent stroking of the fake and the hidden hand with a brush causes the foreign limb to be adopted into the participant's body representation. In addition to the tactile sensations applied to the real hidden hand, falsely being perceived as originating from the rubber hand, the RHI also causes a proprioceptive drift of the participant's perceived hand position in space towards the fake rubber hand. In VR, the same phenomenon can be elicited by rendering a virtual representation of the hand. The ability to actively track and mimic the motion of the users' real hands causes a strong integration into their body schema [22, 32] and deliberate visual displacements alter their sense of proprioception instead of breaking it [30]. Various works have since been investigating the boundaries of detecting these hand redirection in various contexts [9, 12, 37, 42]. Research further explored how this phenomenon can extend the haptic capabilities of systems by having a single physical object stand in for multiple virtual objects and redirecting the user's touch back to the same passive prop [4] or a specific section of it [8]. Furthermore, Samad et al. [31] and Weiss et al. [37] respectively showed that deliberate hand redirections during lifting or pressing alter the perceived weight and stiffness of objects in VR. With active devices, hand redirection has been used to lower encounter-type displays' speed and workspace requirements by diverting users' hands closer to the device's touchable area [13] or to increase the perceived resolution, speed, and size of shape displays by adjusting the redirection while users actively explore the device [2]. However, using hand redirection to conceal hand movements from outside observers during authentication in VR has not been investigated.

3 Methodology

In this work, we propose novel authentication techniques for PIN Authentication in VR, intending to protect user inputs from potential shoulder surfing attacks by masking the user's hand movement.

We deliberately break the one-to-one mapping of the virtual and real movement of users' hands. In line with prior work on hand redirection [4, 12, 42], we adjust the trajectory of the virtual hand representation to drift slightly while users reach out, resulting in users unconsciously adjusting their real hand trajectories to offset the discrepancy. Instead of redirecting users to a physical object, we use this redirection to adjust users' hand movements to target another spot in mid-air randomly offset from the virtual key they originally wanted to press. Thus, while users are putting in their selected passwords on a virtual keypad in front of them, for an outside observer, it appears as if the user is pressing entirely different buttons (see Figure 1).

3.1 Proposed Techniques

We employ three different functions to control how the drift gain is distributed over the hand's entire trajectory. First, we explore a Linear gain function ($f_1(x) = x$) for hand redirection, which is the most commonly used throughout prior work [4, 42]. Geslain et al. [12] discuss that the trajectory of the users' hands during hand redirection can be separated into a ballistic phase at the beginning and a correction phase towards the end of the movement. During the ballistic phase, users have not yet adapted to the redirection and remain on their trajectory to the original target. In the correction phase, users adjust – usually subconsciously – their movements to correct for the offset gained by the hand redirection. While Geslain et al. found no effect of different amounts of gain in ballistic and correction phases on detection thresholds, they stipulate that this may affect visual and proprioceptive accuracy. Consequently, we employed two additional functions: An exponential function (Ease-In) that applies less gain at the beginning of the hand movement and increases while the hand moves closer toward the target ($f_2(x) = x^2$) and an opposing function (Ease-Out) that applies more drift early and decreases the gain over time ($f_3(x) = 1-(1-x)^2$). We selected the Ease-In and Ease-Out functions to evaluate the impact of larger redirection gains in either the ballistic or correction phase on users' performance, experience, and security. In addition to these continuous approaches, we propose a discrete method (Shift), which instead moves the virtual keypad before each key input by the same degree. An overview of the proposed techniques – Linear, Ease-In, Ease-Out, and Shift – as well as the current state-of-the-art without masking techniques (Baseline) can be seen in Figure 2.

These techniques add an additional layer of security to the password entry process, possibly making it more difficult for potential attackers to decipher the password by observation.

3.2 Threat model

We address the threat model of shoulder surfing attacks on a VR user authenticating with PINs in a shared space. In our model, the victim wears a head-mounted display (HMD) and interacts with an immersive virtual environment in public or an otherwise shared space (e.g., at home with roommates or at work with colleagues). The VR user authenticates their account to log in or make a purchase by typing in a PIN with their free hands on a virtual keypad. Because they cannot see the real world around them, they cannot notice a bystander or pre-installed device observing or recording their authentication process. The attacker may then immediately guess the code put in by the VR user or watch the recording back later.

3.3 Research Questions

In this work, we aim to answer the following research questions (RQ):

- **RQ1** – How do techniques masking hand movements of PIN Authentication in VR affect the user's performance and experience?

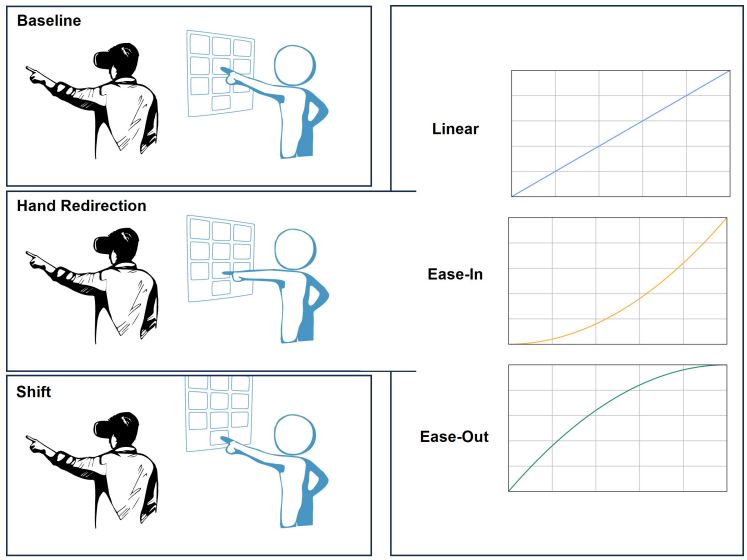


Fig. 2. The techniques proposed to conceal hand movements for PIN authentication in VR. **Baseline** involves no masking of movement. We employ three different hand redirection techniques, using a **Linear** ($f_1(x) = x$), **Ease-In** ($f_2(x) = x^2$) and **Ease-Out** ($f_3(x) = 1 - (1 - x)^2$) function for redirection gains. These functions describe the offset gain as the user's hand moves closer to the keypad plane. Additionally, we propose a **Shift** technique, in which the entire virtual keypad moves before each key input. The total redirection or shifting offset is determined based on the keypad size (see Section 4.1).

- **RQ2** – How do techniques masking hand movements of PIN Authentication in VR affect the success of shoulder-surfing attacks from observers?

To address these research questions, we conducted two controlled studies: First, we implemented the proposed techniques in a VR environment and measured the performances and reported experiences of 30 participants while recording their trials. Then, 50 new participants watched the recordings in an online survey while trying to guess the PINs that were entered by the VR users.

4 Study I: Effect on performance and subjective experience

We conducted a controlled experiment to evaluate the impact of different masking techniques and keyboard sizes on users' performance and experience during PIN authentication in VR (*RQ1*). We tasked participants to put in 4-digit PINs on virtual keypads hovering mid-air in the virtual environment while we applied the proposed masking techniques.

4.1 Study Design

We used a within-subject design with two independent variables, varying the levels of the masking technique used (Technique) and the size of the virtual keypad being displayed (Size).

- **Technique**: We evaluate the four masking techniques proposed in Section 3.1 and a baseline without masking. Thus, our first independent variable Technique comprises **Baseline**, **Linear**, **Ease-In**, **Ease-Out**, and **Shift**.
- **Size**: For our second independent variable, we adjust the size of the virtual keypad. We selected three different virtual keypad sizes: **Small** ($\sim 5cm \times 6.7cm$), **Medium** ($\sim 10cm \times 13.3cm$), and **Large** ($\sim 20cm \times 26.7cm$).

The direction of redirection and shifting for the masking techniques is randomly selected before each individual key input. For the selection of the sizes, we considered that the amount of redirection and shifting of the masking techniques needs to always cover at least the distance between two keys to ensure the displacement is large enough to mask the actual key. So, even if a participant tries to press on the top right corner of a key and the redirection is randomly directing them toward the bottom left, the redirected position of their real hand will not be on the same key. Following this requirement, we derived these keypad sizes, which result in necessary redirection magnitudes of 4° (Small), 8° (Medium), and 16° (Large) when starting the grasp from 30cm in front of the keypad. Based on the investigations of Zenner and Krüger [42] – which reported the lowest threshold of the works we found – the redirection applied during interaction with the Small keypad should, therefore, remain below the threshold of 4.5° for detection of participants, while the redirection during the Medium and Large conditions might be perceived by participants more often. However, all three should not be disruptive to the experience based on the findings of Cheng et al. [8].

We evaluated each combination of Technique and Size, which results in a total of 15 conditions (5 Techniques × 3 Sizes). Each condition consisted of eight randomly selected PINs participants had to put in. In line with prior work [11], we used 4-digit PINs. Consequently, each participant put in a total of 60 PINs (5 Techniques × 3 Sizes × 8 PINs). The order of conditions was counterbalanced among participants using a Balanced Latin Square approach with 30 rows.

For each condition, we measured the error rate (number of incorrectly entered keys) and task completion time (time between first displaying the keypad and the 4th digit entered) of inputs by tracking participants' hand movements, and we asked participants to indicate the perceived task load of the condition using the NASA Task Load Index (NASA-TLX) questionnaire [15]. Furthermore, to assess their subjective sense of security, we asked participants: *In a shared space, how concerned would you be that bystanders can guess your input?* Participants rated their concern on a scale from *Very Low* (0) to *Very High* (20).

4.2 Apparatus

We created a virtual environment in Unity3D and deployed it to a desktop computer running Windows 10, equipped with an Intel i7 processor, 16GB RAM, and an NVIDIA GeForce RTX 2070 Super graphics card. The VR scene is rendered on an HTC Vive Pro head-mounted display (HMD). We attached a Leap Motion Controller[4] to the front of the HMD to track the participants' hand and finger movements in real time at 120Hz. The participant's hands are represented by low-poly hand models provided by Ultraleap. The participants were seated on a chair while wearing the HMD. We display a 10-digit numeric keypad placed 50cm in front of participants in the virtual environment that participants can interact with by pushing the buttons using their dominant hand. Hand redirection was implemented by extending the hand redirection toolkit (HaRT) by Zenner et al. [41] with our custom gain functions described in Section 3.1. Additionally, we mounted a webcam[5] with 1080p and 30fps video output on a tripod in front of the participants. The webcam recorded the users' hand movements while inputting the PINs. The recordings were later used to analyze if observers watching them would be able to guess the entered PIN. The full study setup can be seen in Figure 3.

4.3 Procedure

After welcoming the participants, we explained the general objectives of the study and the procedures for data collection and processing. We then asked participants to sign a consent form and to

[4] https://www.ultraleap.com/datasheets/Leap_Motion_Controller_Datasheet.pdf, last accessed: 2024-05-20
[5] https://www.microsoft.com/en/accessories/products/webcams/microsoft-modern-webcam, last accessed: 2023-09-12

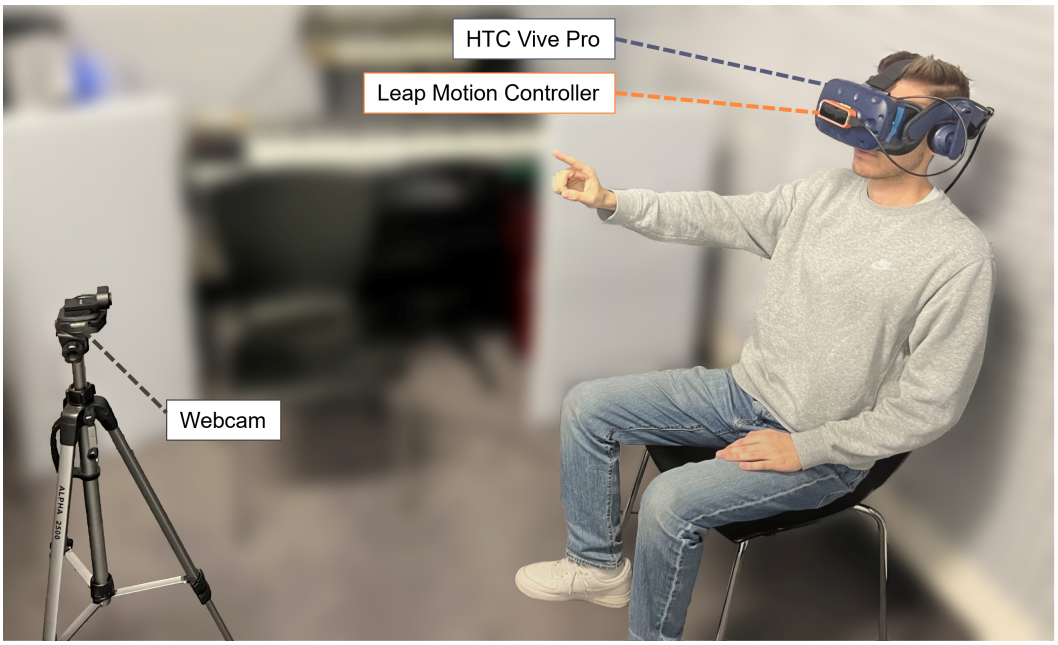
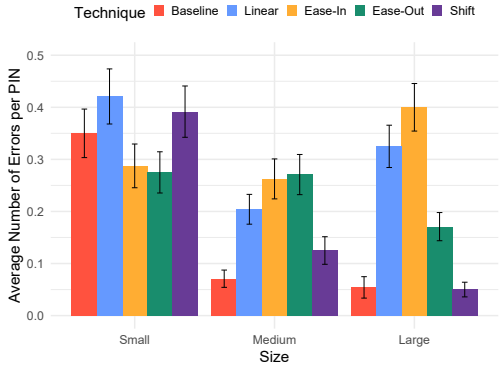


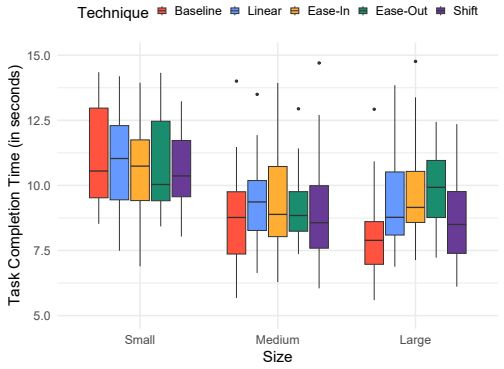
Fig. 3. The experimental setup consists of an HTC Vive Pro HMD with an attached Leap Motion Controller for hand tracking, a desktop computer running the virtual environment, and a webcam placed in front of participants to record the users during their PIN authentication.

fill out a demographic questionnaire. Afterward, participants sat down and put on the VR HMD. We asked each participant to hold their dominant hand straight in front of them, and we adjusted the virtual environment so that their hand marked the horizontal and vertical center of the virtual keypad's plane. This was done to reduce the possible confounding effect of participants' heights and to ensure a comfortable interaction with the system. The distance of the keypad was then set to 50cm in front of the chair's backrest the participants were seated in. Following the calibration, the participants started a training session consisting of eight PINs on the medium-sized keypad without redirection. Following this, the test trials were started.

We asked participants to input eight PINs for each condition, with a 10-second break after every four PINs. Each PIN required the input of four digits, which were displayed above the keypad. The next digit they had to put in was always visually highlighted. Participants had to start with their hands close to their bodies, reach out with their dominant hand, and press the buttons on the virtual keypad. Auditory feedback signaled a successful button press, but no indication about whether the participant hit the right key was given. The participant then had to retract their hand fully before reaching out to put in the next digit of the PIN. This was done to ensure the redirection techniques worked properly and to mitigate possible confounding effects of sequential keys being closer or farther apart from each other. The PINs they had to put in were generated randomly. Depending on the condition, the virtual hand representation of the participant was redirected while reaching out. The redirection gain started 30cm in front of the keypad and reached the maximum redirection amount at the keypad. In the shifting condition, the whole keypad jumped to a new spot after each key input. After putting in eight PINs, participants were asked to answer the questionnaire displayed in the virtual environment by adjusting virtual sliders. They used their non-dominant hand to answer the seven displayed questions to avoid too much strain on the dominant arm from



(a) Average number of incorrectly entered digits per PIN for each combination of Size and Technique. Shown error bars represent the standard error (SE).



(b) Distribution of Task Completion Times in seconds for each combination of Size and Technique.

Fig. 4. Distributions of our performance measurements **Error Rate** (a) and **Task Completion Time** (b).

constantly holding it up. After submitting their answers, the next condition was started. After seven conditions, participants took a short break of five minutes. Overall, the study took $\sim 60min$. With their informed consent, we video-recorded the participants during all trials. This study was approved by the ethics committee of our university.

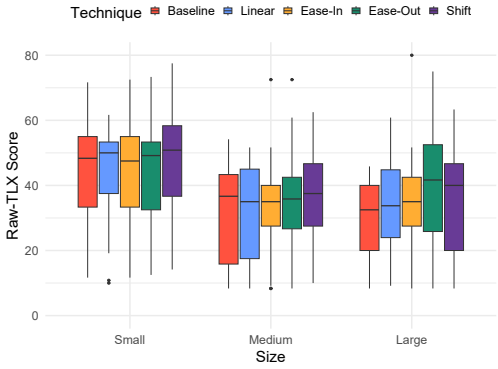
4.4 Participants

We recruited 30 participants (13 female, 17 male) between 19 and 50 years old (M = 25.17, SD = 6.07). 26 participants were right-handed, 3 were left-handed, and one was ambidextrous or unsure about their handedness but used their right hand to complete the study. 26 participants used VR before (12 participants below 2 hours, 9 between 2 and 20 hours, and 5 over 20 hours). Participants' height ranged from 159cm to 193cm (M = 173.07, SD = 7.99). We offered 15€ (~16$) as compensation.

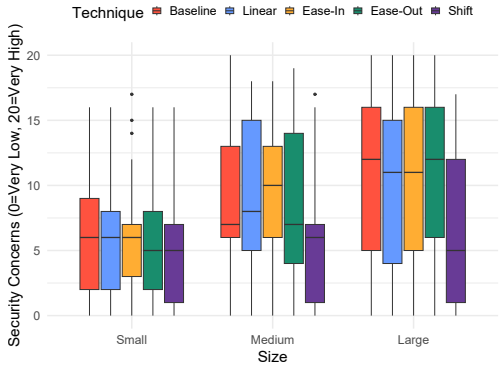
4.5 Results

In the following section, we report our results structured based on the measurements described in Section 4.1. The means and standard deviations of all measurements for each combination of Technique and Size can be found in Appendix A.

4.5.1 Error Rate. To measure participants' performance, we analyzed their accuracy by tracking the number of incorrect digits per PIN. We then fitted Poisson regression models via maximum likelihood (Laplace Approximation) and applied type III Wald chi-square tests for significance testing. The average number of errors per PIN can be seen in Figure 4a and ranges from $M = 0.05$, $SD = 0.22$ (Shift/Large) to $M = 0.42$, $SD = 0.82$ (Linear/Small). We found significant main effects for Technique ($\chi^2(4) = 94.81, p < .001$) and Size ($\chi^2(2) = 8.14, p < .05$) and a significant interaction effect of Technique and Size ($\chi^2(8) = 106.19, p < .001$). Post-hoc tests with Bonferroni correction reveal significantly ($p < .0001$) higher numbers of errors for the Small keypad ($M = 0.35$, $SD = 0.72$) compared to the Medium ($M = 0.19$, $SD = 0.48$) and Large ($M = 0.20$, $SD = 0.51$) keypads. Furthermore, the Linear ($M = 0.32$, $SD = 0.65$), Ease-in ($M = 0.32$, $SD = 0.65$), and Ease-out ($M = 0.24$, $SD = 0.55$) techniques cause significantly ($p < .01$) higher number of errors than the Baseline ($M = 0.16$, $SD = 0.50$) and Shift ($M = 0.19$, $SD = 0.54$) conditions. We found an effect of Size on the Baseline and Shift techniques, where both techniques show significantly ($p < .0001$) increased



(a) Distribution of Raw-TLX scores for each combination of Size and Technique.



(b) Distribution for each combination of Size and Technique of subjective ratings regarding the question: *In a shared space, how concerned would you be that bystanders can guess your input?* Participants rated their concern on a scale from *Very Low* (0) to *Very High* (20).

Fig. 5. Distributions of our subjective measurements **NASA-TLX** (a) and **Subjective Security** (b).

error rates for the Small keypad compared to the Medium and Large one. However, for Ease-in and Ease-out techniques, we see no significant difference between the Small, Medium, and Large groups. Both techniques even reported lower error rates than the Baseline for the Small keypad size, but these differences are not significant. For Linear, we only found a significant ($p < .01$) difference between the Linear Medium and Linear Small group.

4.5.2 Task Completion Time. As another measurement for performance, we collected the time taken to complete each PIN. For analysis, we first performed an Aligned Rank Transform (ART) due to violating the normality of residuals based on Shapiro-Wilk's test. We then performed a two-way repeated measures (RM) ANOVA on the transformed data with Size and Technique as factors. The distribution of task completion tasks are shown in Figure 4b. We found completion times ranging from *M = 8.02s, SD = 2.14s* (Baseline/Large) to *M = 11.37s, SD = 3.52s* (Linear/Small). Analogously to error rates, we found a significant main effect of Technique ($F_{4,116} = 11.49, p < .001$) and Size ($F_{2,58} = 110.34, p < .001$) and a significant interaction effect of Technique and Size ($F_{8,232} = 3.78, p < .001$). Post-hoc tests with Bonferroni-correction show significant ($p < .0001$) higher task-completion times in the Linear (*M = 10.12s, SD = 2.98s*), Ease-out (*M = 10.43s, SD = 3.22s*), and Ease-in (*M = 10.15s, SD = 2.93s*) techniques compared to both the Baseline (*M = 9.36s, SD = 3.10s*) and Shift (*M = 9.71s, SD = 3.34s*) techniques. Furthermore, Shift caused a significantly ($p < .05$) higher completion time than the Baseline. Again, the Small (*M = 11.31s, SD = 3.43s*) keypad performed worse with significantly ($p < .0001$) higher task completion times compared to Medium (*M = 9.33s, SD = 2.86s*) and Large (*M = 9.21s, SD = 2.62s*) sizes. When we look only at the Small keypad, we see no significant differences between Techniques. For both the Medium and Large sizes, the Shift and Baseline conditions required significantly ($p < .05$) less time to complete compared to the other Techniques.

4.5.3 NASA-TLX. We excluded nine participants who skipped a question from the analysis of the NASA-TLX. For the rest (n = 21), we calculated the raw-TLX score [14] by averaging the scores

of the six subscales and mapping it to a 0-100 scale. As suggested by Wobbrock et al. [38] for non-parametric data, we performed an ART on the raw-TLX scores before conducting a two-way RM-ANOVA (factors Size and Technique). We then applied pairwise post-hoc tests with Bonferroni correction. Raw-TLX scores ranged from *M = 29.4, SD = 12.9* (Baseline/Large) to *M = 47.7, SD = 16.6* (Shift/Small). Their distribution is shown in Figure 5a. We found significant main effects of Technique ($F_{4,80} = 3.32, p < .05$) and Size ($F_{2,40} = 38.31, p < .001$) on the raw-TLX scores, but no interaction effects between them. Analogously to the performance, the Small (*M = 44.5, SD = 15.7*) caused significantly ($p < .0001$) higher task loads compared to both Medium (*M = 34.1, SD = 14.8*) and Large (*M = 35.1, SD = 16.3*) sizes. Regarding the Techniques, participants reported significantly ($p < .05$) higher task loads of the Shift (*M = 39.6, SD = 16.8*) and the Ease-out (*M = 39.5, SD = 16.6*) condition compared to Baseline (*M = 34.8, SD = 15.7*). The remainder of Techniques did not significantly impact the reported task load scores.

4.5.4 Subjective Security. For the analysis of the subjective rating of participants' concerns regarding whether bystanders can guess their input, we excluded five participants due to the question being skipped at least once. For the remainder, we followed the same procedure as for the NASA-TLX by applying ART, performing a two-way RM-ANOVA, and using Bonferroni-corrected post-hoc tests on groups. We found reported concerns ranging from *M = 5.64, SD = 5.22* (Shift/Medium) to *M = 10.8, SD = 6.82* (Baseline/Large). Figure 5b shows the distribution of participant's ratings. We found significant main effects of Technique ($F_{4,96} = 7.21, p < .001$) and Size ($F_{2,48} = 19.26, p < .001$) and an interaction effect between both ($F_{8,192} = 3.23, p < .01$). The Small keypad (*M = 5.91, SD = 4.50*) significantly ($p < .0001$) decreased the reported concern compared to other sizes, and the Shift technique (*M = 5.93, SD = 5.32*) significantly ($p < .001$) decreased the reported concern compared to all other Techniques. For the Shift technique, the Size of the keypad did not influence the score significantly, while the Large and Small conditions for all other techniques differ significantly ($p < .05$).

4.6 Discussion

To summarize, we see Size to have a significant impact on performance, the task load, and the perceived security of the system. Participants generally performed worse on the Small keypad and reported higher subjective task loads but also felt less concerned about bystanders being able to guess their inputs. Regarding the techniques, Shift stood out as the one performing the closest to the Baseline, but also introducing the highest subjective task load. In contrast, the hand redirection techniques (Linear, Ease-In, and Ease-Out) performed worse in both error rates and completion time but did not impact the subjective task load significantly over the Baseline – with the exception of the Ease-Out condition. In this section, we will address and discuss our main findings.

4.6.1 Balancing Performance and Task Load. None of the investigated masking Techniques were able to remain on the level of the state-of-the-art Baseline technique when looking at both performance and subjective task load. Instead, we saw the Shift technique performing well while increasing the subjective task load. Conversely, the hand redirection techniques – specifically, Linear and Ease-In – performed worse but did not impact task load as extensively. These discrepancies regarding performance and task load between the discrete shifting and continuous hand redirection could stem from their detectability. Shift always takes place before the participant starts their movement and is always noticeable. Consequently, it does not require active trajectory adaption that would impact performance. This noticeability, however, also disrupts the virtual experience and could be the cause for the higher subjective task loads. Hand redirection works more subtly, which could explain the lower impact on subjective task load – as they often do not notice it actively – and the decrease in performance – because they have to adapt their trajectory during the

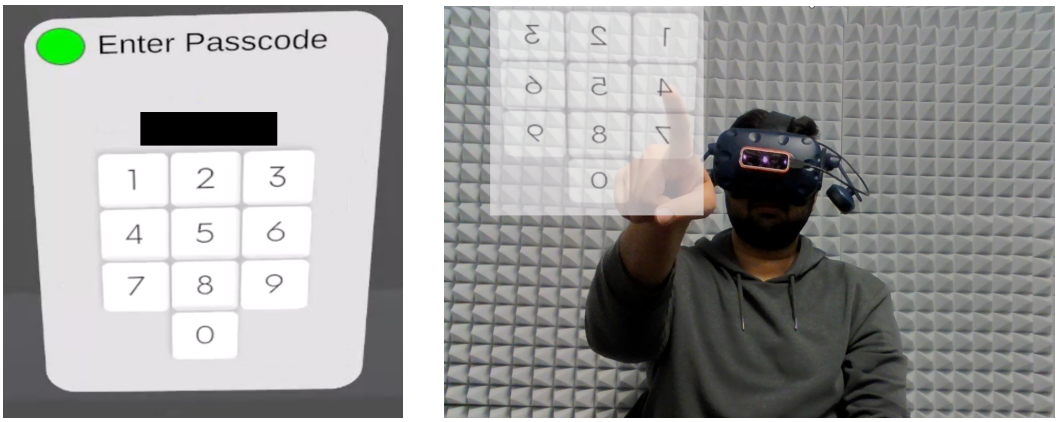
movements. This influence is revealed the strongest in the number of errors for the large condition. Here, we see the Ease-In technique, which applies more redirection gain in the final (correction) phase of the trajectory to perform the worst. This is followed by the Linear technique, which applies constant gains throughout. Lastly, the Ease-Out technique resulted in the least amount of errors from the three hand redirection techniques, which we attributed to the redirection decreasing over the movements, thus causing less drift when the participants were close to hitting their target. Similarly to the perceived task load and performance, the detectability of the techniques seems to be the biggest influence on participants' concerns regarding security against outside observers – with the less noticeable redirection techniques causing more concerns than the clearly observable shift of the keypad. These findings showcase the necessity of carefully weighing the importance of performance and subjective experience of users during PIN authentication. As authentication constitutes only a small portion of the interactions with VR systems, a higher task load or larger completion time might sometimes be acceptable. Making an error during PIN input could lead to higher frustrations, so hand redirection techniques in larger keypads – and consequently larger redirection magnitudes – must be employed carefully.

4.6.2 Effect of Redirection and Shifting Magnitude. To ensure that masking techniques obscure the entered digits, the amount of redirection or shifting must be at least equal to the distance between the closest keys. Smaller amounts would cause the user and observer to perceive different positions, but they would be on the same key. For this study, we, therefore, adjusted the magnitudes of redirection and shifting to the keypad sizes. In the Small condition, the applied hand redirection of 4° did not significantly increase the subjective task load or decrease performance. This suggests that the redirection remained undetected by users. This is consistent with previous work by Zenner and Krüger [42], which reported a detection threshold for horizontal and vertical hand redirections (as in our cases) of 4.5°. Other work generally found even higher detection thresholds when users were distracted by other tasks [4, 8]. Thus, we conclude that hand redirection techniques with magnitudes below the perceptual threshold can be employed without affecting performance or subjective experience. The same cannot be said for Shifting, which works discretely before every press and whose task load results indicate that it disrupted participants' experience. The Medium and Large sizes resulted in redirection angles of 8° and 16° , which potentially have been detected more frequently by participants. However, these angles remained well below the threshold where they become unacceptable, identified by Cheng et al. [8] at 40°. This finding is supported by the subjective task loads of our participants for the Linear and Ease-In redirection techniques, which do not significantly differ from the Baseline.

4.6.3 Effect of Keypad Size. Our user study demonstrates the critical importance of keypad size. The results indicate that smaller keypads tend to result in slower authentication times, an increased number of errors, and an increased task load. The notable decline observed for the Small keypad suggests that it may be inherently too small for accurate input despite its dimensions ($\sim 5cm \times 6.7cm$) exceeding common keypad sizes on smartphones. This may be attributed to limitations in optical hand-tracking, which – while remaining the prevailing technology to facilitate free-hand interactions – still struggles with finger-sized inputs. This necessitates the usage of larger virtual interfaces, which result in more pronounced and, therefore, easier-to-read hand and body movements, intensifying the risk of shoulder surfing. Consequently, like in our Medium and Large keypads, larger magnitudes of redirection or shifting are required to mask the entered PINs.

5 Study II: Security against human observers

To evaluate the efficacy of the investigated masking techniques against shoulder surfing attacks of observers (*RQ2*), we conducted an online survey using the videos recorded during the PIN input in



(a) VR user's perspective (b) Camera's perspective

Fig. 6. The layout of the virtual keyboard as shown to participants of the online survey (Left) and an example for survey participants showing what the keypad would look like if overlayed from the camera's perspective (Right). These two images were displayed next to every video to remind participants. The videos themselves had no overlays and only showed a VR user typing in mid-air.

the first study and asked the survey participants to guess the PINs that were typed in. We employed a within-subject design in which each survey participant saw a video of every condition from the first study. The independent variables remain the levels of Technique used to mask the inputs and Size of the virtual keypad.

5.1 Survey Construction

First, we removed all videos where the PIN was put in incorrectly. We then randomly selected 150 of the remaining recordings so that each of the 15 conditions (5 Techniques × 3 Sizes) is represented by 10 different videos. Additionally, we balanced the videos so that each recorded VR user (n = 30) was shown exactly 5 times. We aimed the survey at 50 participants so that each of the 150 videos would be watched exactly 5 times, and each participant would watch 15 videos, one for each condition. We automated the selection of the videos to avoid introducing bias but checked each selected video before uploading to ensure there were no privacy violations (e.g., the recorded participant took off the headset, and their uncovered face was shown) or recording errors occurred. We marked 18 videos with issues and let the automation replace and balance them. We removed the sound from all videos.

5.2 Procedure

First, the survey informed the participants about the general aim and procedure of the study and asked for their consent and demographic data. Then, detailed instruction and information regarding their tasks was given. Participants were asked to watch 15 short videos corresponding to the 15 conditions of the first study (5 Techniques × 3 Sizes). Each video was less than 15 seconds long, but the survey participants could stop and replay the videos as often as they liked. In each video, a VR user was shown putting in a 4-digit PIN by typing it in mid-air with free hands. The task was to try to guess the 4-digit PIN by observing the movements of the hand. Each survey participant had five guesses that they could put in for each video. No feedback about correctness was given during the study. Survey participants were informed that the PIN is always exactly four digits long and put

in correctly in the video. Furthermore, we informed them about the Size (small, medium, or large) of the virtual keypads in the videos. The survey explained that masking techniques might conceal some of the inputs, but no further details regarding the Techniques were given to ensure that the survey participants had equivalent insights into the system as a malicious observer would have. We recorded the videos from the front of the user. Although this results in the keyboard being mirrored for observers, it ensures that all of the VR user's hand movements are always in frame, unobstructed by their own body, and less affected by perspective distortions. The chosen perspective requires mental rotation, which could affect the success rate and completion times. However, as participants had no time or repetition constraints, we favored this perspective over others that might contain incomplete information. To remind the survey participants, we showed an image of the keypad layout from the VR user's perspective (see Figure 6a) and another image showing what it would look like from the outside camera's perspective for each video (see Figure 6b). Before starting the test conditions, we gave an example with a VR user typing in a PIN on a large keypad without masking techniques. We showed the video first without any additional overlays, like the videos they will see for each condition, and followed this with a video where we overlayed a virtual keypad at the right position to show how the input worked. We also gave them the solution of the PIN for the example video. Afterward, the survey participants proceeded through the 15 videos. The order of videos was randomized. The study was approved by the ethics committee of our university.

5.3 Participants

We recruited 50 participants (25 female, 25 male) between the ages of 20 and 53 (M = 28.16, SD = 9.25) on Prolific. None of the participants took part in the first study. The participants took, on average, 40 minutes (SD = 18.2min) to complete the survey and were compensated with 4.5£ (~5.6$). To give an extra incentive to try to solve the PINs correctly, we offered bonus compensation of 1£ (~1.3$) for each PIN guessed right.

5.4 Results

To measure the accuracy of participants in guessing the PINs being put in, we first checked for completely correct guesses of the entire PIN. We found the Baseline to provide by far the highest chance for guessing the PIN. Our 50 participants could guess the PIN of the Baseline technique 21 times (14× for Large, 6× for Medium, and 1× for Small). The Techniques employing redirection or shifting reduced these numbers to 4× for Ease-Out (1× for Small, 3× Medium), 3× for Linear (2× for Small, 1× for Large) and only 1× for each Ease-In and Shift, both with the Large keypad.

For a more in-depth comparison, we counted the number of correctly guessed digits at the correct position within the PIN for each guess. We then fitted Poisson regression models via maximum likelihood (Laplace Approximation) and applied type III Wald chi-square tests for significance testing. Figure 7 shows the average number of correct digits per PIN for each combination of Technique and Size. Randomly guessing each digit of the PIN would result in an expected rate of 0.4 correct guesses per PIN. We found an average number of correctly guessed digits per PIN ranging from *M = 0.53, SD = 0.76* (Baseline/Small) to *M = 1.42, SD = 1.33* (Baseline/Large). We found significant main effects for Technique ($\chi^2(4) = 73.37, p < .001$) and Size ($\chi^2(2) = 85.40, p < .001$) and a significant interaction effect of Technique and Size ($\chi^2(8) = 57.80, p < .001$) on the number of correctly guessed digits. Pairwise post-hoc tests with Bonferroni correction display a significantly ($p < .0001$) higher number of correctly guessed digits for the Large (*M = 1.11, SD = 1.05*) condition compared to Medium (*M = 0.81, SD = 0.88*) and Small (*M = 0.64, SD = 0.79*) conditions and a significantly higher number for the Medium compared to the Small size. Regarding the Techniques, Shift (*M = 0.66, SD = 0.73*) produces significantly ($p < .01$) fewer correctly guessed digits compared to all others. Only looking at the Small size, we see no significant differences between Techniques,

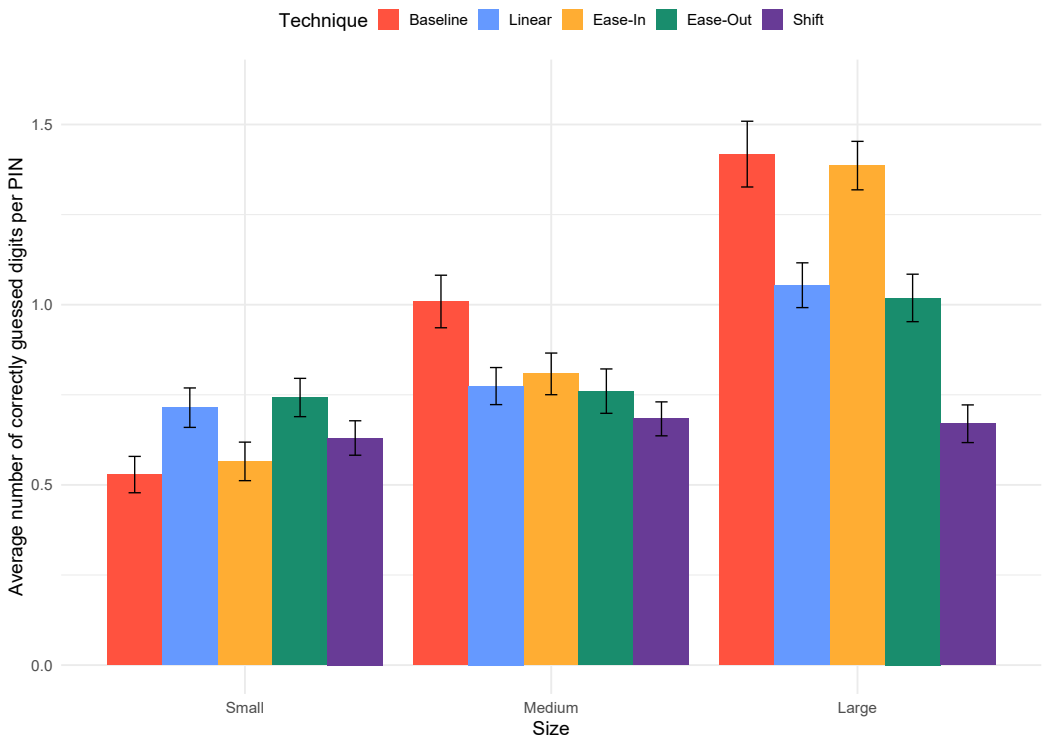


Fig. 7. Average number of correctly guessed digits (in their right positions) per PIN for each combination of Size and Technique. Shown error bars represent the standard error (SE).

with all of them being generally low. However, for the Large keypad, we see significant ($p < .05$) decreases in the number of correct guesses for Linear, Ease-Out, and Shift compared to the Baseline. Furthermore, for the Large conditions, Shift produces significantly ($p < .05$) fewer correct guesses than all others. Within the Medium size, only Shift and Baseline resulted in significantly ($p < .05$) different rates. We see Size mostly having an effect on the Baseline and Ease-In techniques; they show the highest numbers of successful guesses in the Large conditions, which significantly ($p < .01$) decrease for Medium and Small.

5.5 Discussion

Our findings highlight the necessity of mechanisms protecting PIN authentication in VR from shoulder surfing, with 14 survey participants being able to guess the PIN on the Large keypad, and still 6 participants being able to do so for the Medium size in the unprotected Baseline conditions. A smaller size or using our proposed masking techniques dropped the number of successful attacks drastically, although not entirely to zero. The more in-depth analysis revealed that while all Techniques lowered the success rate of attackers, Shift outperformed the others. This likely originates from the fact that its discrete displacement before the hand movement has started does not induce online adaptions of trajectories but simply pivots the hand's target from the beginning. In contrast, hand redirection techniques, which apply this drift continuously, cause the user to divert from their original target during the trajectory. This leaves a short period at the beginning of the redirection, where the physical hand still moves analogously to an unredirected

hand until the visual drift is unconsciously compensated. In the Ease-In technique, this effect is the strongest due to it accumulating drift more slowly at the beginning of the hand movement, thus leaving the originally planned trajectory visible for a longer duration. This could explain the high number of correctly guessed digits in the Ease-In condition for the Large keypad, as the VR users in the video made large hand movements that were adjusted late. Conversely, the Linear and Ease-Out techniques do not exhibit this issue and significantly decrease the success chance of our attackers. The Small keypad also successfully prevented guessing even without explicit protection using a masking technique. These insights show that large, unconcealed hand movements made during PIN input in VR are a critical added risk compared to authentication on a smartphone or ATM machine. Having no or an insufficient masking technique lets observers have a much higher chance of guessing partial digits of a PIN and employing no method of concealing hands for large keypads might result in every third to fourth layperson being able to simply guess a user's 4-digit PIN entirely.

6 Design Recommendations

In this section, we provide design recommendations and general considerations for PIN authentication in VR derived from the findings of our two experimental studies. First, we address the necessary balancing of performance, user experience, and security (6.1), followed by recommendations regarding the context in which these techniques could be employed (6.2) and a discussion of users' perceived and actual security while entering PINs in VR (6.3).

6.1 Use Shift for the Highest Security and Linear Redirection to Prioritize Usability

While we already addressed the necessity of weighing performance and user experience (see section 4.6.1), security is the third important dimension that has to be balanced to allow for a usable *and* secure authentication system. In prior work, we have observed different ends of this spectrum. George et al. [11] showed that the established PIN and pattern techniques work efficiently and intuitively but show great risks when it comes to shoulder surfing. Knowledge-based approaches using randomization and more complex input methods, such as controller gestures, instead favor security over performance and subjective user experience, as shown by Abdelrahman et al. [1]. Behavioral biometric approaches, on the other hand, are generally applied more implicitly, with methods such as BioMove by Olade et al. [27] able to authenticate users effectively without disrupting their experience. However, they also demonstrated the system's vulnerability to mimicking attacks, especially with attackers of similar build and height as the victims.

Similarly, the results of our two experimental studies indicate that there is no one-size-fits-all approach for concealing PIN authentication. Shifting the keypad discreetly was found to have the least impact on performance, as it avoids online adjustments that could lead to user errors. This was reflected in our security evaluation, where Shift demonstrated the greatest reduction in attacker success rates. However, it notably increased the subjective task load for users, compromising its usability. On the other hand, hand redirection techniques, particularly on smaller keypads, had less impact on the task load but resulted in longer completion times and more errors. While all techniques drastically reduced the likelihood of attackers guessing the complete PIN, further analysis on individually guessed digits revealed significant differences, with the Ease-In technique performing the worst among masking methods, likely due to the slow application of redirection during the initial grasping phase. Ease-Out and Linear were more effective but still fell short of Shift. Among hand redirection methods, Linear emerged as the most favorable across all aspects, surpassing Ease-Out in usability and Ease-In in security. Consequently, we recommend Shift to achieve the highest security and Linear redirection to take users' subjective task load into account.

6.2 Context of Use and Data Sensitivity Change the Required Security & Usability

Ultimately, the optimal balance among performance, experience, and security should be modeled to the specific use case and context in which PIN authentication takes place. Authentication generally takes up only a small sliver of our time spent in virtual environments but also poses the largest threat to our security. Therefore, in environments where protecting security-sensitive data is crucial, such as public transit or communal areas, minor inconveniences to the virtual experience or requiring users to type in their PIN with heightened attention might be valid trade-offs when it disrupts a possible attack on security-critical data, such as our payment or personal information. In these cases, smaller keypads masked by discrete shifting provide the most secure PIN authentication. However, to secure highly sensitive data, masked PIN authentication could be further supported by a second-factor authentication, such as physiological or behavioral biometrics. The Apple Vision Pro, for example, already allows for iris-based biometric authentication[6]. However, similar to smartphones, where identification by face or fingerprint is common, PIN authentication is still required if biometric authentication is unsuccessful or the device has been restarted or not unlocked for a while. Alternatively, a single-factor PIN authentication may suffice if the authentication is intended for less safety-critical purposes or employed in a more trusted environment. For instance, it could be employed to differentiate between various users, such as roommates, colleagues, or parents and their children who share the same device. While a certain level of trust can be assumed among them, users should still have the means to safeguard their privacy. Consequently, such verifications would need to occur more frequently – i.e., each time someone puts on the headset or logs into a social platform. Here, more seamless integration and reduced disruption of the virtual experience are crucial, rendering subconscious hand redirection techniques on larger, easier-to-use keypads more appropriate.

6.3 Be Aware of Users' Security Awareness

While most of the masking techniques significantly increased the security of the authentication against shoulder surfing attacks, only the Shift technique was perceived to be very secure subjectively by our VR participants using them. Trust in these systems may, however, be a double-edged sword. While we might not want to use a system that we deem insecure, putting too much trust in a system to protect us – in this case, from outside observers – may cause us to take more risks [35], e.g., typing in our PIN with large hand movements on mass transit or in a public building without checking for observers or pre-installed devices. Here, it is interesting to note that while the Shift technique is clearly noticeable, the employed hand redirection techniques oftentimes remain unnoticed, which could explain the lower ratings for their subjective security. Thus, the transparency to users about how these redirection techniques work may influence their perceived security, leaving it to the designers of these PIN authentication systems to decide how cautious or improvidently they want the users to be.

7 Limitations & Future Work

Although we build on the intuitiveness of the established method of PIN entry via a keypad, the use of the presented masking techniques nevertheless causes additional constraints, as well as challenges and opportunities for future work.

[6] https://support.apple.com/en-us/118483, last accessed: 2023-05-20

7.1 Future Advancements in Hand Tracking Could Enable Smaller Necessary Trajectories

For hand redirection to work properly, we require a certain travel distance of the user's hand for the drift to increase continuously. We, therefore, required participants of our first study to pull back their hands entirely before putting in another digit on the virtual keypad. Solely using finger movements instead of large hand movements would decrease the amount of effort and required time for entering the keys. Redirecting small finger movements analogously to hand redirections is plausible, as prior work has shown visual hand pose manipulations to be accepted into our body schema [5, 6]. However, current optical hand tracking technology still presents limitations in regard to precise tracking of all fingers, which was also evident for our smallest keypad size tested. Although we employed additional hardware for hand tracking, which surpasses the accuracy of most systems built into mobile VR or AR headsets, we saw significant decreases in performance for the smaller targets, even without any redirection or shifting techniques. We are confident that future improvement will make free-hand input more effortless and, consequently, single-finger redirections feasible.

7.2 Optimizing or Randomizing Algorithms Could Increase Performance, Experience, and Security

In order to accurately compare and evaluate the presented techniques, we opted to keep the total amount of redirection and shifting constant between individual PINs and key entries and solely constrained it to the keypad size. This may have implications regarding the security of these techniques, as a constant magnitude of the offsets should be easier to determine by observers. A straightforward solution for this in future PIN authentication systems would be to randomize the amount of total offset being applied to the hand or keypad in addition to the randomly chosen direction of the drift. Furthermore, we limited ourselves to three different functions of hand redirection to test the commonly used approach as well as the effect of more redirection in the beginning (ballistic phase) and end (correction phase) of the trajectory. We employed these algorithms to acquire insights into the impact of different redirection functions on users' performance, experience, and security to serve as a first step towards optimizing redirection functions for PIN authentication or other purposes requiring free-hand key input. The effect of other and more optimized redirection functions or the random switching between them for each PIN or individual key warrants further investigation.

7.3 Scaling to Larger Keyboards Can Increase Security But Poses Challenges Regarding Usability

Lastly, while we investigated the feasibility and efficacy of the proposed masking techniques on PIN authentication with a 4-length PIN on a 10-key keypad (numbers 0 to 9), this approach can be extended to any virtual keyboard and PIN length. We selected a numeric keypad over a larger keyboard based on prior works [11, 24, 39] and the fact that these keypads are still remarkably pervasive despite their vulnerabilities. Extending the PIN length is possible without adjustment and can significantly increase the security of masked PIN authentication methods. Using an alphanumeric keyboard with a larger number of possible keys would also inherently increase security, but presents challenges regarding the amount of redirection and shifting necessary and the needed accuracy of input. In this case, redirection could be used in smaller sections of a larger keyboard to conceal the individual keys typed among those nearby, eliminating the necessity for redirection to encompass the entire keyboard. However, how the findings of our two experimental studies may be extended to other types of keyboards remains to be explored.

8 Conclusion

Mobile headsets have extended VR and AR interaction outside our homes, prompting the need for robust security measures to safeguard our personal data. However, the standard PIN entry method via virtual keypads is vulnerable to shoulder-surfing, where an unnoticed observer can monitor our movements. To counter this, we proposed and investigated masking techniques that conceal hand movements during PIN authentication through hand redirection or shifting of the virtual keypad. Through two experimental studies, we found that these methods can substantially increase a user's protection against shoulder-surfing attacks without disrupting their experience or performance. Specifically, we found discrete shifting of the virtual keypad to perform the most efficiently and accurately, though causing adverse effects on user experience. In contrast, hand redirection techniques did not adversely impact user experience but did show a negative influence on performance. By weighing these aspects and adapting the masking techniques to individual contexts, we can seamlessly integrate added protection without disrupting the virtual experience. These findings mark an important step towards enhancing the security of future VR authentication.

Acknowledgments

This project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 425869442 and is part of Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments.

References

[1] Yomna Abdelrahman, Florian Mathis, Pascal Knierim, Axel Kettler, Florian Alt, and Mohamed Khamis. 2022. CueVR: Studying the Usability of Cue-Based Authentication for Virtual Reality. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces* (Frascati, Rome, Italy) *(AVI 2022)*. Association for Computing Machinery, New York, NY, USA, Article 34, 9 pages. https://doi.org/10.1145/3531073.3531092

[2] Parastoo Abtahi and Sean Follmer. 2018. Visuo-Haptic Illusions for Improving the Perceived Performance of Shape Displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173724

[3] Karan Ahuja, Rahul Islam, Varun Parashar, Kuntal Dey, Chris Harrison, and Mayank Goel. 2018. EyeSpyVR: Interactive Eye Sensing Using Off-the-Shelf, Smartphone-Based VR Headsets. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 57 (jul 2018), 10 pages. https://doi.org/10.1145/3214260

[4] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. 2016. Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1968–1979. https://doi.org/10.1145/2858036.2858226

[5] Yuki Ban, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2014. Displaying Shapes with Various Types of Surfaces Using Visuo-Haptic Interaction. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (Edinburgh, Scotland) *(VRST '14)*. Association for Computing Machinery, New York, NY, USA, 191–196. https://doi.org/10.1145/2671015.2671028

[6] Raoul Bickmann, Celine Tran, Ninja Ruesch, and Katrin Wolf. 2019. Haptic Illusion Glove: A Glove for Illusionary Touch Feedback When Grasping Virtual Objects. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC'19)*. Association for Computing Machinery, New York, NY, USA, 565–569. https://doi.org/10.1145/3340764.3344459

[7] Matthew Botvinick and Jonathan Cohen. 1998. Rubber hands 'feel' touch that eyes see. *Nature* 391, 6669 (01 Feb 1998), 756–756. https://doi.org/10.1038/35784

[8] Lung-Pan Cheng, Eyal Ofek, Christian Holz, Hrvoje Benko, and Andrew D. Wilson. 2017. Sparse Haptic Proxy: Touch Feedback in Virtual Environments Using a General Passive Prop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3718–3728. https://doi.org/10.1145/3025453.3025753

[9] Martin Feick, Kora P. Regitz, Anthony Tang, Tobias Jungbluth, Maurice Rekrut, and Antonio Krüger. 2023. Investigating Noticeable Hand Redirection in Virtual Reality using Physiological and Interaction Data. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 194–204. https://doi.org/10.1109/VR55154.2023.00035

[10] Ceenu George, Mohamed Khamis, Daniel Buschek, and Heinrich Hussmann. 2019. Investigating the Third Dimension for Authentication in Immersive Virtual Reality and in the Real World. In *2019 IEEE Conference on Virtual Reality and*

3D User Interfaces (VR). IEEE, New York, NY, USA, 277–285. https://doi.org/10.1109/VR.2019.8797862

[11] Ceenu George, Mohamed Khamis, Emanuel Zezschwitz, Marinus Burger, Henri Schmidt, Florian Alt, and Heinrich Hussmann. 2017. Seamless and Secure VR: Adapting and Evaluating Established Authentication Systems for Virtual Reality. In *Network and Distributed System Security Symposium (NDSS 2017)*. https://doi.org/10.14722/usec.2017.23028

[12] Benoit Geslain, Simon Besga, Flavien Lebrun, and Gilles Bailly. 2022. Generalizing the Hand Redirection Function in Virtual Reality. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces* (Frascati, Rome, Italy) *(AVI 2022)*. Association for Computing Machinery, New York, NY, USA, Article 33, 9 pages. https://doi.org/10.1145/3531073.3531100

[13] Eric J. Gonzalez, Parastoo Abtahi, and Sean Follmer. 2020. REACH+: Extending the Reachability of Encountered-type Haptics Devices through Dynamic Redirection in VR. *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (2020).

[14] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (Oct. 2006), 904–908. https://doi.org/10.1177/154193120605000909

[15] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[16] Mohamed Khamis, Carl Oechsner, Florian Alt, and Andreas Bulling. 2018. VRpursuits: Interaction in Virtual Reality Using Smooth Pursuit Eye Movements. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces* (Castiglione della Pescaia, Grosseto, Italy) *(AVI '18)*. Association for Computing Machinery, New York, NY, USA, Article 18, 8 pages. https://doi.org/10.1145/3206505.3206522

[17] Vrishab Krishna, Yi Ding, Aiwen Xu, and Tobias Höllerer. 2019. Multimodal Biometric Authentication for VR/AR Using EEG and Eye Tracking. In *Adjunct of the 2019 International Conference on Multimodal Interaction* (Suzhou, China) *(ICMI '19)*. Association for Computing Machinery, New York, NY, USA, Article 6, 5 pages. https://doi.org/10.1145/3351529.3360655

[18] Manu Kumar, Tal Garfinkel, Dan Boneh, and Terry Winograd. 2007. Reducing Shoulder-Surfing by Using Gaze-Based Password Entry. In *Proceedings of the 3rd Symposium on Usable Privacy and Security* (Pittsburgh, Pennsylvania, USA) *(SOUPS '07)*. Association for Computing Machinery, New York, NY, USA, 13–19. https://doi.org/10.1145/1280680.1280683

[19] Alexander Kupin, Benjamin Moeller, Yijun Jiang, Natasha Kholgade Banerjee, and Sean Banerjee. 2019. Task-Driven Biometric Authentication of Users in Virtual Reality (VR) Environments. In *MultiMedia Modeling*, Ioannis Kompatsiaris, Benoit Huet, Vasileios Mezaris, Cathal Gurrin, Wen-Huang Cheng, and Stefanos Vrochidis (Eds.). Springer International Publishing, Cham, 55–67.

[20] Sukun Li, Sonal Savaliya, Leonard Marino, Avery M. Leider, and Charles C. Tappert. 2019. Brain Signal Authentication for Human-Computer Interaction in Virtual Reality. In *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*. IEEE, New York, NY, USA, 115–120. https://doi.org/10.1109/CSE/EUC.2019.00031

[21] Jonathan Liebers, Patrick Horn, Christian Burschik, Uwe Gruenefeld, and Stefan Schneegass. 2021. Using Gaze Behavior and Head Orientation for Implicit Identification in Virtual Reality. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (Osaka, Japan) *(VRST '21)*. Association for Computing Machinery, New York, NY, USA, Article 22, 9 pages. https://doi.org/10.1145/3489849.3489880

[22] Lorraine Lin and Sophie Jörg. 2016. Need a Hand? How Appearance Affects the Virtual Hand Illusion. In *Proceedings of the ACM Symposium on Applied Perception* (Anaheim, California) *(SAP '16)*. Association for Computing Machinery, New York, NY, USA, 69–76. https://doi.org/10.1145/2931002.2931006

[23] Florian Mathis, Hassan Ismail Fawaz, and Mohamed Khamis. 2020. Knowledge-Driven Biometric Authentication in Virtual Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3334480.3382799

[24] Florian Mathis, John Williamson, Kami Vaniea, and Mohamed Khamis. 2020. RubikAuth: Fast and Secure Authentication in Virtual Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3334480.3382827

[25] Robert Miller, Natasha Kholgade Banerjee, and Sean Banerjee. 2020. Within-System and Cross-System Behavior-Based Biometric Authentication in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, New York, NY, USA, 311–316. https://doi.org/10.1109/VRW50115.2020.00070

[26] Tahrima Mustafa, Richard Matovu, Abdul Serwadda, and Nicholas Muirhead. 2018. Unsure How to Authenticate on Your VR Headset? Come on, Use Your Head!. In *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics* (Tempe, AZ, USA) *(IWSPA '18)*. Association for Computing Machinery, New York, NY, USA, 23–30.

https://doi.org/10.1145/3180445.3180450

[27] Ilesanmi Olade, Charles Fleming, and Hai-Ning Liang. 2020. BioMove: Biometric User Identification from Human Kinesiological Movements for Virtual Reality Systems. *Sensors* 20, 10 (2020). https://doi.org/10.3390/s20102944

[28] Ilesanmi Olade, Hai-Ning Liang, Charles Fleming, and Christopher Champion. 2020. Exploring the Vulnerabilities and Advantages of SWIPE or Pattern Authentication in Virtual Reality (VR). In *Proceedings of the 2020 4th International Conference on Virtual and Augmented Reality Simulations* (Sydney, NSW, Australia) *(ICVARS 2020)*. Association for Computing Machinery, New York, NY, USA, 45–52. https://doi.org/10.1145/3385378.3385385

[29] Ken Pfeuffer, Matthias J. Geiger, Sarah Prange, Lukas Mecke, Daniel Buschek, and Florian Alt. 2019. Behavioural Biometrics in VR: Identifying People from Body Motion and Relations in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300340

[30] Dario Pittera, Elia Gatti, and Marianna Obrist. 2019. *I'm Sensing in the Rain: Spatial Incongruity in Visual-Tactile Mid-Air Stimulation Can Elicit Ownership in VR Users*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300362

[31] Majed Samad, Elia Gatti, Anne Hermes, Hrvoje Benko, and Cesare Parise. 2019. Pseudo-Haptic Weight: Changing the Perceived Weight of Virtual Objects By Manipulating Control-Display Ratio. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300550

[32] Valentin Schwind, Lorraine Lin, Massimiliano Di Luca, Sophie Jörg, and James Hillis. 2018. Touch with Foreign Hands: The Effect of Virtual Hand Appearance on Visual-Haptic Integration. In *Proceedings of the 15th ACM Symposium on Applied Perception* (Vancouver, British Columbia, Canada) *(SAP '18)*. Association for Computing Machinery, New York, NY, USA, Article 9, 8 pages. https://doi.org/10.1145/3225153.3225158

[33] Yiran Shen, Hongkai Wen, Chengwen Luo, Weitao Xu, Tao Zhang, Wen Hu, and Daniela Rus. 2019. GaitLock: Protect Virtual and Augmented Reality Headsets Using Gait. *IEEE Transactions on Dependable and Secure Computing* 16, 3 (2019), 484–497. https://doi.org/10.1109/TDSC.2018.2800048

[34] Manimaran Sivasamy, V.N. Sastry, and N.P. Gopalan. 2020. VRCAuth: Continuous Authentication of Users in Virtual Reality Environment Using Head-Movement. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, New York, NY, USA, 518–523. https://doi.org/10.1109/ICCES48766.2020.9137914

[35] Steeven Villa, Thomas Kosch, Felix Grelka, Albrecht Schmidt, and Robin Welsch. 2023. The placebo effect of human augmentation: Anticipating cognitive augmentation increases risk-taking behavior. *Computers in Human Behavior* 146 (Sept. 2023), 107787. https://doi.org/10.1016/j.chb.2023.107787

[36] Emanuel von Zezschwitz, Paul Dunphy, and Alexander De Luca. 2013. Patterns in the Wild: A Field Study of the Usability of Pattern and Pin-Based Authentication on Mobile Devices. In *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Munich, Germany) *(MobileHCI '13)*. Association for Computing Machinery, New York, NY, USA, 261–270. https://doi.org/10.1145/2493190.2493231

[37] Yannick Weiss, Steeven Villa, Albrecht Schmidt, Sven Mayer, and Florian Müller. 2023. Using Pseudo-Stiffness to Enrich the Haptic Experience in Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 388, 15 pages. https://doi.org/10.1145/3544548.3581223

[38] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Vancouver BC Canada, 143–146. https://doi.org/10.1145/1978942.1978963

[39] Zhen Yu, Hai-Ning Liang, Charles Fleming, and Ka Lok Man. 2016. An exploration of usable authentication mechanisms for virtual reality systems. In *2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*. IEEE, New York, NY, USA, 458–460. https://doi.org/10.1109/APCCAS.2016.7804002

[40] Nur Haryani Zakaria, David Griffiths, Sacha Brostoff, and Jeff Yan. 2011. Shoulder Surfing Defence for Recall-Based Graphical Passwords. In *Proceedings of the Seventh Symposium on Usable Privacy and Security* (Pittsburgh, Pennsylvania) *(SOUPS '11)*. Association for Computing Machinery, New York, NY, USA, Article 6, 12 pages. https://doi.org/10.1145/2078827.2078835

[41] André Zenner, Hannah Maria Kriegler, and Antonio Krüger. 2021. HaRT - The Virtual Reality Hand Redirection Toolkit. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 387, 7 pages. https://doi.org/10.1145/3411763.3451814

[42] André Zenner and Antonio Krüger. 2019. Estimating Detection Thresholds for Desktop-Scale Hand Redirection in Virtual Reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 47–55. https://doi.org/10.1109/VR.2019.8798143

[43] Yongtuo Zhang, Wen Hu, Weitao Xu, Chun Tung Chou, and Jiankun Hu. 2018. Continuous Authentication Using Eye Movement Response of Implicit Visual Stimuli. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 177 (jan 2018), 22 pages. https://doi.org/10.1145/3161410

[44] Huadi Zhu, Wenqiang Jin, Mingyan Xiao, Srinivasan Murali, and Ming Li. 2020. BlinKey: A Two-Factor User Authentication Method for Virtual Reality Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 164 (dec 2020), 29 pages. https://doi.org/10.1145/3432217

A Tables of Means and SD for all Measurements

	technique	size	mean	sd
1	baseline	large	0.05	0.32
2	baseline	medium	0.07	0.26
3	baseline	small	0.35	0.72
4	ease_in	large	0.40	0.71
5	ease_in	medium	0.26	0.59
6	ease_in	small	0.29	0.65
7	ease_out	large	0.17	0.42
8	ease_out	medium	0.27	0.60
9	ease_out	small	0.28	0.61
10	linear	large	0.32	0.63
11	linear	medium	0.20	0.44
12	linear	small	0.42	0.82
13	shift	large	0.05	0.22
14	shift	medium	0.12	0.41
15	shift	small	0.39	0.76

Table 1. **Error Rate:** Mean values and standard deviations for every combination of Technique and Size for the number of incorrectly entered digits per PIN.

Received February 2024; revised May 2024; accepted June 2024

	technique	size	mean	sd
1	baseline	large	8017.34	2141.91
2	baseline	medium	8799.47	2593.59
3	baseline	small	11245.26	3447.96
4	ease_in	large	9674.65	2175.23
5	ease_in	medium	9463.05	2441.31
6	ease_in	small	11298.73	3621.90
7	ease_out	large	10326.04	3151.48
8	ease_out	medium	9639.82	2853.03
9	ease_out	small	11315.78	3441.52
10	linear	large	9344.02	2355.16
11	linear	medium	9652.81	2507.93
12	linear	small	11365.52	3522.92
13	shift	large	8699.67	2516.30
14	shift	medium	9086.05	3669.31
15	shift	small	11338.28	3121.57

Table 2. **Task Completion Time:** Mean values and standard deviations for every combination of Technique and Size for the task completion time in milliseconds.

	technique	size	mean	sd
1	baseline	large	29.44	12.93
2	baseline	medium	31.79	14.79
3	baseline	small	43.10	16.31
4	ease_in	large	34.44	16.45
5	ease_in	medium	34.01	15.24
6	ease_in	small	43.81	15.82
7	ease_out	large	39.09	17.52
8	ease_out	medium	35.20	16.71
9	ease_out	small	44.33	14.93
10	linear	large	37.22	17.43
11	linear	medium	33.41	13.67
12	linear	small	43.57	15.74
13	shift	large	35.08	16.85
14	shift	medium	35.91	14.58
15	shift	small	47.70	16.65

Table 3. **NASA-TLX:** Mean values and standard deviations for every combination of Technique and Size for the raw-TLX score.

	technique	size	mean	sd
1	baseline	large	10.84	6.82
2	baseline	medium	8.48	5.30
3	baseline	small	6.20	4.43
4	ease_in	large	10.80	6.45
5	ease_in	medium	9.28	4.83
6	ease_in	small	5.96	4.56
7	ease_out	large	10.68	6.43
8	ease_out	medium	8.48	5.87
9	ease_out	small	5.72	4.60
10	linear	large	10.00	6.30
11	linear	medium	9.20	5.40
12	linear	small	6.00	4.24
13	shift	large	6.48	5.87
14	shift	medium	5.64	5.22
15	shift	small	5.68	5.00

Table 4. **Subjective Security:** Mean values and standard deviations for every combination of Technique and Size for the subjective security concern. We asked them: *In a shared space, how concerned would you be that bystanders can guess your input?* Participants rated their concern on a scale from *Very Low* (0) to *Very High* (20).

	technique	size	mean	sd
1	baseline	large	1.42	1.33
2	baseline	medium	1.01	1.09
3	baseline	small	0.53	0.76
4	ease_in	large	1.39	1.02
5	ease_in	medium	0.81	0.86
6	ease_in	small	0.57	0.81
7	ease_out	large	1.02	0.96
8	ease_out	medium	0.76	0.92
9	ease_out	small	0.74	0.81
10	linear	large	1.05	0.93
11	linear	medium	0.77	0.76
12	linear	small	0.71	0.83
13	shift	large	0.67	0.78
14	shift	medium	0.68	0.70
15	shift	small	0.63	0.71

Table 5. **Success Rate:** Mean values and standard deviations for every combination of Technique and Size for the number of correctly guessed digits at their right positions per PIN (from the second study).