CoordAuth: A Two-factor Authentication Method in Virtual Reality Leveraging Head-Eye Coordination

ANONYMOUS AUTHOR(S)*

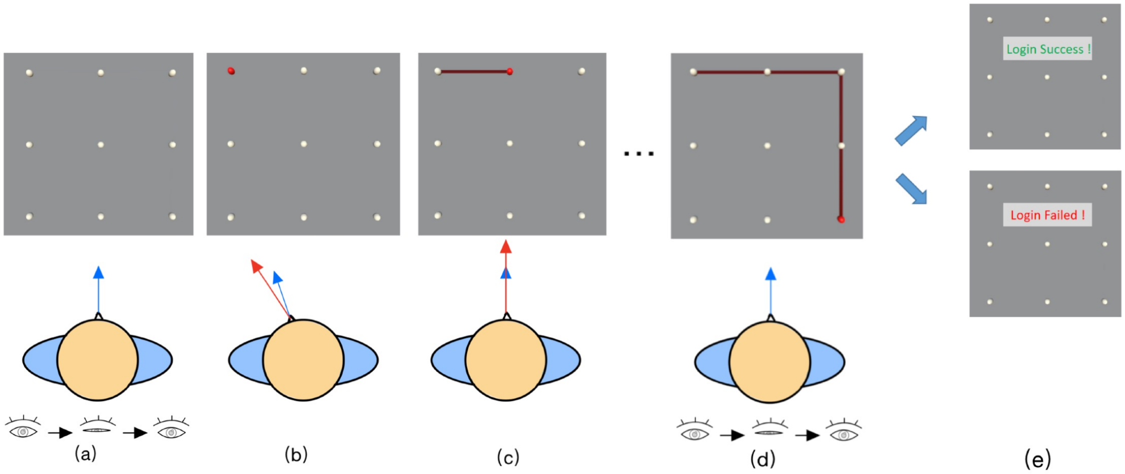


Fig. 1. A storyboard illustrating an authentication attempt saccading "12369" with CoordAuth. (a) The interface consisted of 3X3 grid pattern with a 60° FOV. The user blinks and reopens the eyes over 0.3s to start the authentication attempt. (b) To input "1", the user rotates the head and eyes, coordinating both to focus their gaze on the hit test region (as Figure 3(a) shown). The blue arrow represents the direction of the user's head, and the red arrow represents the direction of eye gaze. (c) Then the user simultaneously rotate their head and eyes, guiding their gaze **across** the center point of the first row to input "2" (as Figure 3(b) shown. (d) The user completed the pattern and used the blink and repoening of the eyes to terminate the authentication as (a). (e) The system assesses the success or failure of the login, considering both the entered pattern and the behavioral biometrics associated with head-eye coordination movements.

With the advance of virtual reality, there is an increasing appeal to protect users' privacy and security via authentication. However, existing knowledge-based authentication is lengthy and biometric authentication may potentially reveal users' data. Hence, we proposed CoordAuth, a two-factor authentication technology leveraging head-eye coordination features. To facilitate CoordAuth's design, we first conducted a study to analyze the effect of grid size (40° – 100° FOV) on users' input performance. We found the medium-small size (60°) was the most suitable setting and identified unique head-eye coordination features during input. Leveraging these features, CoordAuth combined pattern-based classifiers and majority voting-based behavioral biometric classifiers for two-factor authentication, achieving 0.04% False Acceptance Rate (FAR) and 0.88% False Rejection Rate (FRR) during leave-one-out simulation.

Manuscript submitted to ACM

CoordAuth also exhibited longitudinal stability with a 0.32% FAR and 2.73% FRR across 7 days. The subsequent usability and shoulder-surfing attack study proved CoordAuth's usability and robustness, where CoordAuth achieved 3.82s authentication time, 2.50% Error Rate, and 0.60% Attack Success Rate (ASR) comparable to knowledge-based and behavioral-biometric-based baselines.

CCS Concepts: • **Human-centered computing → Interaction techniques;Virtual reality**; • **Security and privacy;Two factor authentication**;

Additional Key Words and Phrases: Authentication, Behavioral biometrics, Gaze interation, Fixation and saccade, Virtual Reality

ACM Reference Format:

Anonymous Author(s). 2018. CoordAuth: A Two-factor Authentication Method in Virtual Reality Leveraging Head-Eye Coordination. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 29 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

With the rapid development of VR technology, VR is finding increasingly diverse applications in fields such as e-commerce [10, 50, 63], social networking [27, 70], healthcare [84], and education [5, 29]. As VR devices become more widespread across various domains, the challenge of developing secure and user-friendly authentication technology within the VR environment is becoming increasingly vital [35].

Existing widely used authentication methods could be classified into knowledge-based, possession-based and property-based methods. Knowledge-based methods included alphanumeric passwords (including visual passwords [79], 3D visual passwords [21, 23], novel passwords [53, 54, 94]), PINs [24, 86, 93] and patterns [24]. Typical possession-based methods included using ears (e.g., EarEcho [22]), iris[1] or other features (e.g., ElectricalAuth [12]) for authentication. However, knowledge-based methods faced a higher probability of being shoulder-surfed and a lengthy authentication time, whereas possession-based methods were prone to leak personally identifiable information, which resulted in privacy concerns. Traditional property-based methods used other pairing devices or physical keys for authentication, which was lengthy, inconvenient and vulnerable.

To address these limitations, some researchers proposed behavioral biometrics-based authentication, which leveraged sensors to collect users' behavioral data (e.g., eye movement [44], head motion [85], full-body motion [65] and hand gesture [80]). Upon behavioral-biometrics features, some constructed two-factor authentication such as GlassGesture [90], RubicBiom [51] and Blinkey [95]. However, these methods faced the problem of unstable features (e.g., full-body motion[65] and hand gesture [80]) and additional devices' requirements (e.g., controllers [51], pupil trackers [95]).

Hence, we propose CoordAuth, a two-factor authentication technology leveraging the correlation of head-eye coordination. Through entering the 3X3 pattern password in VR, CoordAuth captured the correctness of pattern as the first authentication factor and the biometric biometrics features of head and eye fixation as well as saccade as the second authentication factor.

To concretize CoordAuth, we first conducted a study to investigate users' head-eye coordinated movement pattern. We also aimed to determine the UI design of CoordAuth. Through setting 4 grid sizes (40° – 100°), we concluded medium-small size (60°) was the most suitable regarding input speed, error rate and subjective ratings.

CoordAuth leveraged the motor features (e.g., velocity, acceleration) of head-eye coordination elicited by glancing at different patterns to facilitate two-factor authentication. The first factor consisted of comparing the drawn pattern and ground-truth pattern using the built-in tracking system of gaze interaction. The second factor designed a voting

[1] https://learn.microsoft.com/en-us/hololens/hololens-identity

mechanism combining different features (e.g., position and rotation) and recognizers (e.g., fixation recognizer and saccade recognizer) for authentication. Ensembling was proved effective in CoordAuth's design through experiments and using Random Forest as the base classifier, CoordAuth reached a 0.88% FRR and a 0.04% FAR.

To prove the stability and robustness of CoordAuth, we conducted the longitudinal study and tested the case where pattern collides. The former case showed an FRR of 12.6% and an FAR of 2.6% and an FRR of 2.7% and an FAR of 0.3% with updating strategy. While the latter case resulted in an FRR of 0.88% and an FAR of 0.042%, proving the effectiveness of behavioral biometrics features.

We further examined the usability of CoordAuth and its robustness to shoulder surfing through two real-life studies. CoordAuth achieved 3.80s authentication time, 2.50% Error Rate, and 0.60% Attack Success Rate (ASR) comparable to knowledge-based and behavioral-biometric-based baselines. Besides, through letting users authenticate in standing and sitting positions, we found standing resulted in a even faster speed, proving the advantage of CoordAuth in real-world scenarios.

To sum up, the main contributions of our work are three-folded:

- To the best of our knowledge, we are the first to exclusively utilize gaze interaction for pattern unlocking in VR. Specifically, we derived a set of reference points and parameters for hit tests based on previous research. We conducted a user study to assess the impact of FOV on gaze-based input. The design extends beyond authentication, making it widely applicable to various input scenarios.
- We present CoordAuth, a two-factor authentication methodology based on gaze input for pattern passwords with head-eye coordinated movements. We conducted feature engineering based on the two mentioned behavioral biometric features, and achieved a 0.045% FRR and a near-zero FAR using an ensemble learning model that integrates saccade classifiers and fixation classifiers with a majority voting mechanism.
- Even in the presence of password collision, CoordAuth maintained an FRR of 0.88% and an FAR of 0.04%. It demonstrates stability over time, robustness to contextual factors, a lower error rate compared to traditional patterns, and an ASR as low as 0.6% in shoulder-surfing attacks. Additionally, all sensors used in CoordAuth are readily available in contemporary VR headsets, facilitating its adoption without the need for additional hardware.

2 RELATED WORK

2.1 Exploring the Correlation Between Head and Eye to Interact in Virtual Reality

Eye movement has been recognized as a natural means of interaction in the field of Human-Computer Interaction, as individuals naturally gaze at objects they are attending to [49]. Occasionally, unconscious gaze behaviors may occur during gaze input, leading to the Midas Touch problem [34]. Numerous studies have focused on developing interaction strategies to prevent the Midas Touch problem, such as dwell time [58, 77], simulating a single-button mouse [14] or a proper hit test region [26, 58, 91]. Furthermore, there are also some interaction techniques based on head movements. For example, technologies like HeadGesture [88, 89] allow users to interact with devices through head movements, while others detect users' interaction intentions through head movements [87].

Currently, there is a lot of research comparing gaze input with head rotation input, as observed in augmented reality and virtual reality head-mounted displays (HMDs) [6, 31, 39, 66]. Studies have found that gaze-only techniques are faster and more accurate compared to head-only interaction [39, 66], but using only head movements can be overall more fatigued [66]. However, none of them used the concept as a feature for authentication.

Combining eye tracking with head rotation is a natural and comfortable interaction method [58]. For example, some technologies enable navigation in virtual environments without the need for hand movement involvement [61, 62, 67]. Other research explores different combinations of eye movements and head movements for text input [17] or object selection. Overall, combining eye tracking with head rotation can address calibration drift [67] and result in improved interaction performance [39], which served as the inspiration for our design and technique.

2.2 Head-Eye Coordination Modeling

The model of head-eye coordination has been present in biological research [18] and has been extensively studied in the field of HCI in recent years. In VR, this model refers to the coordinated collaboration between the eyes and the head during gaze shifts [75]. When individuals solely use their eyes for observing without moving their heads, the Field of View is limited, with an approximate eye movement range of 50° [41]. VR enables the exploration of virtual scenes beyond the constraints of Head-Mounted Display visibility, extending into a broader field of view. People often expand their field of view by rotating their heads to gather more comprehensive information. Therefore, studying the model of head-eye coordination becomes crucial in the VR domain to more naturally simulate human visual behavior [75].

The relationship between eye and head movements is complex. During gaze shifts, the combined movement of the head and eyes forms the gaze in the world coordinate system. The faster the synchronous movement of the head and eyes during gaze shifts, the smaller the required eye movement [18]. When the gaze target is reached, the head typically continues to move, while the eyes fixate on the target by executing compensatory eye movements in the opposite direction. This allows the eyes to rotate back to a more central and comfortable position relative to the head [83]. The current research has demonstrated that the magnitude of gaze shifts significantly influences head movement. Studies indicate that gaze shifts smaller than 20° are primarily executed through eye movements [18, 25, 41]. Despite the eyes having a physical range of 50°, research has found that they rarely rotate more than 30° relative to the head [41]. In natural tasks, the head contributes to approximately one-third of gaze shifts below 30° and even more for larger gaze shifts [71]. Factors such as initial eye position, amplitude, and the location of the next target have been identified to influence the decision to move the head [19, 20, 60]. Furthermore, individual differences in the degree of using head movements have been observed [20, 60, 78, 81, 82]. Fuller introduced the concept of "head movers" and "non-head movers" [20], while other studies indicated that these differences in head movement tendencies exist in both controlled and real-world environments [81]. These individual differences are particularly significant for us. Specifically, we conducted a study into the users' head involvement when entering pattern passwords within a specific Field of View (FOV). This also provides a solid explanatory foundation for the selection of features in CoordAuth's behavior biometric factor.

2.3 Behavioural Biometric Authentication in Virtual Reality

In virtual reality, high availability and accurate authentication technologies are crucial. Research in the VR domain focuses on developing novel and VR-specific authentication technologies that fall into five major categories [79]: knowledge-based [21, 23, 24, 36, 40, 53, 54], physical biometric [7, 22, 37, 72], behavior biometric [3, 11, 38, 42, 56, 57, 59, 64, 68, 74, 76], token-based [9], and multi-factor [51, 90, 95].

Among these, behavior biometrics have garnered widespread attention due to lower privacy concerns and their ability to resist shoulder surfing. These biometric features leverage unique individual behavioral characteristics such as motion, coordination, and body dynamics [30]. There is a considerable body of research on various aspects of behavior biometrics, including head movement [3, 42, 45, 52, 59, 65, 68, 69], eye movement [46–48, 59], hand gestures [3, 38, 56, 59, 64], blink patterns [69, 95], and full-body motion [55, 59, 74].

It is noteworthy that while some studies use both head and eye movement features, it is more common to focus on a single feature [42, 46–48, 59, 68]. Although [59, 69] simultaneously employed both, but did not delve into the characteristics of head-eye coordination movement. Additionally, currently widely adopted behavioral biometric verification methods typically involve the direct use of machine learning models or proprietary mathematical models for classification, lacking reasonable interpretability. Our head-eye coordination model provides theoretical support for behavioral biometric features in virtual reality and offers explanatory insights.

Behavioral biometric features are often employed as one of the factors in two-factor authentication systems. For instance, the specific changes in pupil size under Blinkey's [95] distinctive blink pattern, hand movement features during interaction with RubicBiom's [51] knowledge-based authentication system and head movement features in GlassGesture [90] when inputting specific head gestures can serve as auxiliary authentication factors. Our proposed model offers higher input speed, lower error rates, and enhanced security.

3 INTERACTION DESIGN OF COORDAUTH

We developed the CoordAuth's virtual unlocking interface using Unity and deployed it on the Meta Quest Pro headset, which is equipped with built-in cameras and eye trackers, as shown in Figure 2. In practical applications, CoordAuth can also be deployed on other devices equipped with eye trackers, such as HTC Vive Pro Eye [2] and Pico Neo [3].

Fig. 2. Experiment apparatus

3.1 UI Design

Our design aims for minimal, comfortable gaze interaction in VR pattern entry. We adopted a widely recognized 3x3 grid pattern lock design in the VR environment, inspired by smartphone interfaces. Upon entering the VR authentication interface, users will encounter a vertically arranged 3x3 grid. Initially, each point on the grid is represented by a white sphere. When a user gazes at a point, it transits to red, serving as the feedback to indicate selection. A red line is created when the user draws a pattern from one point to another. Each grid point can only be selected once per pattern password entry. Selecting two endpoints on a straight line is equivalent to selecting all three points.

In our pilot study, we determined the size of the points. While larger grid points make it easier for the gaze point to fall into the grid point area, the point size cannot be too large, as oversized objects in VR may induce psychological

[2] https://www.vive.com/eu/product/vive-pro2-full-kit/overview/

[3] https://business.picoxr.com/us

pressure on users [26]. Combining insights from the pilot study and previous work [58], we established an appropriate point size with a FOV of 3°.

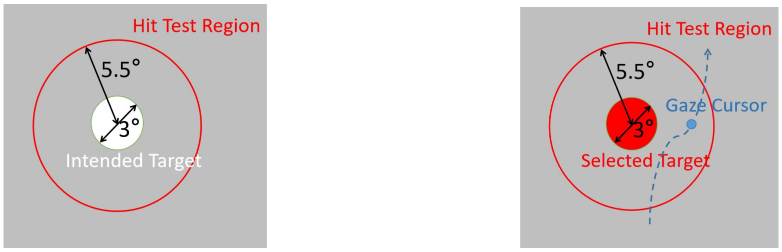
3.2 Input Triggering

Our design aims to keep the user’s interaction as simple as possible without requiring additional hand movements, so we used non-spontaneous eye blinks to trigger the start and end of input. We defined one complete closure of the upper and lower eyelids as one blink. In our pilot study, participants were instructed to intentionally close their eyes to initiate the input. We found that non-spontaneous eye blinks by users typically lasted for over 0.35 seconds, with the longest duration exceeding 1 second. Previous research also indicated that spontaneous eye blinks in humans last approximately 0.1-0.15 seconds [8]. Therefore, we established the following criteria for the start and end of input: complete closure of the upper and lower eyelids within the past 0.3 seconds, i.e., non-spontaneous blink, to trigger the beginning and end of input. This effectively prevents accidental triggering of input due to spontaneous blinking.

3.3 Hit Test Detection

We used an enlarged area with goal-crossing strategy for hit-test detection. Due to the presence of the "Midas touch" issue [26, 34] in eye gaze interactions, it is crucial to design appropriate selection methods and determine the size of the detection area. For eye gaze object selection in VR environments, dwelling and goal crossing are two main approaches. We aimed for an efficient and swift interaction, hence we opted for goal crossing, where users only need to have their gaze point fall within the hit-test area for a single frame [58].

In our pilot study, we tested hit test regions of a ±4°, ±5.5°, ±7° and ±8.5° FOV relative to the object’s center. The results indicated that a ±5.5 size yielded a sufficiently low error rate. This finding contrasted with some prior studies [58, 75], where our results were larger. However, considering gaze drift in eye-tracking devices [17], we found this to be reasonable. The specific parameters and interaction details are illustrated in Figure 3.



(a) (b)

Fig. 3. (a) Parameters for the hit test region. (b) Selection of targets based on goal crossing.

4 STUDY 1: EXAMINING THE EFFECT OF GRID SIZE ON USERS’ AUTHENTICATION BEHAVIOUR

We conducted a user experiment to determine the optimal size based on performance and subject feedback during authentication tasks. We also aimed to determine the behavioral patterns and features for authentication algorithm.

4.1 Participants and Apparatus

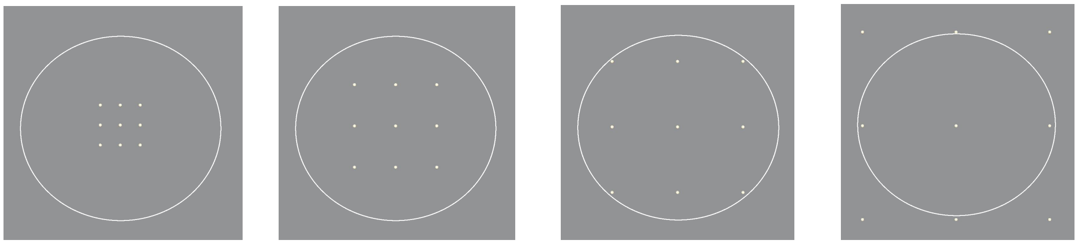
We recruited 16 participants from the campus (7 females, 9 males, average age=23.8, SD=2.1). All participants had normal vision or corrected vision. The average reported VR usage experience was 2.6 (SD=0.56, out of a maximum score of 5). We also asked participants whether they had used VR authentication techniques before, and 2 of them reported prior experience with VR authentication. Each participant received a payment of $15. The study had been reviewed and approved by the Ethics Committee of our institute.

We used a Meta Quest Pro headset as the apparatus. The headset had a refresh rate of 72Hz and a resolution of 1800 × 1920 per eye, with 106° horizontal, 96° vertical claimed field of view. In experiments, we used eye tracking with high tracking frequency offered by the Oculus Integration SDK [4]. The experimental platform was developed in C# using Unity 3D 2021.3.21f1c1. The coordination data of eye gaze and head was recorded for each frame.

4.2 Study Design

To simulate actual pattern entry, we instructed participants to input 20 pre-designed patterns [13] with different grid sizes. These 20 patterns are among the most commonly used in Android, so we deemed them sufficiently representative. We asked participants to input these patterns as quickly and accurately as possible. We employed a within-subjects design, with **grid size** being the only factor.

When the FOV reaches 40-50 degrees, head movement emerges [18, 41], and at this point, head-eye coordinated movement occurs. Due to our desire to better leverage the features of head-eye coordination, we chose a 40° FOV as the minimum grid size. For similar studies [75] involving the modulation of head-eye coordination for gaze selection, it chose to use intervals of 20° FOV each. Thus we selected four different grid layout sizes as shown in Figure 4, with FOV 40°, 60°, 80°, 100°, which are small, medium small, medium large and large.



(1)Small: 40° FOV. (2)Medium small: 60° FOV. (3)Medium large: 80° FOV. (4)Large: 100° FOV.

Fig. 4. Grid layouts of different sizes, with the white ellipse representing the FOV of the head-mounted display.

4.3 Procedure

We initially introduced participants to the experiment platform and usage of CoordAuth, allowing them 5 minutes for practice. Then the main experiment was divided into four sessions differed by the grid size (see Section 4.2) with a counter-balanced order (through Latin Square design). For each session, participants would underwent 20 patterns' instruction and input in a random order using shuffle function of Python. For each pattern, we first showed the drawing of the pattern using green lines and points for one time. After the showing disappeared, users needed to repeat the pattern by themselves. They started through a non-spontaneous blink, drawed the pattern and ended also with a

[4] https://developer.oculus.com/downloads/package/unity-integration/

non-spontaneous blink. The pattern drawn was rendered in red. We showed no final results to users regarding whether the pattern was correct. After the experiment, we collected participants' subjective feedback through questionnaire and interviews.

4.4 Results

In total, we conducted 16 participants x 4 sizes x 20 PINs = 1280 trials from all participants. We used Repeated Measures Analysis of Variance (RM-ANOVA) parametric or non-parametric tests and corresponding post-hoc tests for statistical analysis.

4.4.1 Input time. The input time refers to the average duration required for users to draw each pattern and is measured using the timestamp of the last non-spontaneous blink minus the timestamp of the first. We included the duration of non-spontaneous blink to simulate the actual situation, which according to our threshold is at least 0.3+0.3=0.6 seconds in total.

We removed unreasonable input time outliers, which are shorter than 1s or longer than 10s due to trigger detection errors. We determined the threshold by observation of the experiment and the collected data. To avoid data unbalance which is illegal in RM-ANOVA, we replaced these outliers with corresponding median values of that size.

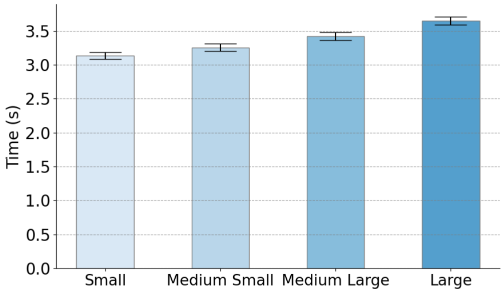
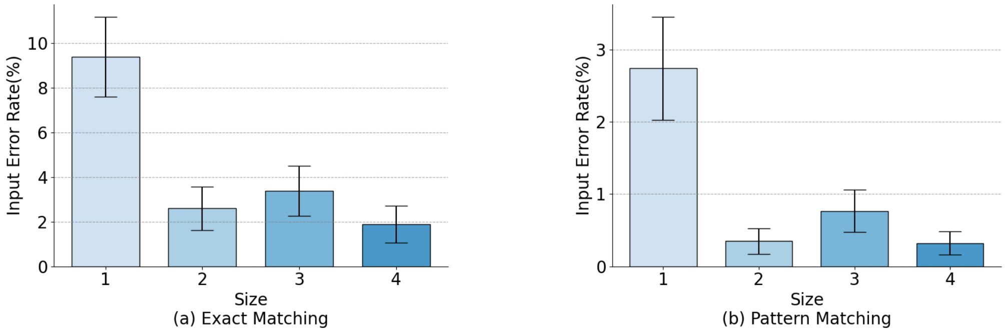


Fig. 5. Bar plots showing main effects of size on the input time. The error bar shows standard error.

RM-ANOVA revealed a statistically significant main effect of size on input time ($F_{3,45}$ = 16.5, p < .001). We further averaged the input time of different PINs under the same size. Figure 5 illustrated the input time across each size. A notable effect of the size on input time ($F_{3,45}$ = 16.5, p < .001) was observed. Post-hoc comparison showed that the input time monotonically increases with significant differences among all sizes except for **small** (M=3.13s, SE=0.05s) and **medium small** (M=3.26s, SE=0.05s) (p>.05), indicating that smaller size won't contribute to faster input due to more careful but time-consuming target selection under smaller sizes. A significant difference between **medium small** and **medium large** (M=3.42s, SE=0.06s), **medium small** and **large** (M=3.65s, SE=0.06s) indicated that larger size than **medium small** would increase input time significantly, which reduce usability. To sum up, in terms of input time, **medium small** outperformed 3 other sizes.

4.4.2 Input Error Rate. We calculated input accuracy with Exact Matching (total error rate) and Pattern Matching (digit-wise) strategies. Exact Matching refers to scoring 1 if the input pattern and given pattern are identical and 0 if they are not the same. Pattern Matching refers to the proportion of user-input points that deviate from the designated points. Figure 6 illustrates the average input error rate across 4 sizes.



(a) Exact Matching

(b) Pattern Matching

Fig. 6. Bar plots showing average input error rate calculated by Exact Matching and Pattern Matching. The error bar shows standard error. Left: Average Input Error Rate by Exact Matching, right: Average Input Error Rate by Pattern Matching

RM-ANOVA found a significant difference of sizes on input error rate calculated by both Exact Matching ($F_{3,48} = 8.06$, p<.001) and Pattern Matching ($F_{348} = 4.09$, p<.01) strategies. In terms of Exact Matching, post-hoc comparison with Bonferroni correction shows that only between **small** (M=9.4%, SD=1.79%) and **medium small** (M=2.6%, SD=0.97%) (p<.001), **medium large** (M=3.38%, SD=1.11%) (p<.01), **large** (M=1.89%, SD=0.84%) (p<.001) do sizes bear significant difference. In terms of Pattern Matching, post-hoc comparison shows that only between **small** (M=2.74%, SD=0.71%) and **medium small** (M=0.35%, SD=0.18%) (p<.001), **medium large** (M=0.76%, SD=0.29%) (p<.01), **large** (M=0.32%, SD=0.16%) (p<.001) do sizes bear significant differences. These results indicate that **small** is the most error-prone overall, from which we speculate that excessively **small** is head and eye control demanding. There is no significant difference among **medium small**, **medium large**, and **large**, illustrating that **medium small**'s input accuracy performance is as good as larger sizes'.

4.4.3 Subjective Ratings. Overall, participants found the designed interface intuitive and straightforward. With some practice, they perceived drawing patterns using gaze as a user-friendly input method. Figure 7 showcased participants' mean subjective ratings.

Among the four sizes, a significant difference was found only in terms of **physical demand**($\chi^2(3) = 14.9$, p<.01) and **input speed**($\chi^2(3) = 13.4$, p<.01). In these two questions, subjective ratings for the four sizes decrease monotonically, which aligns with the actual sensory experience, although no significant difference was observed between **small** and **medium small**.

4.4.4 Case Analysis: Head-Eye coordination. The previous research [75] has proposed that for objects with an amplitude smaller than 25° (corresponding to a FOV less than 50°), the majority of head rotation contributes less than 10%. In our experiments, we observed this phenomenon in **small** with 40° FOV. Therefore, we opted for **medium small** that induces a noticeable contribution from the head. Based on head-eye coordination modeling discussed in subsection 2.2, analyzing data from study1, we observed:

(1) Looking into a certain participant, we found when the participant drew different patterns with an overlapped spatial trajectory part, head movement amplitude was stable across different patterns.

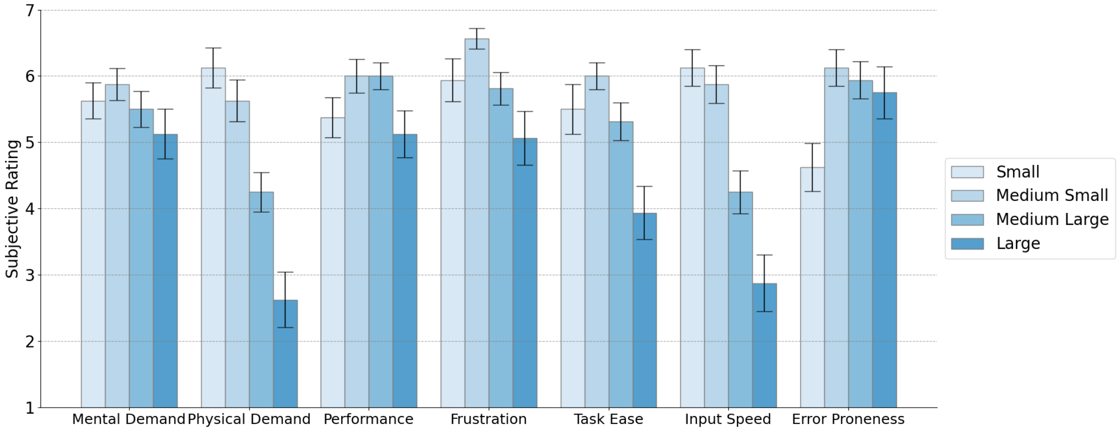


Fig. 7. Participants' subjective ratings (7: most positive, 1: most negative). The error bar shows standard error.

(2) Compared between different participants drawing the same group of patterns, we found some participants are head movers, and some are non-head movers [20].

For this point, we select some input examples with overlapped spatial trajectories for illustration. We plot the trajectories of three selected patterns each drawn by one of the four participants shown as Figure 8.

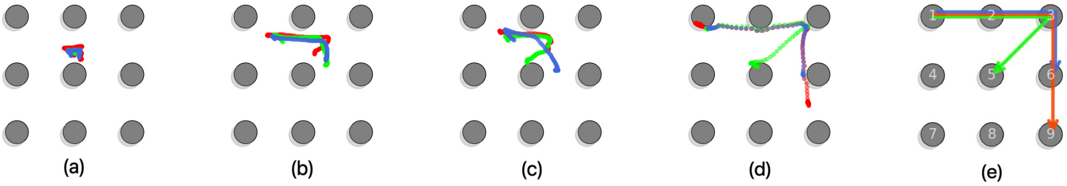


Fig. 8. 4 participants' trajectories of selected 3 patterns and 3 selected patterns with an overlapped spatial trajectory "123". (a), (b), (c), and (d) are drawn by 4 different participants. 3 selected patterns corresponding to different colors are shown in (e).

From Figure 8, we observed that trajectories across 3 colors overlap stably in the north row. From (a) to (d), The amplitude of head motion increased monotonically. This aligned with Fuller's head-eye coordination modeling: head movers and non-head movers [20]. (a) is a non-head mover to this extent, while (b), (c) and (d) are head movers.

This result indicates under **medium small**, CoordAuth can trigger behavioral features that can be implemented for authentication. section 5 will further discuss this head-eye coordination feature.

4.5 Discussion and Design Decision

We determined our size to **medium small** (60° FOV) for the following reasons. In subjective ratings, **medium small** and **medium large** were among the highest. However, regarding input time and input error rate, medium small has the highest cost-effectiveness. Regarding input time, **medium small** is as quick as **small** and significantly quicker than **medium large**, and **large**. Regarding input error rate, **medium small** performs as well as **medium large** and **large**, and significantly better than **small**. Additionally, under **medium small**, features of saccades and fixation segments and head-eye coordination are well observed, which will be used in Section 5.3.

5 COORDAUTH'S ALGORITHM DESIGN AND IMPLEMENTATION

5.1 Algorithm Flow

Similar to most traditional authentication methods [92], CoordAuth consists of two stages: registration and login. In the registration phase, users are required to input their desired patterns multiple times. During this process, CoordAuth's IMU and eye tracker collect head and eye movement data as behavioral biometrics. Simultaneously, the device records the user-inputted patterns as a knowledge-based password.

During the login process, after users input the pattern password, CoordAuth's two-factor authentication mode checks the correctness of both the pattern and the head-eye movement features. Figure 9 shows a visual representation of CoordAuth's two-factor authentication algorithm.

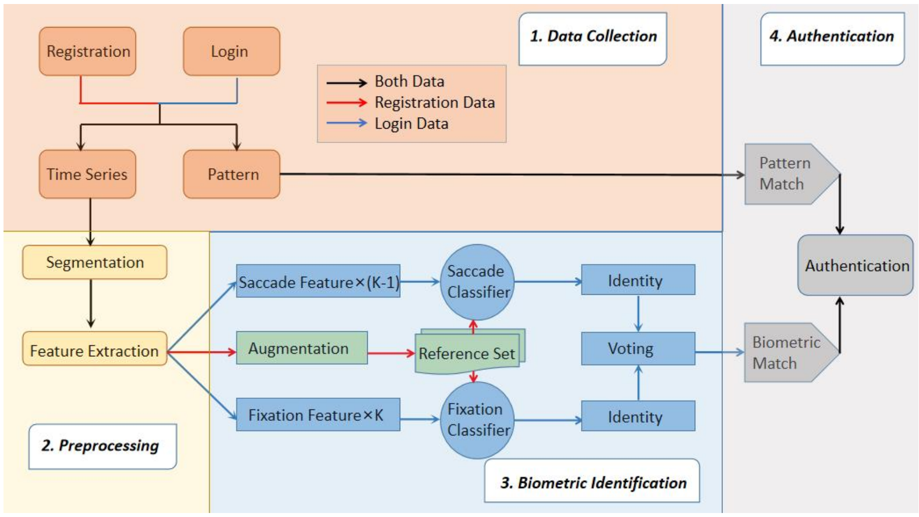


Fig. 9. Authentication algorithm design of CoordAuth.

In summary, during user registration, CoordAuth follows the following steps:

(1) Collect the pattern password and record head and eye movement features.
(2) Segment the inputted data sequences into segments corresponding to saccades and Fixations and extract the relevant features.
(3) Use Gaussian noise for data augmentation and train the saccade and fixation classifiers. These classifiers will be updated when a new user registers.

During the login process, CoordAuth verifies the user's identity as follows:

(1) Compare the inputted pattern with the registered pattern.
(2) Segment the head and eye data sequences, extract the corresponding features, similar to step 2 during registration.
(3) Use the classifiers trained during the registration phase to classify the features from step 2.

(4) Access is granted only if both the inputted pattern and the biometric features match the registered identity; otherwise, access is denied.

Since the comparison of pattern passwords is straightforward and simple, our focus here is on exploring the behavioral biometric-based part of the authentication algorithm.

5.2 Segmentation of Head and Eye Sequences

Some related studies [1, 58] verified that the time between two saccades during the user's input is 100-200 ms. Based on this, we establish our segmentation criterion: if a user's gaze remains within a point and its hit test area for more than 200 ms, this sub-time sequence is considered a fixation segment. Therefore, the entire time sequence can be divided into K-1 saccade segments by fixation segments. The starting and ending points of the pattern naturally form fixation segments.

For example, we visualized the input time series data of participant P14 drawing pattern No.13 under size **medium small** study 1, as shown in Figure 10. Grey intervals represent fixation segments during input.

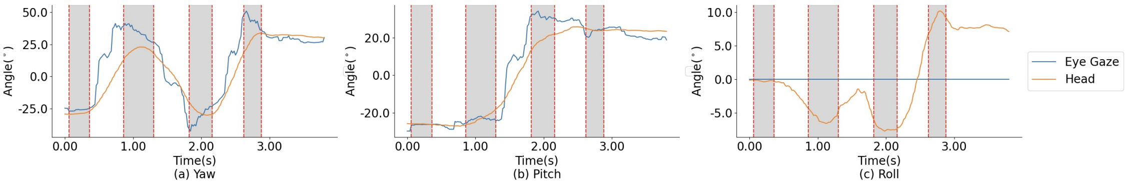


Fig. 10. Time series data for participants P14, Size medium small (60° FOV), Pattern No.13, Study1. Left is Yaw angle, middle is Pitch angle and right is Roll angle. Grey intervals represent fixation segments during inputting.

Additionally, for a pattern-based password, assuming it consists of N points, there are N-1 input lines connecting these N points. During the user's input process, their gaze needs to pass through these N points sequentially, forming a pattern composed of N-1 connected line segments. This process can be divided into K fixation segments and K-1 saccade segments. As mentioned earlier, in the input process, if the gaze sequentially selects two endpoints on a straight line, the midpoint of the line will be automatically chosen as the point between these two points. As a result of this, we observed that some users, when inputting a straight line formed by three points, tend to use their gaze to select the two endpoints of the line as input. In this case, although the midpoint is automatically included in the pattern password, they do not pause their gaze on the midpoint. Therefore, for a password consisting of N points, we can conclude that K must be less than or equal to N. Since we stipulate that each point cannot be reused, N can be at most 9, and K cannot exceed 9 either.

5.3 Feature Extraction

For the fixation and saccade segments, we collected real-time motion data. To minimize the impact of the user's initial position and posture, We collected data about the rotation and position of the head and eyes when the user is facing the 3*3 unlock UI. Based on this, we calculated the position and rotation of the head and eyes in the coordinate system of the initial posture. The sensor data included head position, head rotation, and eye rotation.

$$\text{Head Position} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad \text{Head Rotation} = \begin{bmatrix} q_w \\ q_x \\ q_y \\ q_z \end{bmatrix}, \quad \text{Eye Rotation} = \begin{bmatrix} q_w \\ q_x \\ q_y \\ q_z \end{bmatrix}$$

As shown, in the raw data from both the IMU and eye tracker, the position is represented as a three-dimensional vector (x, y, z), and the rotation is represented as a four-dimensional quaternion (w, x, y, z) with its corresponding rotation matrix:

$$\mathbf{R}(\mathbf{q}) = \begin{bmatrix} 1 - 2q_y^2 - 2q_z^2 & 2q_xq_y + 2q_zq_w & 2q_xq_z - 2q_yq_w \\ 2q_xq_y - 2q_zq_w & 1 - 2q_x^2 - 2q_z^2 & 2q_yq_z + 2q_xq_w \\ 2q_xq_z + 2q_yq_w & 2q_yq_z - 2q_xq_w & 1 - 2q_x^2 - 2q_y^2 \end{bmatrix}$$

We can use the following formula to convert the quaternion to represent Turning (Yaw), nodding (Pitch), and tilting (Roll) [16, 85].

$$\text{Yaw} = \text{atan2}\left(\mathbf{R}_{21}, \mathbf{R}_{11}\right)$$

$$\text{Pitch} = -\text{asin}(\mathbf{R}_{31})$$

$$\text{Roll} = \text{atan2}\left(\mathbf{R}_{32}, \mathbf{R}_{33}\right)$$

In study 1, we extensively explored the behavioral biometric difference among individuals when using the pattern password of head-eye coordinated movements. These differences effectively serve as features for our model. Additionally, drawing inspiration from previous research [76, 95] and referring to methods for processing biometric signals [92], we identified the following features:

- F1: Fixation Features. In study 1, we observed diverse proportions of head-eye coordination during fixation across different individuals. Some users significantly rotated their heads toward the direction they were looking, while others primarily relied on eye rotation. To capture these features, for each fixation segment, we transformed eye rotations in world coordinates into rotations relative to the head coordinate system, represented as $\mathbf{Yaw}_{\text{Diff}} = Yaw_{\text{Eye}} - Yaw_{\text{Head}}$, $\mathbf{Pitch}_{\text{Diff}} = Pitch_{\text{Eye}} - Pitch_{\text{Head}}$, $\mathbf{Roll}_{\text{Diff}} = Roll_{\text{Eye}} - Roll_{\text{Head}}$. We obtain $\mathbf{Rot}_{\text{Diff}} = \left[Yaw_{\text{Diff}}, Pitch_{\text{Diff}}, Roll_{\text{Diff}}\right]$, $\mathbf{Pos}_{\text{Head}} = \left[x_{\text{Head}}, y_{\text{Head}}, z_{\text{Head}}\right]$, and $\mathbf{Rot}_{\text{Head}} = \left[Yaw_{\text{Head}}, Pitch_{\text{Head}}, Roll_{\text{Head}}\right]$. Thus, for a fixation segment, we obtain 9 sub-time series. Additionally, we computed the second derivatives (representing acceleration or angular acceleration) of these series. Subsequently, statistical features (maximum, minimum, mean, median, variance) were computed for each dimension of these sequences. Therefore, for each fixation segment, these features collectively constitute $(3 + 3 + 3) \times 2 \times 5 = 90$ dimensions.
- F2: Saccade Features. In study 1, we observed varying degrees of head lag behind the eyes during the fixation process across different individuals. To capture this phenomenon, we utilized head rotation in world coordinates, represented as $\mathbf{Rot}_{\text{Head}} = \left[Yaw_{\text{Head}}, Pitch_{\text{Head}}, Roll_{\text{Head}}\right]$, and eye rotation matrices $\mathbf{Rot}_{\text{Eye}} = \left[Yaw_{\text{Eye}}, Pitch_{\text{Eye}}, Roll_{\text{Eye}}\right]$, along with their first-order derivatives (indicating velocity), to compute the cross-correlation function lists L_{type} for respective sub-time series [15]:

$$L[k]_{\text{type}} = \text{CCF}(X = \text{type}_{\text{Head}}, Y = \text{type}_{\text{Eye}}, k) = \frac{\sum_{t=1}^{N-k}(X_t - \bar{X})(Y_{t+k} - \bar{Y})}{\sqrt{\sum_{t=1}^{N-k}(X_t - \bar{X})^2 \cdot \sum_{t=1}^{N-k}(Y_{t+k} - \bar{Y})^2}}$$

Here, 'k' represents the time lag, ranging from the first timestamp of the sub-time series to the last, which is N. The calculated $L[k]_{\text{type}}$ for each 'k' lag forms the list L_{type}, 'type' refers to the category of the sub-time series. For example, $L[k]_{\text{Yaw}}$ is computed as $L[k]_{\text{Yaw}} = \text{CCF}(X = \text{Yaw}_{\text{Head}}, Y = \text{Yaw}_{\text{Eye}}, k)$. Subsequently, we computed the statistical features (maximum, minimum, mean, median, variance) for each type of these cross-correlation function lists L_{type}. Additionally, we treated $\mathbf{Rot}_{\text{Eye}} = \left[Yaw_{\text{Eye}}, Pitch_{\text{Eye}}, Roll_{\text{Eye}}\right]$, $\mathbf{Pos}_{\text{Head}} = \left[x_{\text{Head}}, y_{\text{Head}}, z_{\text{Head}}\right]$, and $\mathbf{Rot}_{\text{Head}} = \left[Yaw_{\text{Head}}, Pitch_{\text{Head}}, Roll_{\text{Head}}\right]$ as features for each saccade segment. Furthermore, we calculated the first derivatives (representing velocity or angular velocity) of each series. Then, statistical features (maximum, minimum, mean, median, variance) were computed for each dimension of these sequences. Consequently, for each saccade segment, these features collectively constitute 3 × 2 × 5 + (3 + 3 + 3) × 2 × 5 = 120 dimensions.

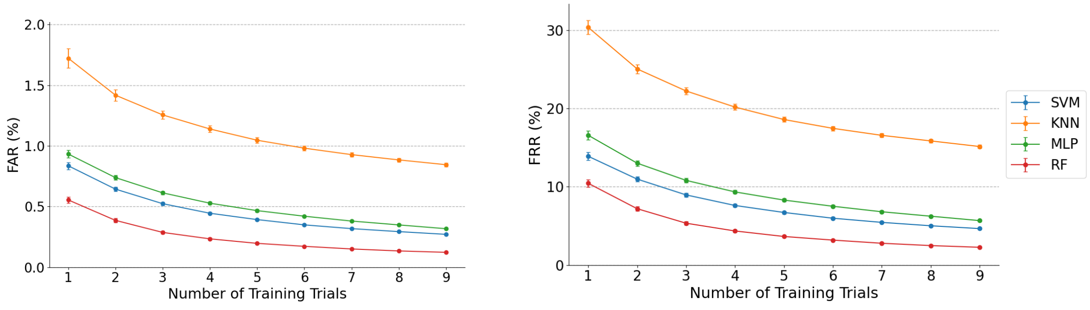
In conclusion, the feature dimensions for each fixation segment are 90, and for each saccade segment are 120.

Furthermore, for each fixation segment and saccade segment, the biometric features vary among different users. Therefore, we separate these two types of segments, using the $i\,th$ fixation segment to train the $i\,th$ fixation classifier and the $j\,th$ saccade segment to train the $j\,th$ saccade classifier. For a pattern password of length N, there can be K fixation classifiers and K-1 saccade classifiers. Due to the constraint that $K \leq N \leq 9$, theoretically, we only need to train a maximum of 8 saccade classifiers and 9 fixation classifiers to cover predictions for all segments.

We removed outliers from the original data and applied data augmentation techniques, adding Gaussian noise to the original data's standard deviation. The augmentation ratio is 20 [43, 92]. Each feature in the training set is normalized to have a mean of 0 and a standard deviation of 1. The test set is normalized using the same mean and standard deviation as the training set.

5.4 Model Selection

After feature extraction, we employed classification models to implement the authentication. We utilized four commonly used methods [51, 59, 95]: Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Multilayer Perceptron (MLP). We partitioned the training set and test set based on data collected to determine the optimal method. Using the scikit-learn machine learning library in Python, we trained classifiers with default hyperparameter values.

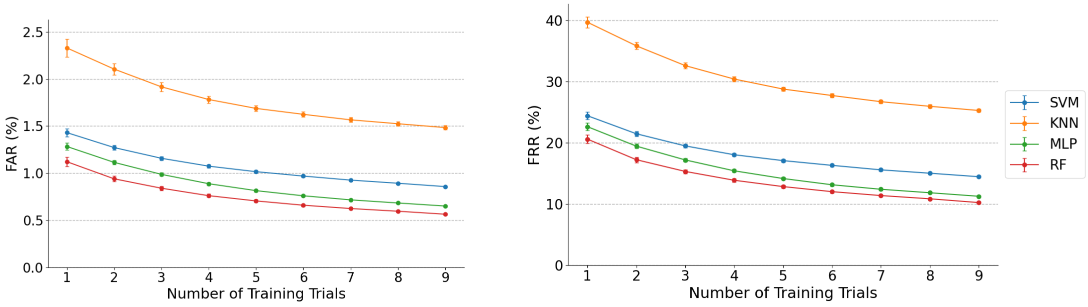


(a). FAR of Single Saccade classifier

(b). FRR of Single Saccade classifier

Fig. 11. For a single saccade classifier, the (a) FAR and (b) FRR as a function of T across different models. The error bar shows standard error.

Based on the UI layout determined in Study1, we additionally recruited 18 participants (9 males, 9 females) with an average age of 23.6 (SD = 2.4) from the campus. To alleviate the burden, we selected the top 10 likely used patterns [13],



(a). FAR of Single Fixation classifier

(b). FRR of Single Fixation classifier

Fig. 12. For a single fixation classifier, the (a) FAR and (b) FRR as a function of T across different models. The error bar shows standard error.

each entered 10 times, resulting in a total of 18 × 10 × 10 = 1800 samples. We utilized the Gaussian noise mentioned earlier for data augmentation. To assess the performance of the saccade classifier and fixation classifier under different training sizes, we selected T samples for the training set, and the remaining 10-T samples for the test set. For each password, each person took turns being the legitimate user, while others are considered illegitimate users. We took turns using one classifier from these fixation classifiers and saccade classifiers to perform binary classification on the corresponding segmented sequences. To evaluate the authentication performance of each classifier, we used two widely adopted metrics [92, 96]:

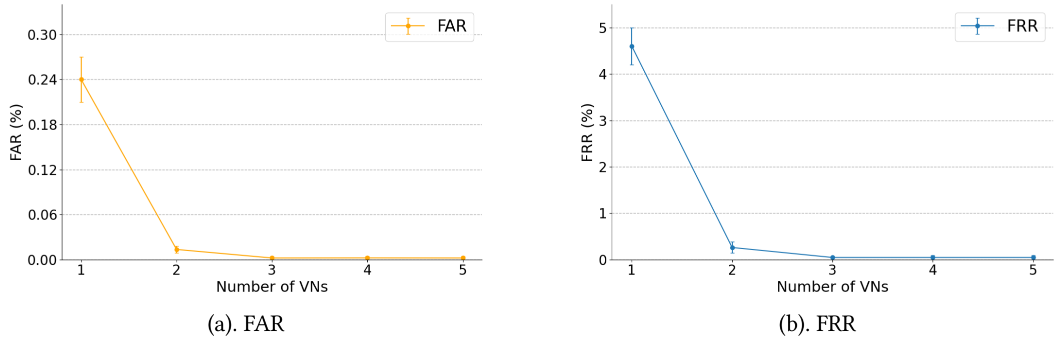
- False Rejection Rate (FRR): The percentage of legitimate samples incorrectly rejected.
- False Acceptance Rate (FAR): The percentage of illegitimate samples incorrectly accepted.

Figure 11 and Figure 12 respectively display the variations of FRR and FAR of the saccade classifier and fixation classifier. The performance of the saccade classifier excels over that of the fixation classifier, showing lower FAR and FRR. With the change of T, the Random Forest algorithm exhibits the lowest FRR and FAR, performing exceptionally well under both classifiers. Specifically, at a training set sample size of $T = 9$, the FAR for the saccade classifier is 0.12% (SE=0.01%), and the FRR is 2.27% (SE=0.08%). For the fixation classifier, the FAR is 0.56% (SE=0.01%), and the FRR is 10.2% (SE=0.2%). We choose the Random Forest algorithm for both the saccade classifier and the fixation classifier.

5.5 Ensemble Learning Model

To fully leverage behavioral biometrics, we design an ensemble learning model by combining the two mentioned classifiers. As mentioned earlier, a pattern password consists of K fixation segments and K-1 saccade segments. For simplicity, we mix fixation and saccade classifiers in a 1:1 ratio. For a sample to be recognized, it is first transformed by segmentation into fixation and saccade segments. Then, these segments are converted into the input format by through feature engineering for the fixation and saccade classifiers. Subsequently, the legality of each segment is predicted using the corresponding fixation and saccade classifiers. Finally, the identity is determined through majority voting, selecting the class with the highest number of votes. Due to the voting nature, we refer to the combined classifier of the $i\,th$ fixation classifier and the $i\,th$ saccade classifier as the $i\,th$ "Voting Neural" (VN).

The $i\,th$ VN is responsible for predicting the $i\,th$ saccade and fixation segments, casting two votes. Therefore, the change in the number of VNs will impact the amount of information used in the classification. For example, if only one



(a). FAR (b). FRR

Fig. 13. With the change in the number of VNs, the variations in the values of (a) FAR and (b) FRR. The error bar shows standard error.

VN is used, it implies that for all input patterns, we only focus on the first saccade and fixation after segmentation. This results in a significant loss of behavioral information. However, having too many VNs is also unnecessary, as an excessive number of VNs implies a heavy training burden.

We aimed to determine the impact of the number of VNs on the performance of the Ensemble Learning Model. Figure 13 illustrates the changes in FAR and FRR of the model as the number of VNs increases. As the number of VNs increases from 1 to 3, both FAR and FRR exhibit a sharp declining trend. When the number of VNs is 3, the FAR and FRR of the model have already reached the lowest values, at 0.003% and 0.045%. It is worth noting that when the number of VNs is large, exceeding the number of fixation and saccade segments in some samples, the excess VNs will not operate on those samples. Overall, for our Ensemble Learning Model, we chose 3 VNs as our parameters. This implies that we recommend using patterns composed of 4 or 5 points because additional fixations and saccades will not impact the voting mechanism of CoordAuth. On the contrary, they may increase the input burden and provide more pattern information, which might aid shoulder-surfing attackers in guessing the specific password [95].

6 STUDY 2: SIMULATION ON AUTHENTICATION PERFORMANCE

For a two-factor authentication system, its stability and robustness are crucial. In this section, we conducted two simulations:

- S1: Robustness to Gesture Collision
- S2: Long Term Stability

Our simulations use real user data, and the algorithmic framework and parameters align with those in subsection 5.5.

6.1 Participants and Apparatus

We recruited 24 participants (13 males, 11 females; average age=25.3, SD=1.6), all from our campus. None had participated in our previous studies. The self-reported average virtual reality usage experience was 2.9 (SD=0.6, out of a maximum score of 5). Each participant received a compensation of $15. We used the Meta Quest Pro Headset and the Oculus Quest Link to establish communication between the computer and the headset.

6.2 Study Design

We have chosen the layout finalized in Study 1 for simulation. Similar to study 1, we provide visual feedback whenever a user successfully inputs a point.

In simulation 1, to assess the robustness of our system against password collisions in extreme scenarios, we must ask all participants to input the same password. This allowed us to analyze the performance of CoordAuth. We selected the same 20 patterns as in Study 1. These 20 patterns are among the most common in Android smartphones, ensuring they are representative and diverse enough.

In simulation 2, we wanted to assess their memory of their own password, so we asked users to personally design and remember two pattern passwords to closely replicate real-world scenarios, and then asked them to input these two patterns on different days. Before their input, we asked them to write down the pattern password they remember and compare it with the password they registered on day 0. If they forgot, we displaced their pattern password from day zero to help them input.

6.3 Procedure

We split the experiment into three sessions. Each participant needed to accomplish the first two sessions. Some of the participants needed to accomplish the third session. Before the experiment, we informed users about the aim of the study and their tasks. We then gave participants 5 minutes to become familiar with the input and the feedback of CoordAuth.

In the first session, participants are provided with sketch paper and pen. We asked participants to independently design two pattern passwords, subsection 5.5 led us to recommend that the pattern should consist of a minimum of 4 points. Following this, participants were instructed to put on the VR headset and throughout 10 consecutive rounds, input their first and second patterns. Short breaks were provided between each round. The entire experiment took approximately 10 minutes. We then forced the participants to take a 5-minute break before the second session.

In the second session, participants were asked to input these 20 passwords in 10 consecutive rounds, with the appearance order randomized in each round. Similar to Study 1, the UI would display the pattern to be input. We provided participants with short breaks between each round. The entire process took approximately 40–50 minutes.

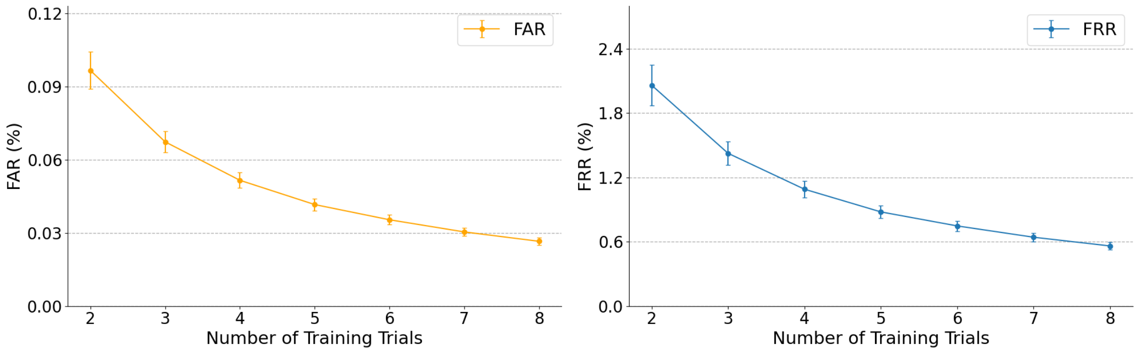
Finally, on the third and seventh days following the initial experiment (day 0), we randomly chose 10 (6 females, 4 males) participants with an average age of 23.9 (SD= 2.6) to test the system's longitudinal robustness. Upon the collection, participants were asked to input their two self-created pattern passwords ten times each, simulating the act of unlocking a mobile device multiple times daily [32].

6.4 Results

For simulation 1, we obtained a total of 20 PINs × 10 times × 24 participants = 4800 samples. For simulation 2, we obtained a total of 2 PINs × 10 times × 10 participants × 3days = 600 samples.

6.4.1 Simulation 1: Robustness to Password Collision. In Simulation 1, to assess how different registration frequencies impact performance, we varied the number of training trials, using the data from the first T trials as the training set and the remaining 10 - T trials as the test set. We validated the extreme case of password collisions, and Figure 14 illustrates the impact of the training set size on FRR and FAR in simulation 2. With an increase in the number of training trials, the FRR of CoordAuth significantly decreased, dropping from 2.06% (SE = 0.19%) for 2 training samples to 0.56% (SE = 0.03%) for eight training samples. We achieved a remarkably low FAR ranging from 0.096% (SE = 0.007%) to 0.027% (SE

= 0.001%). These results highlight the accuracy of CoordAuth in validating legitimate inputs and rejecting illegitimate ones.



(a). FAR

(b). FRR

Fig. 14. With the change in the number of training trials, the variations in the values of (a) FAR and (b) FRR. The error bar shows standard error.

According to the FRR and FAR curves, during the actual user registration process, performing 5 repetitions is sufficient to achieve a low FRR of 0.88% (SE=0.06%) and an FAR of 0.042% (SE=0.001%). This allows for a balance between user fatigue and accuracy.

6.4.2 Simulation 2: Long Term Stability. In simulation 2, we used "Number of passwords" to represent the total number of passwords that need to be recalled in this round, while "total memory recall rate" indicated the proportion of passwords recalled by the users in this round. Only one user forgot all the two passwords designed by themselves on the third day, resulting in an 18 / 20 = 90% memory recall rate on that day. Since we had already shown the passwords to that user after the forgetfulness on the third day, just to help him to finish the simulation study. We did not include this user in the statistics for the seventh day. For the remaining nine users, they maintained a 100% accuracy rate in memory on both the third and seventh days.

We split the data from Day 0 into 5 training trials and 5 test trials. The classifier was trained using the training set obtained from the split on Day 0. Subsequent login data from Day 3 and Day 7 were used to evaluate the model's stability for the long term. The changes in FRR and FAR over time are illustrated in the Table 1. If we continually use the classifier trained on day 0 data, the FAR gradually increases over time, rising from 0% on day 0 to 1.6% (SE = 0.1%) on day 3, and further to 2.6% (SE = 0.2%) on day 7, the FRR initially rapidly increases from 0% on day 0 to 11.4% (SE = 3.0%) on day 3, and then gradually grows to 12.6% (SE = 3.7%) on day 7.

We consider updating the classifier over time. Assuming that on the third day, the user performs five logins, by the end of the third day, we retrain the classifier using the data from the third day combined with the previous data. Based on this approach, the performance on day 7 is shown in Table 1. We achieved a lower FAR of 0.32% (SE=0.01%) and a lower FRR of 2.73% (SE=0.11%). In summary, CoordAuth demonstrates effective validation of legitimate users and rejection of unauthorized users over time, maintaining stable long-term performance.

Data Source	FAR (%)	FRR (%)
Day 0	0.0 (SE=0.0)	0.0 (SE=0.0)
Day 3	1.6 (SE=0.1)	11.4 (SE=3.0)
Day 7 (Unchanged Classifier)	2.6 (SE=0.2)	12.6 (SE=3.7)
Day 7 (Retrained Classifier)	0.32 (SE=0.01)	2.73 (SE=0.11)

Table 1. As time progresses, the values of FRR and FAR

7 STUDY 3: EVALUATING COORDAUTH'S USABILITY

In addition to security, usability is another critical criterion for evaluating the authentication framework. We measured the usability of CoordAuth based on aspects such as registration and login time, error rates and robustness against contextual factors.

7.1 Participants and Apparatus

We recruited 20 participants (11 males, 9 females) with a mean age of 24.1 (SD=1.9), 6 of them participated in previous experiments. Each participant received a compensation of $10. We used the Meta Quest Pro Headset and the Oculus Quest Link to establish communication between the computer and the headset.

7.2 Study Design

Based on the typical usage scenarios of VR, we defined two contextual factors: standing and sitting postures. This is because, during the authentication using VR, users are mostly stationary, rarely walking or running. For each contextual factor, we collected multiple instances of users' input patterns, along with raw data from IMU and eye tracker. We employed the same login time metric as in Study 1 and utilized the exact match criterion from Study 1 as the metric for the total error rate.

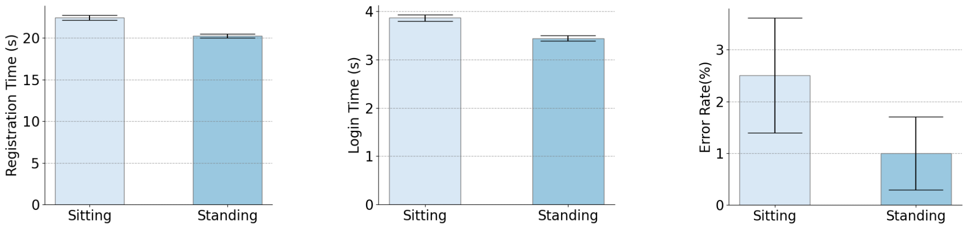
7.3 Procedure

We introduced the experiment task to the participants and gave them 5 minutes to become familiarize with the appratus. Then participants needed to undergo two sessions: registration and login. For registration, they needed to perform 5 consecutive attempts for sitting and standing posture each. While for login, they further needed to perform 5 consecutive attempts for each posture. The order of standing and sitting was counter-balanced for different participants.

7.4 Results

We got 20 participants × 1 pattern × 10 times x 2 postures = 400 samples. We used statistical testing with a sequence of first adopting Shapiro-Wilk normality analysis, then adopting RM-ANOVA for parametric test and corresponding non-parametric test, and last adopting post-hoc test methods to examine.

7.4.1 Authentication Time and Error Rate. The registration time, login time and error rate for both sitting and standing postures are shown in Figure 15. RM-ANOVA found the average registration time in standing posture (M=20.24s, SE=0.26s) is significantly shorter than in sitting (M=22.43s, SE=0.29s) ($F_{1,19}$=24.4, p<.001). The same conclusion applies to login as well($F_{1,19}$=306.7, p<.001), with standing (M=3.4s, SE=0.06s) and sitting (M=3.82s, SE=0.06s). This finding aligns with previous work that standing enables larger range and flexibility of head movement than sitting [73]. Additionally, there is no notable effect on the input error rate of postures (p=0.254), with standing (M=1.0%, SE=1.1%) and sitting (M=2.5%, SE=0.7%), emphasizing CoordAuth's cross-contexts usability.



(1) Registration (2) Login (3) Error Rate

Fig. 15. (a) Average registration time for different gestures. (b) Average login time for different gestures. The error bar shows standard error.

Furthermore, for comparing the usability of CoordAuth with other authentication methods, we referenced four baselines: Blinkey (two-factor) [95], RoomUnlock (behavior biometric) [23], and PIN and Pattern (knowledge-based) [24], all utilizing a sitting posture. Therefore, we compared CoordAuth's input time, registration time, and total error rate in the sitting posture with these baselines. Among these baselines, the input time for Blinkey [95] is 9.6s on average, for RoomLock [23] it ranges from 8.58s to 14.33s, PIN [24] ranges from 2.38s to 3.36s, and Pattern [24] ranges from 2.87s to 3.84s. To this extent, we found that CoordAuth is slightly slower than PIN and Pattern, but faster than Blinkey and RoomUnlock, the input time of CoordAuth is less than half of the time Blinkey and RoomUnlock take. Although CoordAuth is slightly slower in terms of input time compared to traditional PIN and Pattern, its security is higher in scenarios of password leakage and shoulder surfing attacks due to the presence of behavior biometrics compared to PIN and Pattern.

Although in terms of registration time, methods like PIN and Pattern, which don't require registration, have an advantage over CoordAuth, CoordAuthwas still approximately 20 seconds shorter than Blinkey [95]. Due to the strategy of updating the classifier over time mentioned in the simulation study, CoordAuth only requires an initial registration. We believed that sacrificing some registration time for device security was worthwhile.

In terms of input error rate, the input error rate of CoordAuth is 2.5%, while Blinkey [95] is 6.1%, RoomLock [23] is 3.5%, PIN [24] is 2.3%, and Pattern [24] is 10.3%. It indicates that the input error rate of CoordAuth is as low as PIN's and outperforms 3 other baselines. It is worth noting that CoordAuth implemented a head-eye coordination approach for pattern input with a 25% error rate compared to traditional Patterns, demonstrating the novelty and usability of CoordAuth.

7.4.2 Robustness against contextual factors. We trained the model using user registration data in either a seated or standing posture and then investigated the performance of CoordAuth during login when the user's posture differed from the registration pose. As shown in Figure 16, we observed that in the case of registration in a seated position, the False Rejection Rate (FRR) during login was 4.4% (SE=1.4%), and the False Acceptance Rate (FAR) was 0.2% (SE=0.1%). Similarly, when registered in a standing position, the FRR during login was 9.5% (SE=5.4%), and the FAR was 0.5% (SE=0.3%). Overall, when registering with one posture and logging in with another, the FRR is below 10% and the FAR is below 0.5% for CoordAuth, demonstrating the robustness of CoordAuth against contextual factors.

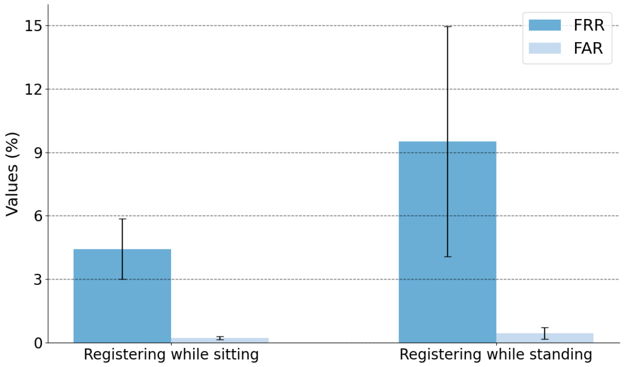


Fig. 16. FAR and FRR under different registration strategies. The error bar shows standard error.

Overall, CoordAuth performs exceptionally well in terms of usability. Participants highlighted that the interaction method of CoordAuth is highly innovative and enjoyable, as mentioned by P13. P8 mentioned that once the input method is mastered, CoordAuth becomes highly efficient, allowing for easy drawing of the desired pattern password.

However, CoordAuth faced some criticisms. P7 pointed out it required constant focus during usage to avoid mis-selections, leading to fatigue. P2 observed occasional delays in inputting points at the corners of the grid. Additionally, P11 mentioned that the overall design of the UI lacked aesthetic appeal.

8 STUDY 4: ROBUSTNESS AGAINST SHOULDER SURFING ATTACKS

We conducted a study to investigate CoordAuth's resistance capability against shoulder surfing attacks. To measure the resistance of CoordAuth against attacks, we introduce the widely adopted metric Attack Success Rate (ASR) [4]. ASR represents the proportion of successful attempts by attackers to impersonate a legitimate user identity in attacks targeted at the authentication system.

8.1 Participants

As CoordAuth introduces a relatively novel authentication method in VR, familiarity with its UI layout and interaction mechanics is limited to individuals who have directly experienced using VR headsets. Therefore, to evaluate CoordAuth under more rigorous attack scenarios, we randomly selected 16 participants (7 males, 9 females, average age=24.1, SD=2.3) from the simulation study and paired them into 8 pairs. These pairs take turns acting as attackers, ensuring that our attackers can more efficiently conduct attempts at attacks. Each participant received a compensation of $5.

8.2 Study Design

To simulate a real-world scenario reflective of potential security vulnerabilities, our study design involved pairs of participants where one acted as the legitimate user and the other as the attacker. The attacker was positioned 1 meter away to the side of the legitimate user [95]. The legitimate user inputs the same authentication pattern three times, allowing the attacker to observe and memorize the sequence. Subsequently, the attacker attempts to access the system five times using the observed pattern. A successful attempt in any one of the five is considered a successful attack.

8.3 Procedure

We first divided each pair of participants (out of 8 pairs) into one legitimate user and one attacker.

(1) **Legitimate User:**
   (a) Introduced to the entire experimental process.
   (b) Sitting on a chair, entering his or her password, with intervals of 10 to 20 seconds between each attempt, totaling three attempts.

(2) **Attacker:**
   (a) Provided with white paper and a black pen to speculate and record the legitimate user's password.
   (b) Observing the legitimate user's three inputs from approximately 1 meter behind and speculating on paper.
   (c) Allowed to take the paper with speculated passwords, wearing a head-mounted display and making five attempts, during which they can review the speculated passwords on the paper.

This process was repeated for two rounds, with different passwords in each round. Feedback on the success of intrusion was provided during this period. Following the completion of both rounds, the participants switched roles—the observer became the legitimate user and vice versa— and the experiment was replicated. Finally, a brief interview was conducted with each participant, lasting approximately 5 minutes. The entire process lasted about half an hour.

8.4 Results

In Study 4, we obtained a total of 16 participants × 2 self-created patterns = 32 patterns and conducted 16 participants × 2 patterns × 5 attack attempts = 160 attack attempts. In these 160 attack attempts, there were 16 attempts where the pattern password was guessed after shoulder surfing, successfully passing the knowledge-based authentication of CoordAuth, accounting for 10% of the total attempts. Among these 16 attempts, only 1 successfully passed the behavior biometric authentication of CoordAuth, resulting in an ASR of 1/160 = 0.6%. 10 out of 32 patterns were guessed correctly at least once in the 5 attack attempts, representing 31.3%, with 1 password (3.1%) ultimately being successfully attacked. It is worth noting that although only 10 unique passwords were correctly guessed by attackers, there were 16 successful attempts involving the correct pattern, indicating that certain passwords were entered correctly more than once. To understand the rationale behind these repeated correct attempts, we conducted interviews with the 3 participants responsible for this situation. One of them mentioned that he was certain that the victim had entered that specific password, leading him to repeat the attempts. The other two participants mentioned that they couldn't think of other possibilities, so they casually entered the same password multiple times.

We selected the ASR from a previous study [95] for PIN, Pattern, and Blinkey at an attack distance of 1m and compared them with CoordAuth. The ASR for these four techniques were 18.0%, 22.5%, 4.7%, and 0.6%. It can be observed that CoordAuth has an ASR that is merely 3.3% of the PIN's ASR, 2.7% of the pattern's ASR, and 12.8% of the Blinkey's ASR, highlighting CoordAuth's robustness against shoulder surfing attack. In the subsequent interviews, we primarily asked attackers about the biggest challenges they faced during the attacks. We summarized the following points:

(1) **Difficulty in simultaneously observing head and eye movements**. Since the input process involves coordinated movements of the head and eyes, attackers need to find a way to observe eye movements while paying attention to head movements. According to participants P2 and P6, *"It's not entirely accurate to guess based solely on head movements."* P2 also highlighted, *"The user's head movements were not obvious during input, so I needed to pay extra attention to eye movements."* Our findings indicate that the requirement for head-eye

coordinated movements in our model's input mechanism poses significant challenges for attackers, often leading to their confusion and distraction.

(2) **Difficulty in determining the start and end positions**. As the start and end of the input are triggered by non-spontaneous blinking signals, it is challenging to determine the starting and ending positions of a pattern password. P2, P5 and P10 said, *"I don't know where he started and finished."* And according to P15, *"She brought her head back to a natural position at the end, but I don't know if she had completed the input."* This difficulty arises because many individuals move their heads while their eyes are closed, making it challenging for attackers to judge the starting and ending positions.

(3) **Difficulty when passwords involve diagonal lines or translational patterns**. Also according to P5, *I am not sure if his password occupies two columns or three columns." "I cannot see clearly whether his diagonal line is a 45-degree connection or a different angle."* P11 and P13 said. And it's worth mentioning that P2 and P5 also said: "I don't know if the pattern occupies the two lines above or the two lines below." This also indicates that classic password translation issues persist in VR input, especially when passwords involve diagonal or translational elements [24].

Overall, we achieved a 0.6% ASR under shoulder surfing attacks and provided reasons why CoordAuth is capable of resisting shoulder surfing. This demonstrates the robustness of CoordAuth against shoulder surfing attacks.

9 DISCUSSIONS

9.1 Feasibility of CoordAuth

9.1.1 Usability. In CoordAuth, we introduced a user-friendly and innovative pattern password entry method in VR, utilizing natural head-eye coordination for gaze input of pattern passwords. Study 1 determined the optimal UI design and the trigger mechanism for this interaction mode. Additionally, it tackled issues related to user fatigue and slow input speed associated with relying solely on head movement input. This collaborative input method is considered the most natural [66].

In simulation 2, we investigated the stability and cumulative learning characteristics of CoordAuth over time. Even without the model being updated, CoordAuth still achieved an FRR of 12.6% and an FAR of 2.6% on Day 7. In study 3, we demonstrated that CoordAuth exhibited robustness to contextual factors, with a low FRR of 4.4% when transitioning from a seated registration to a standing input posture. CoordAuth achieved a total error rate as low as 2.5% and an input time of 3.8 seconds. It achieved a lower error rate than that of patterns [24, 95] and other two-factor and biometric authentication methods [23, 95]. In terms of input time, CoordAuth was over 5 seconds faster compared to authentication methods [23, 95] designed to resist shoulder-surfing attacks. Additionally, it could match the input time of traditional PINs and pattern methods. Additionally, modern knowledge-based methods employed complex randomization mechanisms to defend against shoulder-surfing attacks [2, 28], imposing a significant physical and psychological burden on users. CoordAuth, through the integration of behavioral biometrics, eliminated the need for randomization mechanisms.

CoordAuth relies solely on natural head-eye interaction, requiring no hands, greatly simplifying the complexity of interaction and being friendly to individuals with upper limb and arm movement disorders. Furthermore, CoordAuth's interaction does not involve hand controllers, significantly enhancing the portability of VR devices.

9.1.2 Security. CoordAuth combines pattern passwords with behavioral biometrics of head and eye movements, achieving a high level of security. simulation 1 demonstrates that even in extreme cases of password collisions, CoordAuth can achieve an FRR as low as 0.88% and an FAR as low as 0.042%. This performance surpasses most previous authentication technologies, such as [95]. Additionally, to balance high accuracy while avoiding user fatigue, we recommend users perform 5 repetitions during registration, as shown in Section 6.4.1. Furthermore, given that the FRR sharply decreases with an increase in the number of VNs, as mentioned in Section 5.5, we suggest that the pattern should ideally consist of 4-5 points.

In Study 4, we conducted research on CoordAuth's robustness to shoulder surfing attack. Results indicate that out of 160 attack attempts by 16 attackers, only one attempt successfully breached CoordAuth, achieving an ASR as low as 0.6%. This outcome outperforms PIN and pattern and is superior to two-factor authentication [95], as shown in Section 8.4. In subsequent interviews with attackers, their perspectives on CoordAuth highlighted its robustness to shoulder surfing.

Additionally, compared to the previous direct application of time series classification (TSC) [33] to raw data for behavioral biometric authentication [51, 59, 85], CoordAuth explicitly provides interpretability for behavioral biometrics in a two-factor authentication system. In the discussion of Study 1, we explored the unique behavioral biometric features of users during head and eye input processes, confirming the individual variability in head and eye angle allocation [20, 60, 81, 82]. Additionally, the concepts of "head mover" and "none head mover" persist in VR environments, supporting the effectiveness of CoordAuth's behavior biometric authentication. This contributes to alleviating user privacy concerns and enhancing their trust in behavioral biometric authentication.

9.2 Head-eye Coordination Patterns

We investigated the head-eye coordination patterns during the design of CoordAuth (see Section 4.4.4). We found that many participants exhibited different head-eye coordination patterns that differed in the degree allocated (see Figure 8) in VR. This was consistent with the former work focusing on the head-eye collaboration (e.g. [20, 78, 82] unveiled the head-turning angles during head movements process, [81] indicated these differences in head movements exist in both controlled and real-world environments). However, most prior studies focused on head-eye coordination patterns based on external stimuli, whereas we contribute to research investigating head-eye coordination models during the user's input processes. The results indicate that differences in head-turning angles persist in the head-eye coordination process during typical interactions in VR environments. This insight can provide valuable inspiration for the design of interactions in VR scenarios.

10 LIMITATION AND FUTURE WORK

We acknowledged the limitations in our work, which we also regarded as future directions. First, in our study, we exclusively utilized the Quest Pro VR headset for all experiments due to limitations in laboratory equipment. While CoordAuth is designed for VR headsets, this authentication method can be extended to other head-mounted display devices, such as MR and AR. Additionally, our participant pool was restricted to college students aged 18 to 30, and we acknowledge the need for experiments involving a more diverse audience in the future.

Second, in the simulation study, due to content and article length constraints, we only investigated the stability of CoordAuth within a 7-day timeframe. Future research should further explore the long-term stability of the system. In the usability study, since VR is increasingly being used for social interactions, work, and education[27, 29], with limited use in mobile scenarios for authentication, we only validated CoordAuth for usability in sitting and standing postures. Additionally, replicating other behavioral or two-factor authentication methods proved challenging, leading us to directly reference baselines from their respective studies. A more comprehensive analysis is reserved for future research.

Last, while CoordAuth currently employs a 3x3 grid UI, the potential performance improvement of CoordAuth with a 4x4 or other UI designs is a topic worthy of discussion. Furthermore, given that CoordAuth involves gaze-based biometric features, there is potential for implicit authentication if active input is replaced with stimuli. For instance, integrating gaze-based biometric authentication into gaze calibration might be a promising direction for future research.

11 CONCLUSION

As virtual reality devices become increasingly integrated into our daily lives, designing an authentication method for these devices becomes essential. Therefore, we propose CoordAuth, a two-factor authentication method for head-mounted displays that leverages head-eye coordination. CoordAuth employs gaze interaction, utilizing eye gaze to input 3x3 pattern passwords and incorporates behavioral biometrics for authentication. This approach significantly enhances the security of the authentication process. To implement CoordAuth, we conducted a user study to investigate the impact of *FOV* on the usability of CoordAuth and explored the motion patterns of user head-eye coordination. We implemented CoordAuth on the Quest Pro VR headset and validated its performance under password collision scenarios. The results showed that CoordAuth achieved an FRR of 0.88% and an FAR of 0.042% for 24 participants. A 7-day simulation verified the memorability and long-term stability of CoordAuth. In the usability study, we assessed CoordAuth's robustness to two contextual factors, input time and error rate, demonstrating its strong usability. In the final study, we evaluated CoordAuth's resistance to shoulder surfing. Overall, we believe CoordAuth is a promising authentication method suitable for current virtual reality devices.

REFERENCES

[1] 2007. Eye Tracking Methodology; Theory and Practice. *Qualitative Market Research: An International Journal* 10, 2 (2007), 217–220. https://doi.org/10.1108/13522750710740862

[2] 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

[3] Ashwin Ajit, Natasha Kholgade Banerjee, and Sean Banerjee. 2019. Combining Pairwise Feature Matches from Device Trajectories for Biometric Authentication in Virtual Reality Environments. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 9–97. https://doi.org/10.1109/AIVR46125.2019.00012

[4] Mohammad Al-Rubaie and J. Morris Chang. 2016. Reconstruction Attacks Against Mobile-Based Continuous Authentication Systems in the Cloud. *IEEE Transactions on Information Forensics and Security* 11, 12 (2016), 2648–2663. https://doi.org/10.1109/TIFS.2016.2594132

[5] Muhammad Mujtaba Asad, Aisha Naz, Prathamesh P. Churi, and Mohammad Mehdi Tahanzadeh. 2021. Virtual Reality as Pedagogical Tool to Enhance Experiential Learning: A Systematic Literature Review. *Education Research International* (2021). https://api.semanticscholar.org/CorpusID:244276119

[6] Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2018. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the Workshop on Communication by Gaze Interaction* (Warsaw, Poland) *(COGAIN '18)*. Association for Computing Machinery, New York, NY, USA, Article 1, 9 pages. https://doi.org/10.1145/3206343.3206349

[7] Fadi Boutros, Naser Damer, Kiran Raja, Raghavendra Ramachandra, Florian Kirchbuchner, and Arjan Kuijper. 2020. Iris and periocular biometrics for head mounted displays: Segmentation, recognition, and synthetic data generation. *Image and Vision Computing* 104 (2020), 104007. https://doi.org/10.1016/j.imavis.2020.104007

[8] Davina Bristow, John-Dylan Haynes, Richard Sylvester, Chris Frith, and Geraint Rees. 2005. Blinking Suppresses the Neural Response to Unchanging Retinal Stimulation. *Current biology : CB* 15 (08 2005), 1296–300. https://doi.org/10.1016/j.cub.2005.06.025

[9] Pan Chan, Tzipora Halevi, and Nasir Memon. 2015. Glass OTP: Secure and Convenient User Authentication on Google Glass. (2015), 298–308. https://doi.org/10.1007/978-3-662-48051-9_22

[10] Supply Chain Game Changer. 2020. *Virtual Reality (VR) is Enhancing E-commerce Shopping.* https://supplychaingamechanger.com/how-virtual-reality-vr-is-drastically-enhancing-the-e-commerce-shopping-experience-infographic/ Accessed on: 2024-01-28.

[11] Jagmohan Chauhan, Hassan Jameel Asghar, Mohamed Ali Kaafar, and Anirban Mahanti. 2016. Gesture-based Continuous Authentication for Wearable Devices: the Google Glass Case. arXiv:1412.2855 [cs.CR]

[12] Yuxin Chen, Zhuolin Yang, Ruben Abbou, Pedro Lopes, Ben Y. Zhao, and Haitao Zheng. 2021. User Authentication via Electrical Muscle Stimulation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Yokohama</city>, <country>Japan</country>, </conf-loc>) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 6, 15 pages. https://doi.org/10.1145/3411764.3445441

[13] Geumhwan Cho, Jun Ho Huh, Junsung Cho, Seongyeol Oh, Youngbae Song, and Hyoungshick Kim. 2017. SysPal: System-Guided Pattern Locks for Android. In *2017 IEEE Symposium on Security and Privacy (SP)*. 338–356. https://doi.org/10.1109/SP.2017.61

[14] Carlo Colombo and Alberto Del Bimbo. 1997. Interacting through eyes. *Robotics and Autonomous Systems* 19, 3 (1997), 359–368. https://doi.org/10.1016/S0921-8890(96)00062-0 Intelligent Robotic Systems SIRS'95.

[15] Timothy R. Derrick and J.M. Thomas. 2004. Chapter 7. Time-Series Analysis: The cross-correlation function. In *Innovative Analyses of Human Movement*, N. Stergiou (Ed.). Human Kinetics Publishers, Champaign, Illinois, 189–205. https://dr.lib.iastate.edu/handle/20.500.12876/52528 Posted with permission..

[16] James Diebel. 2006. Representing Attitude : Euler Angles , Unit Quaternions , and Rotation Vectors. https://api.semanticscholar.org/CorpusID:16450526

[17] Wenxin Feng, Jiangnan Zou, Andrew Kurauchi, Carlos H Morimoto, and Margrit Betke. 2021. HGaze Typing: Head-Gesture Assisted Gaze Typing. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 11, 11 pages. https://doi.org/10.1145/3448017.3457379

[18] Edward G. Freedman. 2008. Coordination of the eyes and head during visual orienting. *Experimental Brain Research* 190, 4 (October 1 2008), 369–387. https://doi.org/10.1007/s00221-008-1504-8

[19] Edward G. Freedman and David L. Sparks. 2000. Coordination of the eyes and head: movement kinematics. *Experimental Brain Research* 131, 1 (March 1 2000), 22–32. https://doi.org/10.1007/s002219900296

[20] James H. Fuller. 1992. Head movement propensity. *Experimental Brain Research* 92, 1 (December 1992), 153–164. https://doi.org/10.1007/BF00230391

[21] Markus Funk, Karola Marky, Iori Mizutani, Mareike Kritzler, Simon Mayer, and Florian Michahelles. 2019. LookUnlock: Using Spatial-Targets for User-Authentication on HMDs. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312959

[22] Yang Gao, Wei Wang, Vir V. Phoha, Wei Sun, and Zhanpeng Jin. 2019. EarEcho: Using Ear Canal Echo for Wearable Authentication. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 81 (sep 2019), 24 pages. https://doi.org/10.1145/3351239

[23] 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)*. 277–285. https://doi.org/10.1109/VR.2019.8797862

[24] Ceenu George, M. Khamis, Emanuel von Zezschwitz, Marinus Burger, Henri Schmidt, Florian Alt, and Heinrich Hussmann. 2017. Seamless and Secure VR: Adapting and Evaluating Established Authentication Systems for Virtual Reality. https://api.semanticscholar.org/CorpusID:6671814

[25] H. H. L. M. Goossens and A. J. Van Opstal. 1997. Human eye-head coordination in two dimensions under different sensorimotor conditions. *Experimental Brain Research* 114, 3 (May 1 1997), 542–560. https://doi.org/10.1007/PL00005663

[26] Sven-Thomas Graupner, Michael Heubner, Sebastian Pannasch, and Boris M. Velichkovsky. 2008. Evaluating requirements for gaze-based interaction in a see-through head mounted display. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (Savannah, Georgia) *(ETRA '08)*. Association for Computing Machinery, New York, NY, USA, 91–94. https://doi.org/10.1145/1344471.1344495

[27] Harry Halpin, David J. Zielinski, Rachael Brady, and Glenda Kelly. 2008. Exploring Semantic Social Networks Using Virtual Reality. In *The Semantic Web - ISWC 2008*, Amit Sheth, Steffen Staab, Mike Dean, Massimo Paolucci, Diana Maynard, Timothy Finin, and Krishnaprasad Thirunarayan (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 599–614. https://doi.org/10.1007/978-3-540-88564-1_38

[28] Mudarabilli Harshini, Padigala Lakshman Sai, Singam Chennamma, Thanuja, Alavalapati Goutham Reddy, and Hyun Sung Kim. 2021. Easy-Auth: Graphical Password Authentication using a Randomization Method. In *2021 IEEE Latin-American Conference on Communications (LATINCOM)*. 1–6. https://doi.org/10.1109/LATINCOM53176.2021.9647825

[29] Sandra Helsel. 1992. Virtual Reality and Education. *Educational Technology* 32, 5 (1992), 38–42. http://www.jstor.org/stable/44425644

[30] Chandra Hermawan Heruatmadja, Meyliana, Achmad Nizar Hidayanto, and Harjanto Prabowo. 2023. Biometric as Secure Authentication for Virtual Reality Environment: A Systematic Literature Review. In *2023 International Conference for Advancement in Technology (ICONAT)*. 1–7. https://doi.org/10.1109/ICONAT57137.2023.10080713

[31] Katharina Anna Maria Heydn, Marc Philipp Dietrich, Marcus Barkowsky, Götz Winterfeldt, Sebastian von Mammen, and Andreas Nüchter. 2019. The Golden Bullet: A Comparative Study for Target Acquisition, Pointing and Shooting. In *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. 1–8. https://doi.org/10.1109/VS-Games.2019.8864589

[32] Daniel Hintze, Rainhard Dieter Findling, Sebastian Scholz, and René Mayrhofer. 2014. Mobile Device Usage Characteristics: The Effect of Context and Form Factor on Locked and Unlocked Usage. *Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia* (2014). https://api.semanticscholar.org/CorpusID:7092460

[33] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. 2019. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery* 33, 4 (01 07 2019), 917–963. https://doi.org/10.1007/s10618-019-00619-1

[34] Robert J. K. Jacob. 1995. *Eye tracking in advanced interface design.* Oxford University Press, Inc., USA, 258–288.

[35] John M. Jones, Reyhan Duezguen, Peter Mayer, Melanie Volkamer, and Sanchari Das. 2021. A Literature Review on Virtual Reality Authentication. In *Human Aspects of Information Security and Assurance*, Steven Furnell and Nathan Clarke (Eds.). Springer International Publishing, Cham, 189–198.

https://doi.org/10.1007/978-3-030-81111-2_16

[36] Mohamed Khamis, Ludwig Trotter, Ville Mäkelä, Emanuel von Zezschwitz, Jens Le, Andreas Bulling, and Florian Alt. 2018. CueAuth: Comparing Touch, Mid-Air Gestures, and Gaze for Cue-based Authentication on Situated Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 174 (dec 2018), 22 pages. https://doi.org/10.1145/3287052

[37] Sehee Kim and Euichul Lee. 2018. Periocular Biometric Authentication Methods in Head Mounted Display Device. https://api.semanticscholar.org/CorpusID:251398816

[38] Alexander Kupin, Benjamin Moeller, Yijun Jiang, Natasha Kholgade Banerjee, and Sean Banerjee. 2019. Task-Driven Biometric Authentication of Users in Virtual Reality (VR) Environments. (2019), 55–67. https://doi.org/10.1007/978-3-030-05710-7_5

[39] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. 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, 1–14. https://doi.org/10.1145/3173574.3173655

[40] Pınar Kürtünlüoğlu, Beste Akdik, and Enis Karaarslan. 2022. Security of Virtual Reality Authentication Methods in Metaverse: An Overview. arXiv:2209.06447 [cs.CR]

[41] Michael Francis Land and Benjamin W. Tatler. 2009. Looking and Acting: Vision and eye movements in natural behaviour. https://api.semanticscholar.org/CorpusID:141184531

[42] Sugang Li, Ashwin Ashok, Yanyong Zhang, Chenren Xu, Janne Lindqvist, and Macro Gruteser. 2016. Whose move is it anyway? Authenticating smart wearable devices using unique head movement patterns. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 1–9. https://doi.org/10.1109/PERCOM.2016.7456514

[43] Yantao Li, Hailong Hu, and Gang Zhou. 2019. Using Data Augmentation in Continuous Authentication on Smartphones. *IEEE Internet of Things Journal* 6, 1 (2019), 628–640. https://doi.org/10.1109/JIOT.2018.2851185

[44] 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

[45] 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

[46] Dillon Lohr, Samuel-Hunter Berndt, and Oleg Komogortsev. 2018. An implementation of eye movement-driven biometrics in virtual reality. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (Warsaw, Poland) *(ETRA '18)*. Association for Computing Machinery, New York, NY, USA, Article 98, 3 pages. https://doi.org/10.1145/3204493.3208333

[47] Dillon J Lohr, Samantha Aziz, and Oleg Komogortsev. 2020. Eye Movement Biometrics Using a New Dataset Collected in Virtual Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 40, 3 pages. https://doi.org/10.1145/3379157.3391420

[48] Shiqing Luo, Anh Nguyen, Chen Song, Feng Lin, Wenyao Xu, and Zhisheng Yan. 2020. OcuLock: Exploring Human Visual System for Authentication in Virtual Reality Head-mounted Display. In *Network and Distributed System Security Symposium*. https://api.semanticscholar.org/CorpusID:211266673

[49] Päivi Majaranta and Andreas Bulling. 2014. *Eye Tracking and Eye-Based Human–Computer Interaction*. Springer London, London, 39–65. https://doi.org/10.1007/978-1-4471-6392-3_3

[50] Jesus Martínez-Navarro, Enrique Bigné, Jaime Guixeres, Mariano Alcañiz, and Carmen Torrecilla. 2019. The Influence of Virtual Reality in E-commerce. *Journal of Business Research* 100 (2019), 475–482. https://doi.org/10.1016/j.jbusres.2018.10.054

[51] 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* (<conf-loc>, <city>Honolulu</city>, <state>HI</state>, <country>USA</country>, </conf-loc>) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3334480.3382799

[52] Florian Mathis, Kami Vaniea, and Mohamed Khamis. 2021. RepliCueAuth: Validating the Use of a Lab-Based Virtual Reality Setup for Evaluating Authentication Systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Yokohama</city>, <country>Japan</country>, </conf-loc>) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 534, 18 pages. https://doi.org/10.1145/3411764.3445478

[53] Florian Mathis, John H. Williamson, Kami Vaniea, and Mohamed Khamis. 2021. Fast and Secure Authentication in Virtual Reality Using Coordinated 3D Manipulation and Pointing. *ACM Trans. Comput.-Hum. Interact.* 28, 1, Article 6 (jan 2021), 44 pages. https://doi.org/10.1145/3428121

[54] Florian Mathis, John H. Williamson, Kami Vaniea, and Mohamed Khamis. 2021. Fast and Secure Authentication in Virtual Reality Using Coordinated 3D Manipulation and Pointing. *ACM Trans. Comput.-Hum. Interact.* 28, 1, Article 6 (jan 2021), 44 pages. https://doi.org/10.1145/3428121

[55] Mark Roman Miller, Fernanda Herrera, Hanseul Jun, James A. Landay, and Jeremy N. Bailenson. 2020. Personal Identifiability of User Tracking Data During Observation of 360-degree VR Video. *Scientific Reports* 10, 1 (15 10 2020), 17404. https://doi.org/10.1038/s41598-020-74486-y

[56] 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)*. 311–316. https://doi.org/10.1109/VRW50115.2020.00070

[57] 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

[58] Aunnoy Mutasim, Anil Ufuk Batmaz, Moaaz Hudhud Mughrabi, and Wolfgang Stuerzlinger. 2022. Performance Analysis of Saccades for Primary and Confirmatory Target Selection. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology* (<conf-loc>, <city>Tsukuba</city>, <country>Japan</country>, </conf-loc>) *(VRST '22)*. Association for Computing Machinery, New York, NY, USA, Article 18, 12 pages. https://doi.org/10.1145/3562939.3565619

[59] 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

[60] Brian S. Oommen, Ryan M Smith, and John S. Stahl. 2004. The influence of future gaze orientation upon eye-head coupling during saccades. *Experimental Brain Research* 155 (2004), 9–18. https://api.semanticscholar.org/CorpusID:491888

[61] Benjamin I. Outram, Yun Suen Pai, Tanner Person, Kouta Minamizawa, and Kai Kunze. 2018. Anyorbit: orbital navigation in virtual environments with eye-tracking. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (Warsaw, Poland) *(ETRA '18)*. Association for Computing Machinery, New York, NY, USA, Article 99, 5 pages. https://doi.org/10.1145/3204493.3209579

[62] Yun Suen Pai, Benjamin I. Outram, Benjamin Tag, Megumi Isogai, Daisuke Ochi, and Kai Kunze. 2017. GazeSphere: navigating 360-degree-video environments in VR using head rotation and eye gaze. In *ACM SIGGRAPH 2017 Posters* (Los Angeles, California) *(SIGGRAPH '17)*. Association for Computing Machinery, New York, NY, USA, Article 23, 2 pages. https://doi.org/10.1145/3102163.3102183

[63] Panagiota Papadopoulou. 2007. Applying Virtual Reality for Trust-Building E-commerce Environments. *Virtual Reality* 11, 2 (2007), 107–127. https://doi.org/10.1007/s10055-006-0059-x

[64] Ge Peng, Gang Zhou, David T. Nguyen, Xin Qi, Qing Yang, and Shuangquan Wang. 2017. Continuous Authentication With Touch Behavioral Biometrics and Voice on Wearable Glasses. *IEEE Transactions on Human-Machine Systems* 47, 3 (2017), 404–416. https://doi.org/10.1109/THMS.2016.2623562

[65] 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

[66] Yuan Yuan Qian and Robert J. Teather. 2017. The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 91–98. https://doi.org/10.1145/3131277.3132182

[67] Yuan Yuan Qian and Robert J. Teather. 2018. Look to Go: An Empirical Evaluation of Eye-Based Travel in Virtual Reality. In *Proceedings of the 2018 ACM Symposium on Spatial User Interaction* (Berlin, Germany) *(SUI '18)*. Association for Computing Machinery, New York, NY, USA, 130–140. https://doi.org/10.1145/3267782.3267798

[68] Luis Quintero, Panagiotis Papapetrou, Jaakko Hollmén, and Uno Fors. 2021. Effective Classification of Head Motion Trajectories in Virtual Reality Using Time-Series Methods. In *2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 38–46. https://doi.org/10.1109/AIVR52153.2021.00015

[69] Cynthia E. Rogers, Alexander W. Witt, Alexander D. Solomon, and Krishna K. Venkatasubramanian. 2015. An approach for user identification for head-mounted displays. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers* (Osaka, Japan) *(ISWC '15)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/2802083.2808391

[70] Sam Royston, Connor DeFanti, and Ken Perlin. 2016. A Collaborative Untethered Virtual Reality Environment for Interactive Social Network Visualization. arXiv:1604.08239 [cs.HC]

[71] Sohrab Saeb, Cornelius Weber, and Jochen Triesch. 2011. Learning the optimal control of coordinated eye and head movements. *PLoS Comput Biol* 7, 11 (November 2011), e1002253. https://doi.org/10.1371/journal.pcbi.1002253

[72] Stefan Schneegass, Youssef Oualil, and Andreas Bulling. 2016. SkullConduct: Biometric User Identification on Eyewear Computers Using Bone Conduction Through the Skull. 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, 1379–1384. https://doi.org/10.1145/2858036.2858152

[73] B Shaghayeghfard, Amir Ahmadi, N Maroufi, and J Sarrafzadeh. 2016. Evaluation of forward head posture in sitting and standing positions. *European spine journal* 25 (2016), 3577–3582.

[74] 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

[75] Ludwig Sidenmark and Hans Gellersen. 2019. Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* 27, 1, Article 4 (dec 2019), 40 pages. https://doi.org/10.1145/3361218

[76] 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)*. 518–523. https://doi.org/10.1109/ICCES48766.2020.9137914

[77] Robert Skerjanc and Siegmund Pastoor. 1997. New generation of 3D desktop computer interfaces. In *Stereoscopic Displays and Virtual Reality Systems IV*, Scott S. Fisher, John O. Merritt, and Mark T. Bolas (Eds.), Vol. 3012. International Society for Optics and Photonics, SPIE, 439 – 447. https://doi.org/10.1117/12.274485

[78] John S. Stahl. 1999. Amplitude of human head movements associated with horizontal saccades. *Experimental Brain Research* 126, 1 (April 1 1999), 41–54. https://doi.org/10.1007/s002210050715

[79] Sophie Stephenson, Bijeeta Pal, Stephen Fan, Earlence Fernandes, Yuhang Zhao, and Rahul Chatterjee. 2022. SoK: Authentication in Augmented and Virtual Reality. In *2022 IEEE Symposium on Security and Privacy (SP)*. 267–284. https://doi.org/10.1109/SP46214.2022.9833742

[80] Mei Suzuki, Ryo Iijima, Kazuki Nomoto, Tetsushi Ohki, and Tatsuya Mori. 2023. PinchKey: A Natural and User-Friendly Approach to VR User Authentication. In *Proceedings of the 2023 European Symposium on Usable Security* (<conf-loc>, <city>Copenhagen</city>, <country>Denmark</country>, </conf-loc>) *(EuroUSEC '23)*. Association for Computing Machinery, New York, NY, USA, 192–204. https://doi.org/10.1145/3617072.3617122

[81] Zachary Thumser, Brian Oommen, Igor Kofman, and John Stahl. 2008. Idiosyncratic variations in eye-head coupling observed in the laboratory also manifest during spontaneous behavior in a natural setting. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale* 191 (09 2008), 419–34. https://doi.org/10.1007/s00221-008-1534-2

[82] Zachary C. Thumser and John S. Stahl. 2009. Eye–head coupling tendencies in stationary and moving subjects. *Experimental Brain Research* 195, 3 (May 1 2009), 393–401. https://doi.org/10.1007/s00221-009-1803-8

[83] D. Tweed, B. Glenn, and T. Vilis. 1995. Eye-head coordination during large gaze shifts. *Journal of Neurophysiology* 73, 2 (February 1995), 766–779. https://doi.org/10.1152/jn.1995.73.2.766

[84] Visualise. 2020. *Virtual Reality in Healthcare.* https://visualise.com/virtual-reality/virtual-reality-healthcare Accessed on: 2024-01-28.

[85] Xue Wang and Yang Zhang. 2021. Nod to Auth: Fluent AR/VR Authentication with User Head-Neck Modeling. 1–7. https://doi.org/10.1145/3411763.3451769

[86] Dhruv Kumar Yadav, Beatrice Ionascu, Sai Vamsi Krishna Ongole, Aditi Roy, and Nasir Memon. 2015. Design and Analysis of Shoulder Surfing Resistant PIN Based Authentication Mechanisms on Google Glass. In *Financial Cryptography and Data Security*, Michael Brenner, Nicolas Christin, Benjamin Johnson, and Kurt Rohloff (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 281–297. https://doi.org/10.1007/978-3-662-48051-9_21

[87] Akiko Yamazaki, Keiichi Yamazaki, Yoshinori Kuno, Matthew Burdelski, Michie Kawashima, and Hideaki Kuzuoka. 2008. Precision timing in human-robot interaction: coordination of head movement and utterance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 131–140. https://doi.org/10.1145/1357054.1357077

[88] Yukang Yan, Yingtian Shi, Chun Yu, and Yuanchun Shi. 2020. HeadCross: Exploring Head-Based Crossing Selection on Head-Mounted Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 35 (mar 2020), 22 pages. https://doi.org/10.1145/3380983

[89] Yukang Yan, Chun Yu, Xin Yi, and Yuanchun Shi. 2018. HeadGesture: Hands-Free Input Approach Leveraging Head Movements for HMD Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 198 (dec 2018), 23 pages. https://doi.org/10.1145/3287076

[90] Shanhe Yi, Zhengrui Qin, Ed Novak, Yafeng Yin, and Qun Li. 2016. GlassGesture: Exploring head gesture interface of smart glasses. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*. 1–9. https://doi.org/10.1109/INFOCOM.2016.7524542

[91] Xin Yi, Leping Qiu, Wenjing Tang, Yehan Fan, Hewu Li, and Yuanchun Shi. 2022. DEEP: 3D Gaze Pointing in Virtual Reality Leveraging Eyelid Movement. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 3, 14 pages. https://doi.org/10.1145/3526113.3545673

[92] Xin Yi, Shuning Zhang, Ziqi Pan, Louisa Shi, Fengyan Han, Yan Kong, Hewu Li, and Yuanchun Shi. 2023. Squeez'In: Private Authentication on Smartphones based on Squeezing Gestures. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 532, 15 pages. https://doi.org/10.1145/3544548.3581419

[93] 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)*. 458–460. https://doi.org/10.1109/APCCAS.2016.7804002

[94] Ruide Zhang, Ning Zhang, Changlai Du, Wenjing Lou, Y. Thomas Hou, and Yuichi Kawamoto. 2017. AugAuth: Shoulder-surfing resistant authentication for augmented reality. In *2017 IEEE International Conference on Communications (ICC)*. 1–6. https://doi.org/10.1109/ICC.2017.7997251

[95] 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

[96] Yongpan Zou, Meng Zhao, Zimu Zhou, Jiawei Lin, Mo Li, and Kaishun Wu. 2018. BiLock: User Authentication via Dental Occlusion Biometrics. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 152 (sep 2018), 20 pages. https://doi.org/10.1145/3264962