# Authentication in Immersive Virtual Environments through Gesture-Based Interaction with a Virtual Agent

Daniel Rupp [*] Philipp Grießer Andrea Bönsch Torsten W. Kuhlen
Visual Computing Institute, RWTH Aachen University, Germany

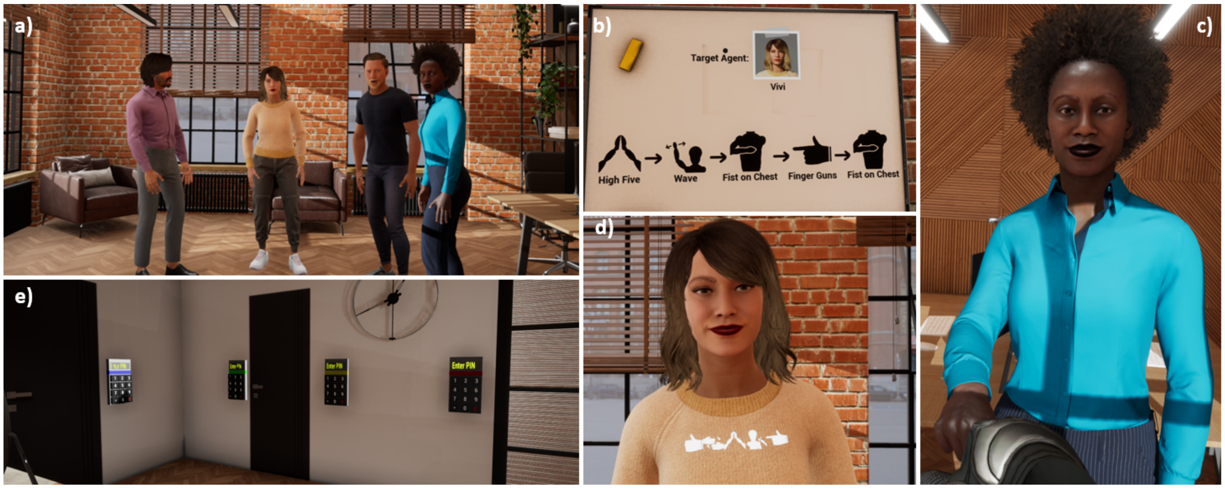

Figure 1: (a) Overview of our virtual office with four potential interaction partners for condition $C_{Gesture}$. During $C_{Gesture}$ the whiteboard (b) is positioned on the wall to the left of the group displaying the password consisting of the target agent and the gesture sequence. (c) The user is performing a *Fist Bump* gesture during the authentication process. After a gesture has been recognized it is displayed as user feedback on the agent's chest (d). (e) The layout of the four virtual keypads during $C_{PIN}$.

## Abstract

Authentication poses a significant challenge in VR applications, as conventional methods, such as text input for usernames and passwords, prove cumbersome and unnatural in immersive virtual environments. Alternatives such as password managers or two-factor authentication may necessitate users to disengage from the virtual experience by removing their headsets. Consequently, we present an innovative system that utilizes virtual agents (VAs) as interaction partners, enabling users to authenticate naturally through a set of ten gestures, such as high fives, fist bumps, or waving. By combining these gestures, users can create personalized authentications akin to PINs, potentially enhancing security without compromising the immersive experience. To gain first insights into the suitability of this authentication process, we conducted a formal expert review with five participants and compared our system to a virtual keypad authentication approach. While our results show that the effectiveness of a VA-mediated gesture-based authentication system is still limited, they motivate further research in this area.

**Index Terms:** Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality; Human-centered computing—Human computer interaction (HCI)—Interaction techniques—Gestural input; Security and privacy—Security services—Authentication—Graphical / visual passwords;

## 1 Introduction

For PC-based VR systems, applications often employ traditional authentication methods such as usernames and passwords, necessitating text input through a physical or virtual keyboard. Physical keyboards, however, prove impractical for portable VR devices due to the inconvenience of carrying them while virtual keyboards tend to be slower compared to their physical counterparts, therefore negatively impacting user experience and efficiency [4].

In our proposed system, we aim to seamlessly integrate authentication within an immersive virtual environment (IVE), leveraging the unique capabilities of VR and 3D input devices. Inspired by social interactions, akin to entering a room and greeting acquaintances, users initiate the authentication process by approaching a designated virtual agent (VA) within a group of VAs upon entering the IVE and performing a predefined action. This interaction mirrors the process of being recognized by a certain acquaintance upon entering a space. By selecting a specific VA and executing a *signature action*, the system controlling the VA confirms the user's identity, completing the authentication process. It is important to note that this one-time authorization, occurring at the outset, grants users the freedom to subsequently interact seamlessly within the IVE.

In a formal expert review with five participants we explored the feasibility of a VA-mediated gesture-based authentication method and compared it to PIN authentication. In summary, our main contribution are:

- the introduction of a novel gesture authentication system
- the exemplary implementation and evaluation of ten gestures to reach the same entropy as numeric PINs based on two types: independent (e.g., waving, thumbs-up) and cooperative (e.g., high five, fist bump)
- first scientific insights from a formal expert review with five participants

[*] e-mail: daniel.rupp@rwth-aachen.de

## 2 Related Work

Our VA-mediated gesture-based authentication technique presented in this paper combines prior research in the realms of both authentication (Sect. 2.1) and VAs (Sect. 2.2) in a VR context.

### 2.1 Authentication

Most modern devices and services used today employ some form of authentication. As attackers gain access to more powerful techniques and devices to breach these security measures, stronger security, such as two-factor authentication, is advised. The widespread availability of established authentication strategies also provides a starting point for newer device classes, such as head-mounted displays (HMDs).

Authentication methods can be classified by different factors [1]: 1) Knowledge (what you know) like passwords, PIN codes, or security questions. 2) Possession (what you have) like hardware tokens or keys. 3) Biometry (what you are) like fingerprints or behavioral biometrics [11].

Combining authentication methods that correspond to different factors is known as multi-factor authentication (MFA). A common example for MFA is a combination of username and password (knowledge) followed by a one-time password received via a secondary device (possession).

Authentication frequently takes place in people's lives, for example when accessing their smartphones. Thus, it should be fast and effortless [11]. So far, widely adapted knowledge-based authentication methods from smartphones, such as PINs and pattern authentication, have been brought into IVEs [8]. To this end, George et al. conducted a user study with HMDs, assessing entry time, error rate and subjective usability ratings for multiple number-based PIN and pattern interface designs. Furthermore, they evaluated their security regarding observation by a third party, using success rate and subjective ratings amongst others [8]. The study showed that on average, number-based PIN authentication took 2.7 seconds and pattern authentication took 3.2 seconds, showing that these established authentication methods are viable for use in VR. The security evaluation, however, showed that 18% of entries were guessed correctly after an observation attack. Existing concepts for improved security could be applied to further leverage the HMD providing a "secret channel" [8], for example, variable layouts for input and the use of randomization to further reduce the risk of a successful observation attack [7].

Biometric authentication requires storing or potentially sharing sensitive personal data, which is generally not required with knowledge-based techniques [10]. This raises privacy concerns as such data has previously been stolen and biometric data cannot be invalidated once it is leaked [6, 7]. Although biometric authentication is highly secure against observation attacks [10], biometric authentication may require a fallback method due to the contextual infeasibility of biometric authentication, e.g., facial recognition while wearing a mask, thus inheriting its weaknesses [7].

Numerous spatial authentication methods have been proposed in VR, however, none have seen widespread adoption. George et al. developed a system in which users select a series of objects in a virtual room through a pointing metaphor [7]. Another proposed technique is RubikAuth, where numbers from one to nine were laid out as a grid on each side of a color-coded six-sided cube attached to the user's hand. By pressing the digits on the faces of the cube, a numeric PIN can be entered [12]. PINs consisted of four digits, and each digit had to be pressed on the correct face of the cube.

### 2.2 Virtual Agents

Computer-controlled anthropomorphic entities present a means for natural interaction between humans and machines. When creating a VA, the designer can control the degree of realism pertaining to its appearance and behavior. Multiple studies indicate that behavioral realism, such as appropriate gaze behavior and gestures, has a strong influence on social presence. Furthermore, while studies evaluating the effect of photographic realism itself show mixed results, the levels of photographic and behavioral realism should be consistent [13].

Realistic gaze is thereby one of the most fundamental behaviors. Observing another's eyes allows humans to interpret their intentions and feelings, while gaze is used for guiding and interpreting social behavior [16]. Before an interaction, the idle behavior can indicate attention, availability and interest as the VA uses gaze to react to the environment, including the user [16]. During a conversation, patterns of directed and averted gaze are a signal for attention and can be used to regulate turn taking, intimacy and to refer to objects [5, 16].

Another fundamental behavior is personal space, the flexible protective zone humans maintain around themselves. It can be subdivided into the intimate, personal, social and public zones [3]. Multiple studies have used VR technologies to show that users prefer the social distance zone, which ranges from 1.2 meters to 3.6 meters, both when approaching and being approached by VAs [3, 15]. When standing in groups, humans use arrangements to signal whether they welcome others to join them. Such an arrangement is called an open formation and can be leveraged when placing VAs in groups. In a study conducted by Rehm et al., users preferred groups standing in an open formation in 84% when tasked to join a group [15].

Another aspect of social interactions is touch, which is an important aspect of our proposed cooperative gestures. In human-to-human communication, touching the communication partner can reduce stress, facilitate bonding and communicate (discrete) emotions [2, 9]. Common touch gestures include handshakes, hugs, and pats on the back [9]. In human-to-agent communication, touch from the VA can be mediated using haptic feedback, which has similar effects as co-located (physical) touch, such as compliance, affect, and social presence [9].

### 2.3 Conclusion

Based on the presented previous work, we propose a system that replaces PINs and patterns with gestures performed in a social setting for more seamless and immersive authentication. Our proposed method combines two established authentication factors, knowledge and biometry, as our authentication process involves physical movements of the user's body and arms to perform a previously memorized gesture sequence with a known interactor. To limit shoulder surfing attacks, as an element of ambiguity, the interaction partner has to be chosen first out of a group of four VAs. The agents in the group were positioned in a semi-open circle close to each other to make it harder for bystanders to identify the VA the user is interacting with and to be able to switch easily between them during the user study. For the VAs appearance, we focused on realistic visuals and fitting fundamental behavior like gaze patterns and animations that respond to the user's behavior. Similar to George et al. a short expert review is conducted comparing PINs to gesture sequences, assessing entry time, error rate, and subjective usability ratings [8].

## 3 System Design

The realization of our VA-mediated gesture-based authentication focuses on the most widespread VR systems, requiring only a HMD with two tracked controllers. No special hand-tracking capabilities are required. All numeric values were selected based on internal testing to fit the IVE and study setup.

The authentication process follows the conceptual model of entering a room and joining a group of VAs of which the user interacts with one, confirming the user's identity. To allow the user to pick the correct target agent for interaction, the VAs are arranged in an open group formation. When the user is within 3 meters of any VA in the group, he is considered part of the group and the VAs gaze shortly (2 to 7 seconds) toward the user to acknowledge his presence without establishing a connection. The interaction partner is determined by the user's proximity and gaze direction. If the user stays within 3 meters and gazes towards a VA (determined by a sphere cast) for at least 1.5 seconds, the VA becomes active, and the user can start interacting with him. To provide the user feedback about what VA is active, the VA smiles and gazes toward the user as long as it is active. If the user looks away from the VA for more than two seconds or if the user moves away further than 3 meters, the VA becomes inactive again.

To be able to perform the gestures, the user has a non-human-like self-avatar with hands, arms, and upper body to provide a sense of self-embodiment without requiring an individual character creation step.

### 3.1 Gestures

In our system, a password can be written as *Agent(Gesture Sequence)*, where *Agent* represents the VA the gesture is directed at and *Gesture Sequence* represents a list of gestures. The reason why we chose to also have *Agent* as part of the password is to tackle the problem of shoulder surfing since an outside observer could see which gestures the user performs, but not with which agent, since this information is hidden in the secret visual channel of the user wearing the HMD.

We implemented ten different gestures, matching the entropy of numeric PINs, which also consist of characters from a set of length ten (0-9). Our chosen gestures are described in Table 1. We chose gestures that are easy to understand, well-known and potentially fun to execute, divided into two groups. *Independent gestures* can be performed without interacting with the VA directly and *cooperative gestures* entailing physical contact between both interactants.

**Independent Gestures**

*Wave:* Waving left-to-right above the shoulder and next to the head.

*Thumbs-Up:* Extending the hand in front of the torso and giving thumbs-up.

*Finger Guns:* Extending the hand in front of the torso and holding a finger guns gesture for one second.

*Fist on Chest:* Making a fist and placing it on the chest (sternum).

*Bow:* Moving the open hand to the chest before bowing forward and down.

**Cooperative Gestures**

*High Five:* Putting the hand up and giving the VA a high five.

*Shake Hands:* Extending the hand, reaching for the VA's extended hand and making a fist.

*Fist Bump:* Extending the fist with the back of the hand pointing up, bumping the knuckles of the VA's extended fist.

*Pat on Shoulder:* Touching the VA's left/right shoulder with the left/right, opened hand.

*Open Hands:* Extending the hand for a handshake, clapping hands together with the VA (the hand moves left and the insides of both hands touch), then clapping the back of both hands together with the VA (the hand moves right and the outsides of both hands touch).

Table 1: Description of the ten gestures implemented to be able to represent numeric PINs, which consist of ten digits (0-9).

Additionally, we added a special *Clear* gesture that is used to delete the currently recognized gesture sequence. To perform the clear gesture the user covers both eyes with their hands.

### 3.2 Gesture Recognition

Within our gesture recognition system, the identification of various gestures is achieved through a comprehensive analysis of the dynamic interplay of the user's hands, considering factors such as motion, orientation, and finger movements. Users seamlessly perform their intended gestures without the need for explicit initiation or termination signals, such as pressing a designated button. However, indirect initiation hints are strategically integrated to improve the gesture recognition process, as detailed in the following.

To start the recognition process, we created different **hand regions** $\mathcal{R} = (Forward, Overhead, Chest, Eyes)$. We assume that when facing and looking at an interaction partner, the location of a person's hand relative to their own body remains unambiguous within a single gesture. For example, the *Shake Hands* gesture always takes place in front of one's body at a comfortable height, while a *Wave* is performed above shoulder height.

Additionally, entering a hand region also triggers an **interaction pose** on the targeted VA. When entering the *Forward Hand Region*, the VA extends his arm to allow the user to perform, e.g., the *Shake Hands* gesture. When entering the *Overhead Hand Region* the VA raises his right or left arm to allow the user to perform a *High Five* or *Wave* gesture.

For the gesture recognition algorithm, we distinguish between three **hand orientations** $\mathcal{O} = (Flat, Upright, FingersUp)$. *Flat* and *Upright* correspond to a controller roll of $-90^\circ$ and $0^\circ$ respectively, while *FingersUp* corresponds to a roll of $-90^\circ$ and a pitch of $90^\circ$. To classify the current hand orientation, we compare the vertical ($Z$) components of the corresponding motion controller's right, up and forward vectors in world coordinates. Then, we find the vector with the largest absolute $Z$ value and use the sign to determine the hand orientation. Other hand orientations are discarded as they are not used by our gestures.

To allow our system to be more generally applicable, we do not rely on individual finger-tracking capabilities. Instead, a standard VR controller is enough. To determine the **finger positions**, capacitive touch on the controller's thumbstick or face buttons determines the flexing of the thumb, pressing the trigger determines the flexing of the index finger and the flexing of the middle, pinky and ring finger is determined as a single unit using the grab button. Although many motion controllers support continuous values and sometimes additional capacitive touch sensing for some buttons, we only consider pressed and released by rounding all values. This simplification is sufficient for our gestures and improves ease of use. It also allows us to represent finger position as bitflag $\mathcal{F} =$ (Bit 2: thumb, Bit 1: index finger, Bit 0: middle, ring and pinky), where 0 corresponds to the fingers being extended and '$X$' indicating either 0 or 1. For example, $\mathcal{F} = 111$ corresponds to a fist, since all fingers are "closed" and $\mathcal{F} = X00$ corresponds to an open hand with either an extended or closed thumb.

We will explain how each gesture is recognized in detail by listing all prerequisites that have to be met in the form of a tuple $\mathcal{P} = (\mathcal{R}, \mathcal{O}, \mathcal{F})$. The gesture recognition process ends either by leaving the respective hand region if the VA becomes inactive or after a successful recognition.

A ***Wave*** ($\mathcal{P} = (Overhead, FingersUp, X00)$) is successfully recognized when the hand has alternated between moving left-to-right and right-to-left at a speed above 90 cm/s at least three times each, without leaving the hand region. Lateral movements are recognized by checking the motion controller's movement direction.

A ***Thumbs-Up*** ($\mathcal{P} = (Forward, Upright, 011)$) is immediately recognized when the prerequisites are met upon entering the designated hand region. The same is true for ***Finger Guns*** ($\mathcal{P} = (Forward, Upright, 001)$) and

***Fist on Chest*** ($\mathcal{P} = (Chest, Flat, 111)$).

For ***Bow*** ($\mathcal{P} = (Chest, Upright/FingersUp, X00)$), upon entering the hand region, the system saves the starting position and view direction vector, which are used to recognize the gesture when both a vertical and horizontal distance threshold of 15 cm have been reached, the user looks down, and the movement direction is roughly forward (as determined by a dot product threshold of 90). To improve recognition, the *Bow* gesture is tracked for at least two seconds, even when the hand leaves the hand region. The remaining time is reset each time the hand enters the hand region again.

For ***High Five*** ($\mathcal{P} = (Overhead, FingersUp, X00)$), as soon as the user raises his hand into the overhead hand region, the VA responds by also raising his hand allowing the user to bring his palm together with the VA's palm. If other prerequisites are met, the *High Five* gesture is recognized.

For ***Shake Hands*** ($\mathcal{P} = (Forward, Upright, X00 \text{ and } 111)$), upon entering the forward hand region, the VA extends his arm allowing the user to grab it by closing his hand into a fist once the user's palm collides with the VA's palm.

For ***Fist Bump*** ($\mathcal{P} = (Forward, Flat/Upright, 111)$), upon entering the forward hand region, the VA extends his arm. As soon as the user closes his hand into a fist, the VA does so as well. When the knuckles of the user collide with those of the agent, the gesture is recognized.

***Pat on Shoulder*** ($\mathcal{P} = (-, -, X00)$) is the only gesture involving touch without requiring an interaction pose, the gesture is recognized as soon as the hand collides with the VA's shoulder.

The ***Open Hands*** ($\mathcal{P} = (Forward, Upright, X00)$) gesture begins like the *Shake Hands* interaction, performing the same initial steps to activate the same interaction pose. When the fingers of both the user's and the VA's right hands overlap, the system stores the hand's position and counts the overlaps. As soon as the fingers stop overlapping, the system queries the current position again to calculate a displacement vector to identify the direction of the hand's movement. The first overlap is expected to correspond to a movement to the left, while the second overlap is expected to occur from the right. At every step of a beginning or ending overlap, the system also checks for an open hand in the *Upright* orientation. If two overlaps occur in this way, the gesture is recognized and the interaction pose ends. The progress of this movement tracking is reset if no change in overlaps occurs for 1.3 seconds.

To allow the user to get feedback about what gestures of the sequence have currently been recognized by the system, we added a display on the VA's chest (see Fig. 1 d)).

For the system to know if a recognized gesture sequence is correct, we assign each gesture a unique number. A correct password then consists of a sequence of numbers, paired with a string that uniquely identifies the VA. When a gesture is recognized, the system compares the currently targeted VA with the correct target specified in the password. If the target matches, the number representing the gesture is appended to the currently recorded password sequence, audio feedback is given and an icon representing the gesture is appended to the display on the VA's chest. Each input immediately triggers a check if the currently recorded password sequence is correct. If the recorded sequence matches the password, the system will acoustically inform the user of the success and clear all recorded inputs. Otherwise, this process remains hidden from the user. If the user makes a false input, the clear gesture can be used to delete the current input.

## 4 Evaluation

Motivated by the prototypical design of our application and following an exploratory approach towards natural authentication, we conducted an expert review to gather first insights. The review followed a within-subjects design and randomized order of conditions to avoid biases. Besides expert feedback, we measured entry times and error rates for password entry and individual gestures using a repeated measures design.

### 4.1 Apparatus and Participants

Five VR experts were chosen for the formal expert review (2 female, 3 male, age: $M = 32.40, \sigma = 2.50$). Each possesses over five years of dedicated research experience in computer science and VR, with a specialized focus on graphical user interfaces (n=3) and interactions with VAs (n=2). The expert review was conducted using an HTC Vive Pro 2 with Valve Index controllers. While we did not use the controller's finger tracking capabilities, we still chose to use these controllers as they attach to the user's hand via a strap allowing participants to fully open their hand, which is more natural for performing gestures. The headset was attached by cable to a PC running Windows 10, which was equipped with an Intel i9-10900X CPU, 32 GB of RAM and an Nvidia RTX 3090 GPU. The square tracking area had a size of around $4m^2$.

### 4.2 Task Design

The expert review consists of two conditions. In $C_{Gesture}$, the gesture authentication method described in Section 3 was used. Gesture sequences had a length of four to five, as this is a common length for PINs used on mobile phones. We created a group of four VAs, two with male and female appearance each, using Unreal's *MetaHuman Creator*[1] and arranged them in an open group formation (see Fig. 1 a)). In the second condition $C_{PIN}$, a virtual keypad was used to enter a numerical PIN of equal length. Instead of four VAs, we arranged four color-coded keypads in a similar arrangement at a different section in the virtual environment (see Fig. 1 e)). Instead of a target VA and a gesture sequence, a password in $C_{PIN}$ consists of a target keypad and a sequence of numbers. To ensure a natural interaction and obtain sufficient, unbiased data for all gestures, we generated ten passwords in advance by generating true-random number sequences on RANDOM.ORG, re-picking sequences until we obtained five passwords of length 4 and 5 each, all digits appeared either four or five times and no more than twice in one password. We designated the numeric passwords as PINs for condition $C_{PIN}$ and mapped the digits to gestures in the order they are defined to obtain gesture sequences for condition $C_{Gesture}$. Furthermore, we generated ten sets containing a single number between 0 and 3 until each number appeared either two or three times. These digits map to the colors of the keypads (red, yellow, green, blue) or the VAs (Kris, Vivi, Aoi, Kioko) respectively. The generated targets were assigned to the generated passwords in the order they were obtained. All participants performed the same passwords from our true-random set. The order in which the sequences and PINs were presented was pseudo-randomized at runtime and different for each participant. The order in which participants performed each condition was counterbalanced.

### 4.3 Procedure

After signing the informed consent, participants were introduced to the VR headset, the controllers, as well as available navigation methods (teleportation, physical turning only). After an initial testing phase to get familiar with the controls and the virtual environment, participants read the task description.

Each condition consisted of three phases: In the *learning phase* of $C_{Gesture}$, each gesture was performed twice with a single VA, while in $C_{PIN}$, two example PINs were entered at a single virtual keypad. Afterward, the part of the study to be analyzed began, where for each password to be performed or entered, the following two phases were iteratively performed: In the *memorization phase*, a password (gesture sequence with icons for $C_{Gesture}$, numbers for $C_{PIN}$) was presented on the whiteboard and had to be accurately performed

[1] MetaHuman Creator: https://metahuman.unrealengine.com/

or entered twice. Subsequently, the password was concealed on the whiteboard, marking the onset of the *blind phase*. This phase aimed to simulate a realistic scenario for measuring entry time and memorability. Removing the password from the whiteboard forced participants to recall it and perform or enter the password from memory, enhancing comparability between both conditions. To this end, participants were explicitly instructed to memorize the password before advancing to this phase. The memorization and blind phase were repeated for ten different passwords of lengths 4 to 5 in both conditions.

During the blind phase, we logged entry time, defined as the interval from the onset of the phase to the moment the password was correctly recognized. The blind phase commenced immediately following the accurate input of the last password in the memorization phase. We only measured entry time during the blind phase, since memorizing the passwords during the memorization phase affects entry time.

After each experimental condition, participants engaged in a semi-structured interview. This interview format was selected to gain first insights into the participants' perceptions of the gesture authentication process's applicability. During the interview, participants were asked about their experiences with the authentication process. For $C_{Gesture}$, we specifically explored their thoughts on the gestures, the display on the chest, and their impressions of the Meta Humans' appearance. The interview also included comparative questions after participants experienced both conditions. The final segment of the interview gathered demographic information from the participants, thereby concluding the expert review. The expert review took about one hour to complete and participants spent about 25 minutes immersed in VR.

## 5 Results

In this section, we will present the results of the logged data and the semi-structured interview.

### 5.1 Logged Data

With five participants supplying ten passwords each, we gathered a dataset of $N = 50$ data points per condition. Figure 2 (a) (orange) $(N = 50, M = 2.82, \sigma = 0.91)$ shows entry time in seconds for $C_{PIN}$. For $C_{Gesture}$, nine data points had to be removed due to gestures not being recognized (7x *Wave*, 2x *Bow*). Figure 2 (a) (blue) $(N = 41, M = 23.67, \sigma = 20.62)$, shows the resulting boxplot (G) illustrating entry time in seconds for $C_{Gesture}$. Furthermore, Figure 2 (b) shows boxplots illustrating entry time in seconds after the removal of further data points. For the orange boxplot ($G_{woError}$) $(N = 31, M = 18.73, \sigma = 17.90)$, we removed ten additional data points. These points correspond to instances where participants made errors during password entry, requiring the use of the *clear* gesture to restart the password entry. These excluded data points, thus, represent incorrectly performed or recognized gestures, else falsifying the entry time. For the green boxplot ($G_{woOutlier}$) $(N = 29, M = 14.29, \sigma = 5.42)$, additionally two large outliers were removed due to extended gesture recognition time (1x *Wave* 93.95 s, 1x *Bow* 72.14 s). There were no recognition issues or input errors made in $C_{PIN}$. A Welch's t-test for unequal variances was used to compare the remaining 29 $G_{woOutlier}$ data points from $C_{Gesture}$ to the 50 data points of $C_{PIN}$ (see Fig. 2 a)). The analysis revealed a significantly lower entry time in $C_{PIN}$ $(t(28.90) = 11.11, p < 0.001)$. Prior to Welch's t-test, normality assumptions were assessed and confirmed using a Shapiro-Wilk test.

We also calculated the entry time in seconds for individual gestures (see Table 2), measuring the time difference between each recognized gesture.

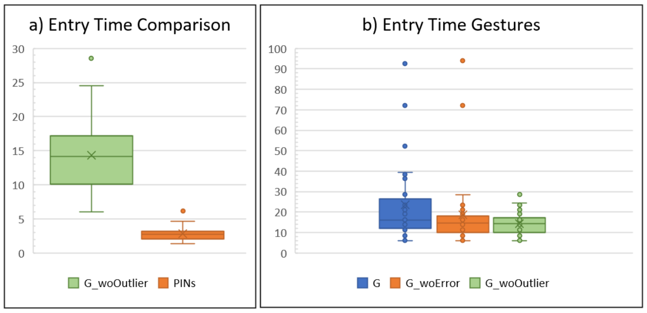

Figure 2: a) Boxplots comparing entry time in seconds between the remaining data points of $C_{Gesture}$ and $C_{PIN}$. b) Boxplots illustrating entry time in seconds of $C_{Gesture}$ for data points without recognition errors (G), input errors ($G_{woError}$) and large outliers ($G_{woOutlier}$).

| Gesture | M | $\sigma$ |
|---|---|---|
| | 13.02 | 9.71 |
| | 2.94 | 1.44 |
| | 3.40 | 1.96 |
| | 4.45 | 2.29 |
| | 5.61 | 6.40 |

| Gesture | M | $\sigma$ |
|---|---|---|
| | 3.11 | 0.97 |
| | 2.87 | 0.92 |
| | 4.40 | 2.49 |
| | 2.13 | 0.33 |
| | 4.59 | 0.33 |

Table 2: Mean entry time with standard deviation in seconds for each gesture.

### 5.2 Interview

In this section, the results of the semi-structured interview are presented in the order in which the questions were prompted. When asked about their **overall impression of the gesture authentication process**, all participants reported that it was fun to use. One noted that it felt algorithmic and two criticized the efficiency and higher mental workload compared to $C_{PIN}$. Furthermore, we asked more specifically about the following aspects: (i) The *general applicability* of $C_{Gesture}$ was unanimously acknowledged by all participants. One participant rates it more immersive than $C_{PIN}$, but also expressed concerns about its unconventional approach sharing (possibly secret) passwords with another (virtual) human. Another participant noted feeling safe and comfortable sharing personal information this way, only when it is evident that the virtual interaction partner was system-controlled, attributing a sense of safety to this gesture authentication process; (ii) Three participants agreed that the *gestures resembled a greeting*, while two others mentioned that their perception varied depending on the specific gesture used; (iii) The *memorability of the gestures* posed a challenge for two participants, primarily due to increased mental workload and occasional distractions caused by technical issues with the gesture recognition process. In contrast, two other participants reported that they used memorization techniques such as associating gestures with letters and therefore had fewer problems; (iv) The *usefulness of the self-avatar for gesture input* was unanimously agreed on. More specifically three participants noted that hands would have been enough, while one participant preferred to have both, hands and arms.

Considering **individual gestures**, we asked participants about their most and least favorite gestures. *Fist Bump* was stated twice as favorite gesture, while *Fist on Chest*, *Finger Guns* and *Pat on Shoulder* were stated once each. Furthermore, the *Thumbs-Up* and *High Five* gestures were highlighted additionally by two participants.

The majority of participants favored these gestures due to their efficient detection capabilities and swift execution. All participants criticized *Wave* due to recognition problems and lengthy execution time, and three also stated it as their least favorite. *Bow* received negative feedback for the same reason by three participants, while two explicitly mentioned it as their least favorite gesture. When asked which gestures they would like to add to the gesture set, two participants wanted more intricate gestures like forming shapes with their fingers. However, they also acknowledged the potential difficulty in accurately tracking these gestures. One participant suggested adding two-handed gestures, while another one proposed additional self-body gestures like touching one's own head or shoulders.

When asked about the **visual appearance of the VAs** all participants felt that the VAs looked visually pleasing. Furthermore, we asked more specifically about the following aspects: (i) All VAs were voted at least once as *the user's favorite VA*, except the leftmost in Figure 1; (ii) The *VA's demeanor conveyed a welcoming attitude* to all participants; (iii) Three participants affirmed *feeling like a part of the group*, whereas two participants expressed disagreement. They reasoned that they had interacted solely with a single VA and positioned themselves more toward the center of the group, which affected their sense of inclusion; (iv) Three participants reported that *the movements of the VAs were helpful*, while one participant wished for feedback about the hand regions. Another one was distracted due to slow or incorrect reaction times for *Wave* and *High Five*. (v) The *display on the VA's chest* was liked by all. One participant reported that technically audio feedback would have been enough.

Although not rated as fun, all participants liked the implementation of the **virtual keypad** as a means of entering number-based PINs, when asked about their general impressions. They all agreed that it was efficient and straightforward to use. Furthermore, when asked about memorability, not a single participant encountered difficulty in remembering their respective PINs.

After both conditions, we asked participants **comparative questions** about $C_{Gesture}$ and $C_{PIN}$. When asked about their overall preferred method, all participants favored $C_{PIN}$ due to the higher efficiency. When asked to name an advantage of gestures-based authentication, participants stated that it felt more innovative, fun, and interactive. In terms of addressing shoulder surfing and safety concerns, three participants expressed that $C_{PIN}$ felt safer due to them not relying on body movements. However, one participant favored $C_{Gesture}$, citing its dissimilarity to traditional password entry as a key advantage. Notably, one participant did not express a preference for either method. When asked about password length for $C_{PIN}$, three participants rated the length as appropriate, while two participants preferred longer PINs, with a maximum length of 7 digits, aiming to strike a balance between security and ease of memorization. For $C_{Gesture}$ all participants rated the length of four to five to be an adequate balance between security and memorability.

The **virtual environment** was rated as visually pleasing by all participants and as suitable for the expert review.

## 6 Discussion, Limitations and Future Work

In this section, we will discuss the results presented in the previous chapter and identify limitations and opportunities for future research.

We were not surprised that entry time was significantly lower in $C_{PIN}$ compared to $C_{Gesture}$. Performing social gestures requires significantly more movements, therefore limiting the minimally achievable entry time. Furthermore, our initial implementation of the gesture recognizer encountered difficulties in accurately identifying certain gestures, as exemplified in Section 5.2. Thus, it required users to repeat the gestures multiple times, albeit not logging the different repetitions, before achieving a correct recognition, leading to an increased entry time. Nevertheless, the fastest gesture sequence was recognized in 6.06 seconds (2x *Pat on Shoulder*, *Fist on Chest*, *Thumbs-Up*), indicating that gesture sequences that follow a natural flow can be performed faster. In a realistic scenario, users would create their own passwords, potentially optimizing flow, and would have more practice than in our expert review, lowering entry time.

While *Wave* and *Bow* had the most recognition problems, also indicated by the large standard deviations, *Thumbs-Up*, *Shake Hands* and *Pat on Shoulder* had entry times of less than 3 seconds each. Together with good recognition, indicated by low standard deviations, these results suggest that a more fine-tuned less movement-intensive gesture set could lower entry time even further, thus increasing the practicality of gesture-based authentication. Moreover, it is crucial to consider cultural differences when establishing a new or extended gesture set, given that certain gestures may carry offensive meanings in specific cultures [14].

Based on the results of the semi-structured interview, overall $C_{PIN}$ was favored over $C_{Gesture}$ for most factors. However, gesture authentication was still considered promising by experts. Future research should focus on improving entry time, and memorability by evaluating a more realistic scenario with personalized, user-created passwords, considering more optimized, less-movement-intensive gesture sets that promote gesture flow and therefore further increase efficiency.

## 7 Conclusion

In our paper, we presented a novel authentication method and provided first insights into the feasibility of such a system. While our implementation has its limitations and the small sample size of our expert review does not allow to generalize our results, we believe that our work shows that VA-mediated gesture-based authentication can be a viable alternative to common authentication methods like PINs and should be explored further in the future.

## References

[1] F. Aloul, S. Zahidi, and W. El-Hajj. Two Factor Authentication Using Mobile Phones. In *2009 IEEE/ACS International Conference on Computer Systems and Applications*, pp. 641–644, 2009. doi: 10.1109/AICCSA.2009.5069395

[2] F. Boucaud, C. Pelachaud, and I. Thouvenin. Decision Model for a Virtual Agent that can Touch and be Touched. In *20th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2021)*, pp. 232–241. Londres (virtual), United Kingdom, May 2021.

[3] A. Bönsch, S. Radke, H. Overath, L. M. Asché, J. Wendt, T. Vierjahn, U. Habel, and T. W. Kuhlen. Social VR: How Personal Space is Affected by Virtual Agents' Emotions. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 199–206, 2018. doi: 10.1109/VR.2018.8446480

[4] T. J. Dube and A. S. Arif. Text Entry in Virtual Reality: A Comprehensive Review of the Literature. In M. Kurosu, ed., *Human-Computer Interaction. Recognition and Interaction Technologies*, pp. 419–437. Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-030-22643-5_33

[5] J. Ehret, A. Bönsch, P. Nossol, C. A. Ermert, C. Mohanathasan, S. J. Schlittmeier, J. Fels, and T. W. Kuhlen. Who's next? Integrating Non-Verbal Turn-Taking Cues for Embodied Conversational Agents. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, IVA '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3570945.3607312

[6] C. George, D. Buschek, A. Ngao, and M. Khamis. GazeRoomLock: Using Gaze and Head-Pose to Improve the Usability and Observation Resistance of 3D Passwords in Virtual Reality. In L. T. De Paolis and P. Bourdot, eds., *Augmented Reality, Virtual Reality, and Computer Graphics*, pp. 61–81. Springer International Publishing, Cham, 2020. doi: 10.1007/978-3-030-58465-8_5

[7] C. George, M. Khamis, D. Buschek, and H. Hussmann. 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)*, pp. 277–285, 2019. doi: 10.1109/VR.2019.8797862

[8] C. George, M. Khamis, E. von Zezschwitz, M. Burger, H. Schmidt, F. Alt, and H. Hussmann. Seamless and Secure VR: Adapting and Evaluating Established Authentication Systems for Virtual Reality. NDSS, San Diego, CA, USA, Feb. 2017. doi: 10.14722/usec.2017.23028

[9] G. Huisman, J. Kolkmeier, and D. Heylen. With Us or Against Us: Simulated Social Touch by Virtual Agents in a Cooperative or Competitive Setting. In *Intelligent Virtual Agents: 14th International Conference, IVA 2014, Boston, MA, USA, August 27-29, 2014. Proceedings 14*, pp. 204–213. Springer, 2014. doi: 10.1007/978-3-319-09767-1_25

[10] J. M. Jones, R. Duezguen, P. Mayer, M. Volkamer, and S. Das. A Literature Review on Virtual Reality Authentication. In S. Furnell and N. Clarke, eds., *Human Aspects of Information Security and Assurance*, pp. 189–198. Springer International Publishing, Cham, 2021. doi: 10.1007/978-3-030-81111-2_16

[11] F. Mathis, H. I. Fawaz, and M. Khamis. Knowledge-Driven Biometric Authentication in Virtual Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, p. 1–10. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3334480.3382799

[12] F. Mathis, J. Williamson, K. Vaniea, and M. Khamis. RubikAuth: Fast and Secure Authentication in Virtual Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, p. 1–9. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3334480.3382827

[13] C. S. Oh, J. N. Bailenson, and G. F. Welch. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI*, 5, 2018. doi: 10.3389/frobt.2018.00114

[14] L. Purnell. *Cross Cultural Communication: Verbal and Non-Verbal Communication, Interpretation and Translation*, pp. 131–142. Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-319-69332-3_14

[15] M. Rehm, E. André, and M. Nischt. Let's come together—social navigation behaviors of virtual and real humans. In *Intelligent Technologies for Interactive Entertainment: First International Conference, INTETAIN 2005, Madonna di Campiglio, Italy, November 30–December 2, 2005. Proceedings 1*, pp. 124–133. Springer, 2005. doi: 10.1007/11590323_13

[16] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell. A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. In *Computer graphics forum*, vol. 34, pp. 299–326. Wiley Online Library, 2015. doi: 10.1111/cgf.12603