Analysis of Reflexive Eye Movements for Fast Replay-Resistant Biometric Authentication

IVO SLUGANOVIC, University of Oxford, UK
MARC ROESCHLIN, University of Oxford, UK
KASPER B. RASMUSSEN, University of Oxford, UK
IVAN MARTINOVIC, University of Oxford, UK

Eye tracking devices have recently become increasingly popular as an interface between people and consumer-grade electronic devices. Due to the fact that human eyes are fast, responsive, and carry information unique to an individual, analyzing person's gaze is particularly attractive for rapid biometric authentication. Unfortunately, previous proposals for gaze-based authentication systems either suffer from high error rates, or require long authentication times.

We build upon the fact that some eye movements can be reflexively and predictably triggered, and develop an interactive visual stimulus for elicitation of reflexive eye movements that support the extraction of reliable biometric features in a matter of seconds, without requiring any memorization or cognitive effort on the part of the user. As an important benefit, our stimulus can be made unique for every authentication attempt and thus incorporated in a challenge-response biometric authentication system. This allows us to prevent replay attacks, which are possibly the most applicable attack vectors against biometric authentication.

Using a gaze tracking device, we build a prototype of our system and perform a series of systematic user experiments with 30 participants from the general public. We thoroughly analyze various system parameters and evaluate the performance and security guarantees under several different attack scenarios. The results show that our system matches or surpasses existing gaze-based authentication methods in achieved equal error rates (6.3%) while achieving significantly lower authentication times (5 seconds).

CCS Concepts: • **Security and privacy** → **Authentication**; **Biometrics**; *Systems security*; • **Human-centered computing** → *Interaction devices*;

ACM Reference Format:
Ivo Sluganovic, Marc Roeschlin, Kasper B. Rasmussen, and Ivan Martinovic. 2018. Analysis of Reflexive Eye Movements for Fast Replay-Resistant Biometric Authentication. *ACM Trans. Priv. Sec.* 21, 5, Article 1 (December 2018), 31 pages. https://doi.org/10.1145/3281745

1 INTRODUCTION

Eye tracking devices capture precise position and movement of the human cornea on a millisecond scale. This, in turn, allows determining the exact location of one's gaze on a screen or on surrounding objects. Since analyzing eye behavior can give insight into our internal cognitive processes and even predict conditions such as autism [30], eye trackers have been used in neurophysiological research for over a century, but until recently their use in everyday life was limited due to prohibitive equipment costs.

2471-2566/2018/12-ART1 $15.00
https://doi.org/10.1145/3281745

However, the speed and responsiveness of eye movements strongly motivate their use as an attractive input channel for human-computer interaction[1]; as a result, recent years have brought a sharp reduction in retail prices of eye tracking devices. While dedicated trackers can be purchased for as little as $100 [1], eye tracking capabilities are also being added to consumer products such as laptops [29], cars [39], tablets, and mobile phones [37]. Given the diverse advantages and applications of eye tracking, its widespread expansion into our everyday lives is only likely to continue.

As we demonstrate in the following sections, tracking a user's gaze is particularly suitable for fast and low-effort user authentication, especially in scenarios where keyboard input is not available. Eye movements exhibit traits distinctive enough that classification algorithms (e.g., [13]) can reliably discern among a large group of individuals. However, despite the advantages, exploiting eye movements for user authentication remains a challenging topic. As we summarize in Section 3, previous work on gaze-based authentication achieves either high error rates (e.g., EER above 15%) or long authentication times (e.g., above 20 seconds). One likely explanation for some of these outcomes are overly complex visual stimuli that result in voluntarily triggered eye movements which are highly dependent on a user's current cognitive state.

In this paper, we show how the reflexive physiological behavior of human eyes can be used to build fast and reliable biometric authentication systems. We utilize the fact that, even though most eye movements are elicited voluntarily, specific *reflexive* movements can be actively triggered using a simple visual stimulus. Measuring and analyzing millisecond-scale characteristics of reflexive eye movements provides several important benefits. Users' eyes naturally and spontaneously react to the shown stimulus so they do not need to follow any instructions or memorize additional information. As a result, elicitation of reflexive behavior requires lower cognitive load and is very fast. This, in turn, enables keeping authentication times short while at the same time extracting large amounts of useful biometric data and achieving low error rates.

Finally, we show a crucial advantage of exploiting reflexive eye movements for authentication: by employing a challenge-response type of protocol, such systems can provide security even under a stronger adversary model than the ones usually considered for biometrics. One of the obstacles for widespread use of biometric authentication in our daily lives is the fact that most biometrics can be captured and replayed relatively easily. Examples include spoofing image recognition systems with photographs from social media and spoofing fingerprint recognition using copies of fingerprints left behind on everyday items. If the visual stimulus can be made unique for each authentication attempt, then the elicited responses will accordingly be different, but still, include user-specific characteristics. By always choosing a new *challenge* (randomly generated stimulus) and verifying if the *response* (measured eye movements) corresponds to it, our authentication system can assert that the biometric sample is indeed fresh. Other biometric systems have to make special provisions to achieve a level of spoofing and replay protection. For example, sophisticated fingerprint readers measure additional attributes like temperature and moisture in order to determine liveness. Our gaze-based authentication system achieves these guarantees without requiring any other information besides the recording of a user's eye movements.

2 BACKGROUND ON EYE MOVEMENTS

We start by giving a short background of the human visual system and describe the necessary terminology related to eye movements; this allows us to introduce main concepts that motivate our research and guide the design of the system in the following sections.

[1]This article is an extension of a previous conference paper [45].

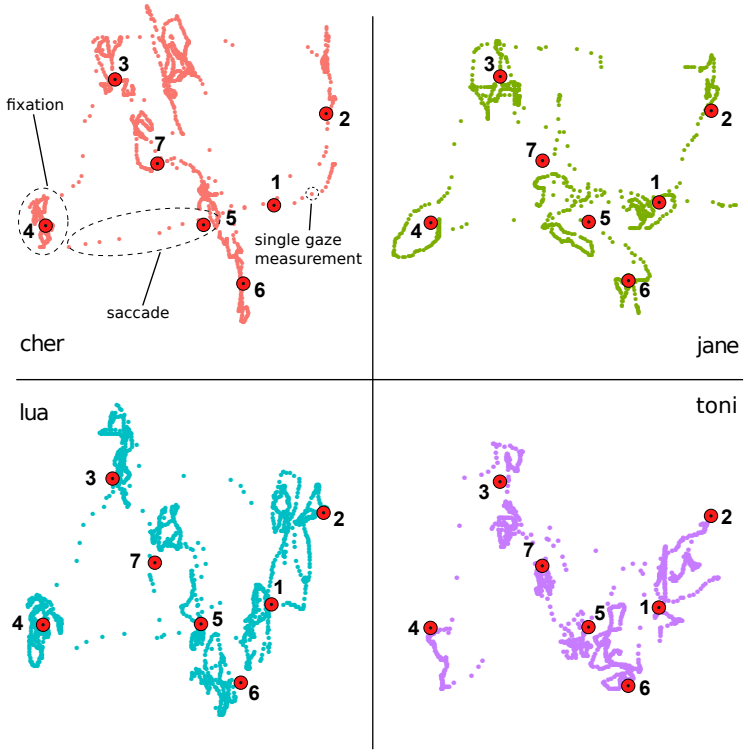


Fig. 1. Eye movements of four users as a response to the same visual stimulus recorded using SMI RED 500 device [43]. Fixations are visible as clustered areas, while saccades consist of series of dots that depict paths. Larger red dots show the positions at which the visual stimulus was shown. Despite their distinct characteristics, all four gaze paths closely match the positions of the stimulus.

Even when one's gaze is firmly fixated on a single stimulus, human eyes are never completely still, as they hey are constantly making hundreds of micro-movements per second. These micro-movements are interlaced with about 3-5 larger movements every second, that amount to more than 100,000 eye movements during the course of one day [2]. During standard visual tasks, such as object search or scene perception, our eyes alternate between *fixations* and *saccades*. Fixations are used to maintain the visual focus on a single stimulus, while saccades reorient the eye to focus the gaze on a next desired position. Saccades are rapid eye movements and they are considered to be the fastest rotational movement of any external part of our body, reaching angular velocities of up to 900 degrees per second, and usually lasting between 20 ms and 100 ms [22]. In Figure 1, fixations can be seen as areas of large numbers of closely grouped points, while saccades consist of series of more spread recordings that depict fairly straight paths.

Reflexive vs Voluntary Saccades. When a salient change happens in our field of vision, our eyes naturally reorient on the target, since this is a necessary first step to provide information for further higher-level cognitive processes [36]. These externally elicited saccades happen reflexively and are considered to be an effortless neuronal response, requiring very low cognitive load from the user. After the stimulus onset, a corresponding *reflexive* saccade is initiated rapidly, with usual latencies of less than 250 ms [47]. In contrast, voluntary saccadic movements were shown to have larger

mean latencies (above 300 ms) which are additionally influenced by different internal and external factors [47].

The analysis of eye movements has been part of medical research for more than a century since it offers valuable information of our cognitive and visual processing [36], [9], [3]. Keeping the goal of reliable biometric authentication in mind, we are interested in extracting and combining multiple characteristics of human eye movements for which there exists supporting research that they offer stable individual differences between users. For example, Castelhano et al. [8] examine stable individual differences in characteristics of both saccades and fixations and provides support for their stable use in biometric authentication. Saccades were also used in [13] to enable stable authentication and identification. Furthermore, several researchers have analyzed eye behavior features of trained shooters [12], professional baseball players [4] and other specific groups of individuals [18], and reported measurable differences between their eye movements characteristics.

Given that reflexive reactions are less dependent on momentary conscious states of an individual than conscious actions, it is expected that biometrics based on reflexive characteristics offer more stable authentication. Furthermore, taking into account the advantage in faster elicitation times, the goal of our research is to design a stimulus that supports the use of reflexive saccades for biometric authentication. For example, prior research has shown that saccade latencies depend on the dominant eye [32] of the individual, which is a stable characteristic and provides strong motivation for using saccade latencies for classification. Finally, it was shown that *saccade latency* varies if anticipation (temporal expectancy) is present [46]. This provides an argument for randomizing the stimulus that is shown to users.

3 RELATED WORK

While different eye tracking methods have been used in medical research for over a century, their use in security is fairly recent. A review paper by Zhang et. al. [49] provides an overview of authentication methods and systems proposed before 2010, while Saeed [42] gives a more recent comparison of methods and results of gaze-based authentication systems proposed up to the year 2013. According to Zhang et. al. [49], existing work in user identification and authentication can be roughly divided into two categories: **1)** using gaze tracking as a human-computer interface (control channel) to support standard security primitives and **2)** using characteristics of the gaze patterns to extract individual biometric traits that enable distinguishing between different users.

In the first line of research, individuals use their eyes to prove their identity by naturally and covertly inputting secret information such as passwords [35], [6] or specific patterns on the screen [5], [11], [31]. Using eyes as a control channel has several advantages, such as prevention of shoulder-surfing and smudge attacks. Unfortunately, these approaches usually share the negative characteristics of passwords, such as requiring the users to learn a procedure or remember and recall different pieces of information, as well as still being susceptible to eavesdropping and replay attacks.

Our work belongs to the second, biometric approach, which uses the characteristics of individual's gaze patterns to discriminate between different users. Such authentication systems usually come with the general benefits, but also challenges typical to biometrics: they usually require no memorization, prevent sharing of credentials and offer high usability, but at the same time, they suffer from irrevocability, which renders replay attacks a serious threat if even a single user's biometric sample is acquired by an attacker.

Biometric approaches to gaze-based authentication can be further divided into two subcategories: those that rely on high-level characteristics of user's gaze patterns (*where* and *what* the user is looking at), and those that analyze the low-level traits of *how* the user's eyes are moving.

Table 1. Comparison to existing biometric authentication systems based on eye-movements

Analysis of	Stimulus	Ref.	Time [s]	EER [%]	Notes
High-level Features					
Scan paths + arch densities	Human faces	[7]	17	25	
Distribution of areas of interest	Human faces	[17]	10	36.1	
Graph matching	Human faces	[40]	4	30	
Fixation density maps	Movie trailer	[41]	60	14	
Low-level Features					
Cepstrum transform of raw signal	Dot, fixed inter-stimulus	[28]	8	N/A	FAR 2%, FRR 22%
Oculomotor plant model	Dot, horizontal sequence	[33]	21	N/A	FAR 5.4%, FRR 56.6%
Scan paths and fixation features	Read section of text	[20]	60	23	
Fixation and saccade features	Read section of text	[21]	60	16.5	
Liveness detection	Dot, horizontal sequence	[34]	100	18	Focus on liveness detection
Fixation and saccade features	Read a poem	[16]	60	2.09-10.16	Feat. selection using ICC
Non-linear DTW	Enter PIN using mouse	[26]	20	6.82	Fusion of mouse and eye movements
Fixation and saccade features	**Dot, interactive**	**this paper**	**5**	**6.3-10.47**	Replay attacks FAR: 0.06%

High-level Characteristics. The first approach is motivated by hypotheses that users exhibit individual behavior during certain tasks, and thus extracts high-level characteristics of users' responses while the users are instructed to freely look at videos, photos of faces, or other specific types of stimuli. Prior work includes analysis of scan paths and arch densities [7], areas of interest on human faces [17], graph matching [40] and fixation density maps [41].

As summarized in Table 1, existing work in this category mostly achieves Equal Error Rates higher than 15%, which is likely due to complex features being more dependent on varying cognitive and physiological states of the user. Furthermore, in order to acquire sufficient data to extract complex features, these systems often require long authentication times (measured in tens of seconds!), so further improvements are needed before they can be applied to real-world systems.

Low-level Characteristics. On the other hand, motivated by psychological and neurophysiological research [8] that suggests stable differences between users [50], several authors researched systems that use low-level characteristics of users' eye movements as features for discrimination, such as eye movement velocity profiles, sizes of fixation areas, saccade latencies, etc.

Kasprowski is one of the first authors to start systematically researching the low-level characteristics of user's gaze for authentication. In his initial paper [28] and corresponding Ph.D. thesis [23], he proposes using features such as the distance between the left and right eye-gaze, Fourier and wavelet transforms of the raw gaze signal and average velocity directions. The used stimulus consists of 9 LED lights arranged in a 3x3 grid, where the position of the single active light changes according to a fixed, equally timed sequence, regardless of the user's gaze. An experimental study showed half total error rates of close to 12%, but with relatively high false reject rates of 22%. In relation to our proposal, such stimulus also leads to eliciting some reflexive saccades, but as Table 1 shows, it results in longer authentication times and higher error rates. This is likely due to periods of time where the user has already gazed at the light but is still waiting for the position of the active LED to change. The authors propose, organize and describe two yearly competitions in eye movements verification and identification using their datasets [25], [27], which have further increased the research interest in gaze-based authentication and its fusion with other biometric modalities [26].

Komogortsev proposes modeling the physiological properties of individuals' oculomotor plant [33] during multiple horizontal saccades and using the estimated model parameters as features for

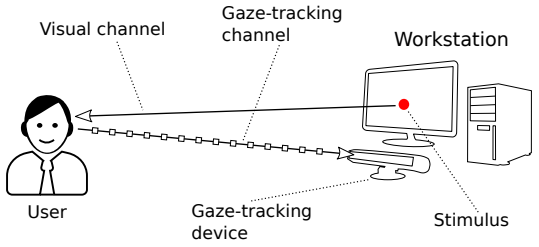


Fig. 2. System model. The workstation uses data acquired by the gaze tracker and user's biometric template to make the authentication decision. The adversary has read-write access to the gaze channel. The visual channel is authenticated and therefore read-only.

classification. Related work by Holland et al. [20] provides an insight into performances of multiple features such as fixation counts and durations during text reading and combines these two approaches to achieve an EER of 23%, while the newer research [21] provides an additional analysis of 13 classification features based on fixations and saccades and achieves an EER of 16.5%.

A recent work by a similar group of authors proposes the use of Intra-Class-Correlation as part of feature selection that specifically optimizes for temporal feature stability [16]. Evaluating several datasets that include multiple biometric modalities, authors report reductions in comparison to previously achieved EER rates, even achieving an EER of 2.01% for an eye tracking dataset in which participants were instructed to read a poem.

Continuous Authentication. In contrast to point-of-entry authentication, in which the classifier must make a single decision about user's identity as quickly as possible, Eberz et al. [13] propose using 21 low-level characteristics of eye movements to continuously re-authenticate users, regardless of their current task, and thus detect intruders whose eye movements differ from the legitimate user over a period of time. For one parameter combination, the authors achieve Equal Error Rates of 7.8% when 40 seconds are chosen as a period before making the first decision. Furthermore, using one-class SVM classification, they are able to detect all but 1% of attackers when the classifier is allowed to make decisions across the span of 30 s [14]. However, due to the requirement of task independence in a continuous authentication scenario, potential replay attacks remain a serious vulnerability. If the attacker is able to capture even a very short recording of legitimate user's gaze, he can continuously rewind and replay it back to the gaze tracking device, and this causes the system to (correctly!) accept the received eye movements as coming from a legitimate user.

4 ASSUMPTIONS AND GOALS

We start by defining the system and adversary model used throughout this paper; we then state the design goals for the visual stimulus and the authentication system.

System Model. We assume the general settings of a user authenticating to a workstation in an office scenario throughout the course of a normal workday. A simple visualization of the system model is shown in Figure 2. The user authenticates to a workstation using a gaze tracking device by looking at a visual stimulus displayed on the screen. The workstation uses data acquired by the gaze tracker and a user's biometric template to make the authentication decision.

A legitimate user is one who is enrolled in the authentication system. The enrollment happens in a secure scenario, where the legitimate user authenticates to the workstation using another

authentication method. During enrollment, the user is shown several visual stimuli and the workstation uses the corresponding recordings of the user's gaze to create a biometric template used for identity verification.

The interaction takes place through three different channels. The *visual channel* is an authenticated channel from the workstation to the user that consists of a screen that displays information, and the *gaze tracking channel* from the user to the gaze tracker allows the workstation to determine characteristics about the user's eyes, including where he is looking on the screen, as well as capture the reflexive eye movements described in Section 2.

The workstation itself cannot be modified or forced to run unintended code.

Adversary Model. The adversary's goal is to impersonate a legitimate user and successfully authenticate to the workstation. The adversary can freely choose his victim from the set of enrolled users. Since he can observe both the visual and gaze channels, the adversary has access to the biometric data from previous authentication attempts by the victim.
We focus on two different types of attacks that the adversary can perform:

- *Impersonation attack.* The adversary tries to gain access to the workstation by positioning himself in front of the gaze tracking device. This is the most common way of evaluating biometric authentication systems, and is usually reported in terms of false reject (FRR) and false accept rates (FAR) as well as equal error rates (EER).
- *Replay attack.* The adversary targets a specific user and replays his previously recorded authentication attempt to the authentication system. This can be done either at the sensor level (e.g. by using a mechanical eye replica) or by bypassing the gaze tracking sensor completely and injecting the recorded samples between the workstation and the sensor.

Biometrics are non-revocable, and we are surrounded by sensors that can be used to steal and replay biometric data. Therefore, we believe that modeling an attacker as having access to legitimate user's previous biometric measurements is a realistic and necessary assumption. Most static biometrics, such as fingerprints or face recognition [5], cannot provide security under such assumptions; the ability to prevent replay attacks is one of the major strengths of our scheme since simply replaying an acquired sample is arguably the most accessible attack vector for most biometrics.

We do not consider a targeted adversary who is able to model and generate arbitrary artificial samples of a user's eye movements in an interactive manner. As we further discuss in Section 10, such attacks require significantly higher levels of complexity and effort from the adversary; a level of commitment against which most biometric systems cannot provide security guarantees.

Design Goals.

- *Low cognitive load:* The system should pose low cognitive load on the users. Ideally, users should not be required to remember credentials, carry tokens, or learn new procedures. Moreover, the cooperation required from the user should be as effortless as possible.
- *Fast:* The duration of a single authentication attempt should be as short as possible.
- *Resistance against replay:* The system should make it difficult for an adversary to replay acquired biometric samples and thereby successfully authenticate.

5 SYSTEM ARCHITECTURE

The proposed authentication system works as follows. The workstation shows an interactive visual stimulus on the screen (we refer to it as *gaze-challenge*). Simultaneously, the gaze tracking device captures eye movements of the user as he watches the screen (*gaze-response*), which the

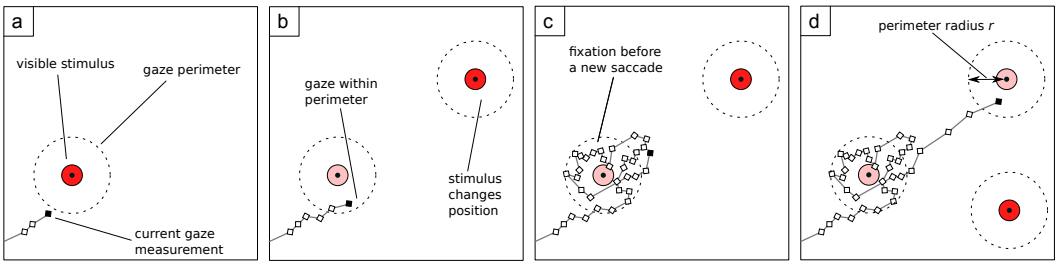


Fig. 3. A visualization of the stimulus for reflexive saccade elicitation. At any given time, only a single red dot is shown; previous positions are shown in this figure to help the reader. Shortly after a red dot appears on the screen (a), a user's visual system starts a reflexive saccade to shift the gaze (dotted path) towards its position. Several milliseconds later, as the user's gaze enters the invisible perimeter around the stimulus (dashed circles), the dot is considered successfully gazed and momentarily changes its position. Before a new saccade start, there is usually a fixation lasting 100-250 ms, during which the visual system processes new input information (saccade latency). In (d), the presented dot is again successfully gazed, and once more changes its position.

workstation uses to adapt the stimulus in real time. Finally, the workstation makes a decision about the user's identity and verifies if the received gaze-response corresponds to the shown gaze-challenge, asserting that the captured eye movements are indeed fresh.

5.1 Stimulus for Reflexive Saccade Elicitation

To achieve stated design goals, a visual stimulus should satisfy several requirements. It should elicit responses that are sufficiently distinctive to allow discrimination between different users. The response should not require high cognitive effort and should not depend on a user's momentary cognitive state. The stimulus should be *unpredictable* to prevent habituation: seeing an image for the first time will likely result in a different response than seeing it for the second and the consecutive times [46]. Finally, in order to allow fast authentication, the stimulus duration should be as short as possible.

Design. Considering that reflexive behavior is more stable and less dependent on a user's transient internal cognitive states than voluntary behavior, our goal is to design a stimulus which allows eliciting and measuring individual traits of user's reflexive saccadic responses. Reflexive saccades are triggered by salient objects that appear in one's field of view; thus our stimulus consists of presenting a single red dot on a dark screen that changes position multiple times. As shown in Figure 3, a user's eyes respond to the change by eliciting a reflexive saccade which reorients the gaze towards the dot. Every time the position of the dot changes, the visual system responds by initiating a new reflexive saccade. Due to saccade latency, this happens after a period of 100-200 ms during which the visual system processes new information.

Ideally, our stimulus should elicit the maximal number of reflexive saccades in a given period of time, and this highly depends on the frequency with which the position of the dot changes. If this frequency is too high, user's eyes will not be given sufficient time to perform a full saccade. If it is too low, the user might get tired of looking at a static screen and start voluntary saccadic movements. Furthermore, each user is slightly different, so there might not exist a unique frequency at all. Using an interactive stimulus ensures an optimum between these trade-offs by interactively changing the location of the dot as soon as the user *successfully gazes* the dot, i.e., when a user's gaze enters a perimeter of radius r around the dot's center. This results in eliciting the maximal

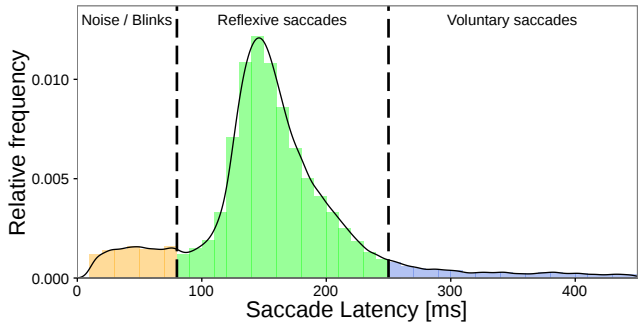


Fig. 4. Relative frequency of saccade latencies for gaze-responses used in this paper. Latencies are computed as the duration between the stimulus change and the start of subsequent saccadic movement. Vertical lines discriminate between reflexive and other types of saccades; latencies of reflexive saccades are usually lower than 250 ms, in contrast to latencies of voluntary saccades that are over 250 ms. Values under 80 ms are likely the result of noise or blinks, or voluntary saccades initiated well before the stimulus change [47].

number of full saccades in any given time interval, which ensures that the user's visual system receives an outside stimulus change as often as possible, consequently reducing the elicitation of voluntary saccades which depend on his current cognitive state. To ensure that the stimulus terminates even if the user is not looking at the screen, the dot is considered to be *unsuccessfully gazed* and moves to the next position after a specific period of $D_{\max}$ milliseconds has passed. This process continues for all N stimulus positions that constitute a gaze-challenge.

Basing an authentication system on reflexive movements provides additional benefits: taking into account that reflexive behavior is significantly harder to consciously control, an adversary is less likely to be able to successfully imitate another user's characteristics. Most importantly, because of the natural and effortless tendency of the human visual system to keep *"catching"* the red dot, the response to such visual stimulus is fully reflexive: users neither need to follow specific instructions nor invest high cognitive effort —*their eyes do the work themselves.*

Effectiveness of the Stimulus. In order to evaluate how effectively our designed stimulus elicits reflexive behavior, we compute saccade latencies for a total of 991 legitimate authentication attempts from the experimental dataset used throughout this paper. Since each of the measurements represents a gaze-response to a stimulus with 25 different positions for the dot, in total, this sums up to analyzing close to 25,000 captured saccades.

Figure 4 shows the distribution and categorization of the measured saccade latencies, dividing them into reflexive saccades, voluntary saccades and saccadic movement caused by blinks. Given that latencies under 80 ms have only been recorded in specifically designed conditions, e.g., when the stimulus position and onset are predictable [46], we consider them to likely be the result of blinks or noise [47]. Remaining latencies predominantly fall below 250 ms, the threshold that characterizes reflexive saccades [47]. This lets us conclude that the stimulus does indeed elicit primarily reflexive behavior.

5.2 Authentication Protocol

We now use the proposed stimulus as a building block in a challenge-response protocol for biometric user authentication that is secure against replay attacks. At the end of the protocol execution, the

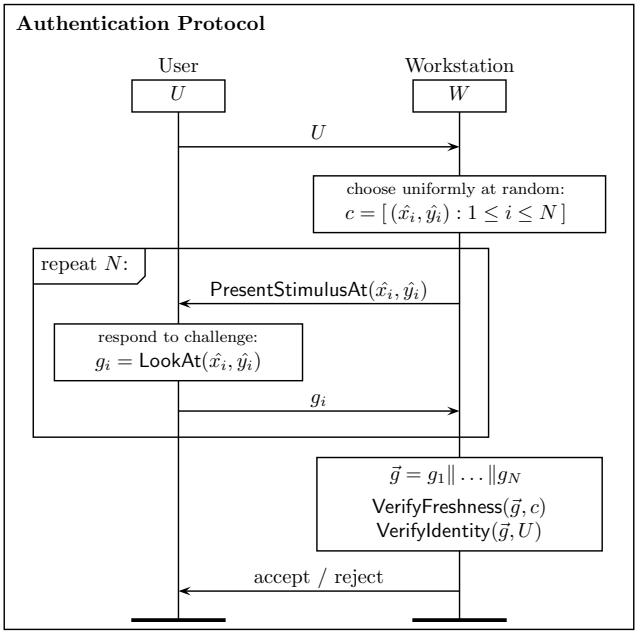


Fig. 5. Biometric challenge-response authentication protocol. User claims his identity, after which the workstation generates a fresh *gaze-challenge* c, an ordered list of positions in which the stimulus is shown. User looks at (LookAt) a screen where the stimulus is shown at N positions $\{(\hat{x}_i, \hat{y}_i)\}$. Meanwhile, the gaze tracking device records the user's gaze paths g_i for all stimulus positions that constitute the *gaze-response* $\vec{g}$. The workstation verifies the freshness of $\vec{g}$, and finally verifies that the biometric features extracted from $\vec{g}$ correspond to the claimed identity.

workstation knows if the user whose identity is claimed is at the moment present in front of the gaze tracking device. To that goal, the workstation must ensure that two properties hold:

Freshness. Freshness of the received biometric data can be ensured by always showing a different randomly generated visual stimulus (gaze-challenge) to which every response will differ in a verifiable way.

Correct Identity. The user has the ability to generate biometric data that corresponds to the claimed user's template which was created during enrollment.

The protocol for local biometric authentication is shown in Figure 5. After the user claims his identity, the workstation generates a fresh visual stimulus, which we refer to as *gaze-challenge* (c_W) in the rest of the paper. c_W consists of a set of n randomly chosen coordinates, which uniquely define the interactive stimulus described in Section 5.1. As the gaze-challenge is presented to the user, his eyes reflexively respond with a series of eye movements, which constitute the *gaze-response* (r_U). Gaze-response is recorded by the gaze tracking device through the gaze channel.

In order to accept or reject the user's authentication request, the workstation performs two verification steps: VerifyFreshness and VerifyIdentity. These are described in detail in Sections 5.3 and 5.4, respectively.

In the final message, the workstation notifies the user if he has been granted or denied access to the system.

5.3 VerifyFreshness

As described in Section 5.1, each visual stimulus is uniquely defined by a list of N coordinates; therefore, it is possible to always present a different random gaze-challenge to the user. Since no visual stimulus shown to users is ever reused, in order to verify the freshness of the response, it suffices to verify if the received gaze-response closely corresponds to the freshly presented gaze-challenge. As visualized in Figure 3, if some gaze-response was recorded while specific gaze-challenge was shown to the user, then the user's eye movements should closely resemble the order and positions in which the stimulus dot was shown. This is visible in Figure 1: despite differences in gaze patterns of different users, all of them correspond to the locations of the stimulus dot.

The system determines if the gaze-response is indeed fresh by ensuring that the user timely gazed at the majority of the stimulus positions. After a stimulus dot is shown in one of the N positions, it is considered *successfully gazed* only if one of the subsequent measurements of the user's gaze position falls within a radius of R pixels from the center of the stimulus dot. Otherwise, if no gaze measurement falls within its radius after $D_{\max}$ milliseconds, a position is considered to be *unsuccessfully gazed* and the dot moves to the next position:

$$g_i \coloneqq \left[(x_j, y_j) : t_i \le t_j < t_i + D_{\max} \right]$$

$$\text{succ. gazed}\,(\hat{x_i}, \hat{y_i}) \iff \exists (x, y) \in g_i : \; \|(x, y) - (\hat{x_i}, \hat{y_i})\|_2 \le R$$

In order to decide on the freshness of the received gaze-response, the system checks if the ratio of successfully gazed stimulus positions is greater or equal to a chosen percentage threshold T.

As the threshold T increases, the possibility that an adversary successfully replays an old recording of a legitimate user's gaze decreases. On the other hand, this also results in more legitimate attempts failing freshness verification, e.g., because of inaccurate gaze measurements. We evaluate the security guarantees of different thresholds T in Section 8.3 and analyze the impact of different values for the threshold D on the classification performance and authentication times in Section 9.3.

5.4 VerifyIdentity

If the received gaze-response passes the freshness verification, the system finally validates that it truly originated from the user whose identity was claimed at the beginning of the authentication. The received gaze-response is first used as input to compute a set of specific feature values that are idiosyncratic and support stable classification between users. Next, the computed features are used as an input to a two-class classifier which is created during user enrollment. The classifier determines whether the calculated features more likely belong to the user whose identity was claimed, or to some internal or external attacker. As the last step, the authentication system makes a final decision and notifies the user of acceptance or rejection.

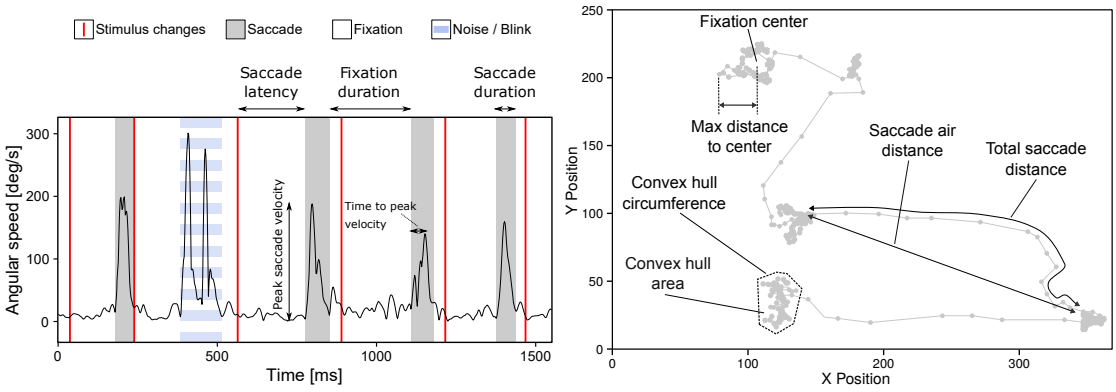
Next section describes the details about the features that we use and how we train the user classifiers.

6 FEATURES FOR GAZE CLASSIFICATION

This section describes the process of extracting individual characteristics from user's gaze-response and training a classifier that can uniquely discriminate between future responses of the same user and any other user's gaze patterns.

6.1 Feature Extraction

Feature extraction is the process of converting the raw measurements into a lower dimensional set of meaningful data that retain most of the useful information to distinguish different output classes



(a) Temporal Features (b) Spatial Features

Fig. 6. Visualization of features on (a) temporal and (b) spatial plots of raw gaze tracking data. In Subfigure (a), the moment when stimulus changes position is depicted with a vertical red line. The period depicted with horizontal stripes is physiologically impossible for a human eye to perform and is caused by a blink. We remove such artifacts with methods described in Section 6.

when used as input to the classification algorithm. When considered in the context of eye tracking, the feature extraction process should take as input the time-stamped positions of one's gaze during the authentication attempt, and compute a significantly lower dimensional set of feature values that allow discrimination between different individuals.

Saccade and Fixation Detection. Following the discussion in Section 5.1, the expected user's gaze will consist of multiple repetitions of a reflexive saccade, which redirects one's gaze at the new position of the red stimulus dot, followed by a fixation that lasts until one's visual system detects the change in stimuli location (saccade latency). Consequently, the first step in the feature extraction process is to split the raw gaze measurement into intervals of saccades and fixations, which we later use to compute specific characteristics of one's eye movements.

We implement an adaptive algorithm [38] that estimates the level of noise in the data to determine the thresholds used to classify the measurements into periods of fixations and saccades. The detection is mainly based on angular velocities and accelerations, taking into account the known physiological limitations of eye movements. As seen in Figure 6a, the algorithm also detects eye movement recordings that could not have been generated by a human eye under known physiological constraints, and are usually the result of blinking. Given that the mean duration of a single blink is close to 200 ms [22], and that head movements and gazes outside of the screen area usually last even longer, it is important to denoise the raw data before further analysis. These artifacts are filtered based on research that shows the peak angular velocity of the human eye to lie between 700 and 900 deg/sec [22], and the peak angular acceleration to not cross 100000 deg/sec^2.

Having grouped the measurements as belonging either to a fixation or a saccade, we proceed to calculate a set of feature values for each recorded gaze sample, ignoring those measurements that are classified as noise by the procedure.

Using Median Values. Since the authentication system always shows a fresh visual stimulus (defined by the N positions of the stimulus dot), the computed features should not be influenced by the positions of the dots in the gaze-challenge, as this would lower the probability that the

Table 2. Relative Mutual Information (RMI_{ID}) of potential features that were considered during system design. Rows in bold numbers show the 16 features that were selected for subsequent classification. Selection was made based on RMI values, except where stated differently. #1, #2, #3, #20, and #33 are included to allow comparison.

#	Feature Description	RMI_{ID}	Comment
1	Mean Y coord. of the corneal reflex position (left eye)	0.3267	Excluded as a static feature
2	Mean the pupil diameter (left eye)	0.2871	Excluded as a static feature
3	Mean X coord. of the corneal reflex position (left eye)	0.2381	Excluded as a static feature
4	**Median air distance vs total distance ratio**	**0.1846**	
5	**Median fixation duration**	**0.1575**	
6	**Median average fixation velocity**	**0.1540**	
7	**Density of fixation convex hulls**	**0.1474**	
8	Mean fixation duration	0.1456	Excluded due to similarity with #5
9	**Median saccade latency**	**0.1453**	
10	Total time of authentication attempt	0.1447	Excluded as similar to 11
11	**Average time per stimulus**	**0.1447**	
12	Mean saccade duration	0.1403	Excluded due to similarity with #15
13	**Mean velocity**	**0.1397**	
14	**Average distance per stimulus**	**0.1382**	
15	**Median saccade duration**	**0.1363**	
16	**Median fixation convex hull area**	**0.1362**	
17	**Median saccade average velocity**	**0.1360**	
18	**Median fixation convex hull perimeter**	**0.1343**	
19	**Median fixation max velocity**	**0.1337**	
20	Ratio of successful gazes	0.1321	Excluded as a static feature
21	**Median saccade max velocity**	**0.1283**	
22	**Median saccade max acceleration**	**0.1257**	
23	Mean saccade latency	0.1227	Excluded due to similarity with #7
24	**Median fixation max distance**	**0.1226**	
25	Median saccade air distance	0.1215	
26	Median fixation Y span	0.1208	
27	Median fixation X span	0.1089	
28	Median fixation X and Y span ratios	0.0983	
29	Median saccade X span	0.0937	
30	Median saccade X and Y span ratios	0.0864	
31	Median saccade Y span	0.0790	
32	Median saccade time to max velocity	0.0698	
33	Random variable	0.0567	Included only for comparison

user reauthenticates with a fresh challenge. As Figure 6 shows, each gaze-response consists of intermixed periods of saccades and fixations and each such period allows us to compute multiple features. However, we are interested in computing a single set of identifiable feature values for a given gaze-response as a whole, irrespective of the number of elicited saccades and fixations; to that end, and to reduce the effect of noise, feature values for a single user's gaze-response (authentication attempt) are computed as the median of feature values computed on individual saccades or fixations in that gaze measurement.

6.2 Feature Quality

In order to support secure and reliable authentication, the features should ideally be chosen so that they are as varied for different users and as similar as possible when computed for multiple authentication attempts of the same user. Extracting stable and distinctive features from real-world user behavior data, especially when the measurement is noisy, as is the case with eye trackers, is a challenging task.

Since all potential features do not contribute the same amount of distinguishing power, we follow a semi-automated approach to select the optimal set of features for the authentication system. Initially, we explore a broader set of fixation and saccade traits, in addition to a range of other metrics that measure overall characteristics of the gaze path.

Relative Mutual Information. In order to choose the final subset of features that we use for classification, we compute the Relative Mutual Information (RMI), a measure that quantifies the reduction in the entropy of the final outcome (user's identity in the context of biometric authentication) as a result of knowing the value of an individual feature [15]. More precisely, RMI can be expressed as:

$$\mathrm{RMI}_{\mathrm{ID}}(F) := \frac{\mathrm{MI}(\mathrm{ID}, F)}{H(\mathrm{ID})} = \frac{H(\mathrm{ID}) - H(\mathrm{ID}|F)}{H(\mathrm{ID})}$$

Mutual Information between two variables A and B is denoted as $\mathrm{MI}(A, B)$, while $H(A)$ represents the entropy of variable A.

Based on RMI, we test the features on randomly chosen subsets of the dataset, measure their classification performance, and exclude those that do not achieve satisfactory results. The chosen features that have a clear spatial or temporal representation are shown in Figure 6, while their RMI values can be found in Table 2.

Chosen Features. As the RMI values in Table 2 show, medians of **average angular speeds** during fixations or saccades, as well as the **duration** of fixations are among the most specific features we tested. This finding is congruent with the feature assessment conducted by Eberz et al. [13, 14], where pairwise speeds exhibit the highest relative mutual information, only outperformed by some of their static features, such as pupil diameter. Contrary to their results, we identify **saccade curviness** (ratio of air distance and total distance of a saccade) and **saccade latency** to be the features that yield the most distinguishing power. Furthermore, we identify several discriminative features based on computing a convex hull of all measurements in a fixation: **convex hull and circumference**, as well as **fixation density**, defined as the ratio of the convex hull area and the number of gaze measurements in that fixation.

Using Dynamic Features. Given the focus on evaluating the feasibility of using reflexive behavior for authentication, this paper only uses dynamic characteristics of eye movements for classification. We thus consciously forego using several features that most gaze tracking devices provide, such as an estimate of user's pupil size and the distances between the user's eyes. In prior work, pupil size was shown to be one of the more discriminative features for gaze-based authentication systems [14], however, the authors raise valid concerns that an adversary could manipulate his pupil size, e.g., by controlling the lighting conditions. Despite potential classification improvements, in this paper we employ only features that can be extracted from raw coordinates of the user's gaze. We further discuss relaxing this assumption in Section 10.

6.3 User Enrollment

During enrollment, several gaze-responses are used to train a dedicated 2-class classifier that the system will use as user's identity verifier. In any subsequent authentication attempt, the same set of feature values are extracted from any gaze-response and the classifier makes a decision whether the values correspond to the claimed user or not.

Besides legitimate user's gaze-responses, the enrollment procedure requires a similarly sized set of gaze-responses belonging to other users that are labeled as negative samples during classifier training.

Choice of the Classifier. In this analysis, we mainly use a Support Vector Machine (SVM) [10] with Radial Basis Function (RBF) kernel as the classifier, since SVMs are known to provide strong classification results for non-linear data sets. In Section 9.4 we also evaluate and discuss several

other classification algorithms for multiple test configurations, confirming that SVMs consistently achieve the lowest error rates on our data.

SVMs with RBF kernels are fully defined by two hyper-parameters: **1)** C, which controls the trade-off between the penalty of incorrect classification and the margin of the decision hyperplane, and **2)** σ, which is a parameter that defines the scale of the radial basis function. The optimal pair of hyper-parameter values is chosen from a predetermined set of potential values, based on the evaluation that uses 5-fold cross-validation: for each pair of potential hyperparameters, 80% of the enrollment data is used to train the resulting classifier, while the remaining 20% of the enrollment data is used to evaluate the classification performance; this is repeated five times.

The pair of hyperparameters that results in strongest classification performance is finally used to derive the final user classifier which is used in future authentication.

7 DATA ACQUISITION

In order to experimentally evaluate the performance of the proposed system and protocol, we developed a prototype and ran a series of user experiments to gather data for analysis.

7.1 System Prototype

Setup. Our prototype setup is composed of a gaze tracking device (SMI RED 500 [43]), a 24-inch LED screen and a desktop computer. The generation of the visual stimulus and the gaze sampling was performed by a custom-built software library that controls the gaze tracking device. We implemented procedures that take care of the internal calibration of the gaze tracker, the validation of the calibration accuracy, and the visual presentation of the stimulus, as well as the acquisition of the gaze samples captured by the gaze tracker.

Parameters. For each authentication attempt, the prototype generated a visual challenge consisting of $N = 25$ random positions on which the stimulus will be shown. The distance between users' eyes and the gaze tracking device (positioned directly underneath the screen) was approximately 70 cm. Red stimulus dot is shown on a plain dark background, with a diameter of 0.7 cm (0.95°). In order to detect that a dot was successfully gazed, we used a perimeter radius of $r = 1.4$ cm (1.25°). If not successfully gazed, the dot changed position after $D_{\max} = 1000$ ms.

7.2 User Experiments

Experiment Design. For the purpose of assessing feasibility and performance of the proposed system, we conducted a series of user experiments that reflect the scenario described in Section 4. We refer to a series of consecutive authentication attempts with the same participant as one session. Each session lasted about 10 minutes and included a briefing and 15 authentication attempts. Before participant's first session we generated a calibration profile that was reused during all subsequent sessions with that participant. To analyze the performance of our system, both from the perspective of a user and an attacker, we divided the participants into two groups: legitimate users who have completed the enrollment procedure, and external attackers, whose gaze characteristics were not known to the system.

In order to show that our system can successfully authenticate users over the course of a normal workday (without re-calibration), we require each enrolled user to take part in a minimum of three (up to four) sessions. The first two sessions are five minutes apart and mimic a legitimate user leaving his desk to take a break or use the restroom. All subsequent sessions are at least 6 hours apart. Participants acting as external attackers are only invited to one session where they are asked to impersonate a legitimate user, i.e., the system uses the calibration profile and biometric template

of the chosen legitimate user. Every external attacker tries to authenticate as 5 different legitimate users, at least 3 times per user. In their last session, legitimate users were asked to act as internal attackers and each performed a minimum of 15 attempts of impersonating other users, analogously to external attackers.

Test Population. Experimental data was acquired from a total of 30 participants aged 21 to 58 who were recruited from the general public through public advertisements, email lists, and social media. The only requirement was a minimum age of 18. The test population consists of 7 women and 23 men. Out of the 30 recruited participants, 22 participants were enrolled as legitimate users and 8 participants represented external attackers whose gaze characteristics were not known to the system. The acquired data set consists of a total of 1602 gaze-responses: 1021 authentication attempts by legitimate users and 581 simulated attack attempts by either internal or external attackers.

Participants were told that their eye movements will be recorded for the purpose of evaluating the feasibility of distinguishing individuals based on their behavioral gaze-based biometrics. They then signed a written consent form in accordance with the experiment ethics review approved by the University's research ethics committee, reference number SSD/CUREC1A/14-226. Names have been replaced with pseudonyms.

Participants who do not have normal vision wore contact lenses or were asked to remove their glasses. This was done to remove the possibility that classification relies on potential specific characteristics of recorded gaze when glasses are worn. For the same reason, lighting conditions were not changed during all experiment sessions.

8 SYSTEM EVALUATION

We now experimentally evaluate the proposed system with respect to the design goals stated in Section 4.

8.1 Varying the Challenge Complexity N

One of the defining parameters of the proposed system is N, the number of stimulus positions in a single gaze-challenge. We first analyze the effect that varying N has on authentication time and overall user classification performance. Incrementing N directly increases the complexity of gaze-challenge, thus requiring more time to respond to the visual stimulus. At the same time, larger N should allow the system to extract more stable features and thus achieve stronger classification results. On the other hand, as N decreases, both the authentication time and the classification performance are likely to decline.

Setup. Since all user experiments were run with gaze-challenges that had $N = 25$ stimulus dot positions, we can evaluate the classifier performance in a scenario where gaze-challenges consist of $K < N$ positions by simulating that the stimulus presentation and gaze recording stopped after the K-th position was gazed. Such an adapted dataset is constructed by only considering gaze measurements that were recorded before the $(K + 1)$-th stimulus position is shown.

The classification performance for each K and for each user is estimated by computing an Equal Error Rate (EER) while performing a five-fold cross-validation of the individual classifiers as follows. In each of five repetitions, four out of five folds of the legitimate user's authentication attempts are provided as enrollment data for user enrollment that was performed as described in Section 6. The remaining fold was used to evaluate classifier performance against other users' authentication attempts as negative samples. The resulting EER for any K is computed as an average across all five folds of all individual users' classifiers for that K.

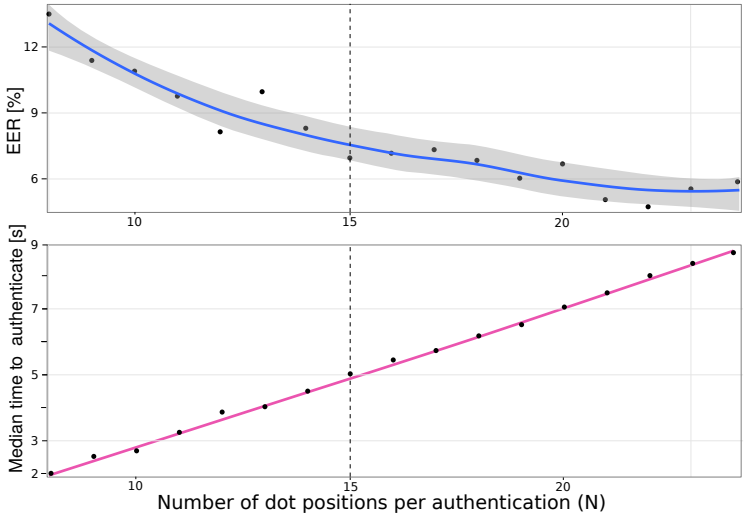


Fig. 7. Measured authentication time and EER as a function of gaze-challenge complexity N. As N increases from 8 to 24, the EER reduces from above 12% to under 6%, while at the same time, the median time to authenticate grows linearly from 2 seconds to about 9 seconds. The vertical line depicts a scenario where 15 positions are used in a challenge: the median authentication time is around 5 seconds, while the EER is close to 7%.

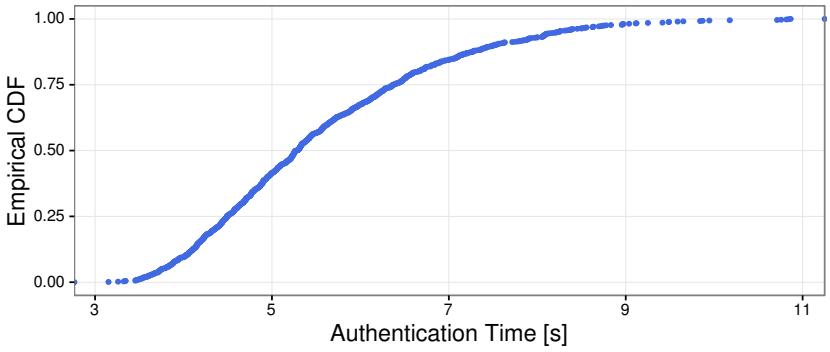


Fig. 8. Empirical cumulative distribution function for the duration of all measured authentication attempts when $N = 15$. Close to 50% of the attempts took less than 5 seconds, while more than 80% of the attempts lasted less than 7.5 seconds.

Results. We show the effect of varying N on authentication time and classification performance in Figure 7. The median time for a single authentication attempt grows linearly from 2 seconds for 8 stimulus positions, to about 9 seconds for 24 stimulus positions. At the same time, the overall EER of the classification falls from around 12% when only 8 stimulus positions are used, to a level of 6% when 24 stimulus positions are used in a challenge.

Since $N = 15$ shows a balanced trade-off between classification performance and median authentication time, we use this value to report results in the remainder of the analysis. In order to provide a more comprehensive estimate of the time required for the majority of users to authenticate than just median, in Figure 8 we show a cumulative density function of the authentication times

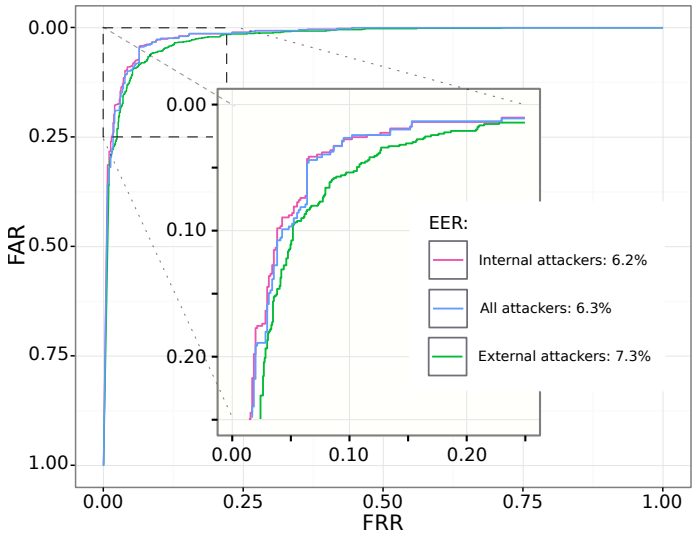


Fig. 9. The ROC curves that show authentication performance under impersonation attacks. Red and green curves represent only internal and external attackers, while blue curve shows the overall combined performance. The EER for internal attackers equals 6.2%, while for external attackers it is expectedly slightly higher and amounts to 7.3%. The overall EER for all attackers is 6.3%.

for all users when $N = 15$. The figure shows that half of the users authenticate in 5 seconds or less, while the authentication for more than 80% of the users takes less than 7.5 seconds. As we discuss in Section 10, these times are favorable to previous related work in gaze-based authentication.

8.2 Impersonation Attacks

Recall that, in an impersonation attack, the attacker targets a specific user with the goal of responding to the gaze-challenge posed by the system, and successfully impersonating the legitimate user in order to gain access. The attacker is permitted to use the gaze-based authentication system in any way he wishes, such as purposely moving or altering the angle of his head to try to increase the chance of gaining access.

As described in Section 7.2, we purposely design the user experiments to simulate this type of attack as closely as possible: all participants were asked to perform multiple "attack attempts", in which they falsely claimed some other user's identity and tried to authenticate with the gaze calibration profile of the legitimate user loaded by the system.

Setup. For each user, we perform a five-fold cross-validation to estimate the performance of the system under such attacks. We enroll the user as described in Section 6, using four out of five folds of legitimate user's samples, and then evaluate the performance of the whole authentication system on the remaining one fifth of the legitimate user's gaze-responses that were not used for enrollment. During the evaluation, legitimate user's samples are labeled as positive, while all attack attempts that other users made while pretending to be the legitimate user are labeled as negative. We consider an authentication attempt accepted by the system only if it passes both the identity verification and the freshness verification. For freshness verification, we use a threshold $T = 50\%$.

Besides overall performance, we also separately evaluate two disjunct subsets of the attack attempts: those originating from external attackers, who are unknown to the system, and those

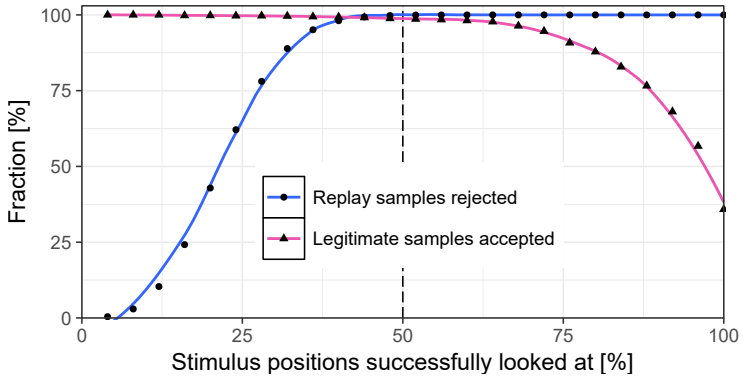


Fig. 10. Performance of the freshness verification procedure depending on the chosen threshold T. As we change the required percentage of successfully gazed stimuli to classify a gaze sample as "fresh" from 0% to 50%, the ratio of successfully detected replay attempts rises from 0 to close to 1. At the same time, the ratio of successfully classified fresh attempts starts declining as the required threshold increases over 60%, showing almost perfect results for the thresholds between 40% and 60%.

originating from internal attackers, whose previous authentication attempts might have been used as negative samples during enrollment.

Results. We show the system performance against impersonation attacks as an ROC curve in Figure 9. Since individual user classifiers output a probability that a given sample belongs to the respective legitimate user, we can achieve different classification performance by varying the threshold above which a sample is considered legitimate. As this threshold increases, so does the likelihood of falsely rejecting a legitimate user (FRR) increase, but at the same time, the likelihood of falsely accepting an attacker (FAR) decreases. Different combinations of FAR and FRR values for three attack scenarios (internal, external, and all attackers) are shown in Figure 9. For all three scenarios, it is possible to achieve low FAR values (under 5%) if FRR is increased closer to 10% and vice-versa.

The Equal Error Rate (EER) is defined as the rate at which FRR and FAR equalize. Given that EER is the single measure most commonly used to compare classification performance, we also use it throughout the rest of the paper. The reported overall system EER is computed using a single, shared, decision threshold for all classification decisions across all users' classifiers.

As expected, in terms of EER, the system achieves slightly stronger performance against internal attackers (6.2% EER) than external attackers (7.3% EER). Overall, the system achieves an EER of 6.3% for impersonation attacks; as we discuss in Section 3, this result is preferable to any previously reported performance of gaze-based authentication systems.

8.3 Replay Attacks

Recall from Section 5.3 that in order to prevent reuse of biometric data, the system verifies that the received gaze-response corresponds to the presented gaze-challenge, i.e., that the user successfully gazed at no less than a chosen percentage T of the stimulus positions presented during authentication.

The result of verifying the freshness of a received response does not depend on the claimed identity during authentication, but only on the positions of the dot in the visual stimulus. Therefore, in order to provide a more comprehensive estimate of the distinctiveness of a challenge-response

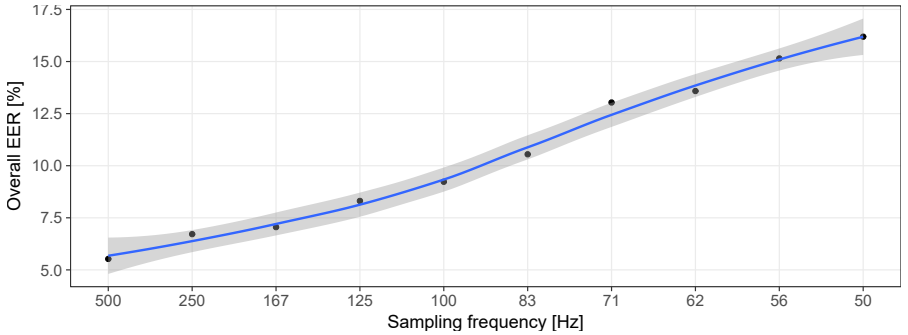


Fig. 11. The effect of the sampling frequency on the overall EER. Sampling frequencies between 50 and 250 Hz were simulated by decimation: applying a low-pass filter before subsampling the original 500 Hz measurements. As the sampling rate reduces from 500 Hz to 50 Hz, the EER increases by about 11%. However, for frequencies close to 120 Hz, which are supported by a range of affordable eye tracking devices, the error rates are still well below 10%.

pair, we report the results for a scenario in which identity verification always returns a positive answer.

Setup. In order to evaluate the probability of success of a replay attack, for each gaze-challenge c_i, we simulate a "replay" of all other gaze-responses r_j to the VerifyFreshness function of the system. We calculate the success rate of replaying r_j to c_i as the percentage of stimulus positions from c_i that would be considered successfully gazed if a user's response was r_j.

Since our dataset consists of 1021 legitimate authentication attempts, each recorded with a unique gaze-challenge, we are able to simulate more than 10^6 potential replay attempts in order to estimate the true reject rate. Furthermore, in order to estimate the true accept rates, we use the same procedure to simulate a total of 1021 legitimate authentication attempts, in which the gaze-response was indeed generated as the user was presented with the matching gaze-challenge.

Results. Figure 10 shows achieved performance of the challenge-response verification for different values of T, which we vary from 0% to 100%. As T increases, so does true reject rate (TRR), the ratio of replay attempts that are correctly rejected. On the other hand, this also causes a decrease of the true accept rate (TAR), the ratio of legitimate, fresh attempts that are correctly accepted.

The desired threshold is the one that detects all replay attempts while accepting all legitimate authentication attempts as fresh. Figure 10 shows a wide range of potential threshold values that lie between 40% and 60% and almost perfectly separate the fresh and the replayed gaze-responses. Such a broad range of thresholds that achieve strong classification is a desirable property for any classification system as it gives strong confidence in reported results.

Since we use $T = 50\%$ to evaluate impersonation attacks, we report specific numeric details for this threshold. The results of simulating more than 10^6 challenge-response pairs as replay attempts show that we achieve close to perfect true reject rates (TRR) of 99.94%. At the same time, very few legitimate attempts are incorrectly rejected: the evaluation shows a true accept rate (TAR) of 98.63%, a result of falsely rejecting only 14 out of 1021 legitimate attempts.

Overall, these experimental results show that our system robustly prevents replay attempts for a wide range of thresholds with very high success rates. Moreover, given that the system can detect repeated authentication attempts, and e.g. lock user's account after a certain number of failed attempts, we finally conclude that our system can effectively prevent replay attacks.

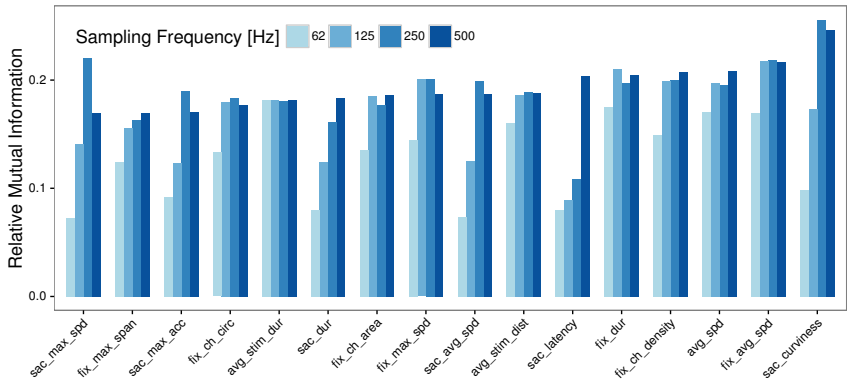


Fig. 12. The impact of sampling frequency on the Relative Mutual Information, $RMI_{ID}(F)$, of used features. Features are ordered according to their RMI at 500 Hz. Due to the velocity of saccadic eye movements, the differences in RMI are most visible in saccade-based features as the sampling frequency is reduced to 125 Hz or 62 Hz. On the other hand, the impact on fixation-based features is significantly smaller.

9 SYSTEM ANALYSIS

After evaluating the proposed system's security guarantees, we now use the dataset to analyze how would the system performance change if some of the crucial design choices had been altered. Such analysis helps strengthen future research on the same topic by providing a better understanding of its behavior, confirming the correctness of choices that were made, or showing potential directions for future improvement.

9.1 Sampling Frequency

While this manuscript focused on an office scenario, in which users authenticate over a course of a normal work day, an important factor in the overall performance of a biometric system is the availability of high-quality data. Even though the sampling frequencies of widely affordable eye tracking devices keep increasing with the proliferation of cheaper high-speed cameras, most consumer-grade eye trackers still predominately capture gaze data between 60 and 240 times per second. As discussed in Section 7, for our experimental data acquisition we use a high-end eye-tracking device that has the ability to capture eye movements at frequencies of up to 500 Hz.

In order to evaluate the feasibility of using reflexive eye movements and the proposed set of features for biometric authentication with a wider range of eye tracking devices, we now simulate a scenario in which data was acquired at lower frequencies.

Setup. We simulate data acquisition at lower frequencies by first applying a low-pass IIR filter (with a suitable limit frequency) before subsampling the data between 2 and 10 times, i.e., by discarding all but every M-th gaze measurement. This results in 9 new datasets, with sampling frequencies ranging from 250 Hz (M=2) to 50 Hz (M=10), and 9 new sets of features computed using the exact same procedure as with original data. Finally, all 10 sets of features are used to repeatedly train and test a classifier for each user, following the procedure outlined in Section 8.2.

Results. The results of the analysis of how the sampling rate influences the overall system error rates are shown in Figure 11. As the sampling rate reduces from 500 Hz to 50 Hz, the EER increases by about 11%. However, the measured difference between the error rates at 500 Hz and 125 Hz,

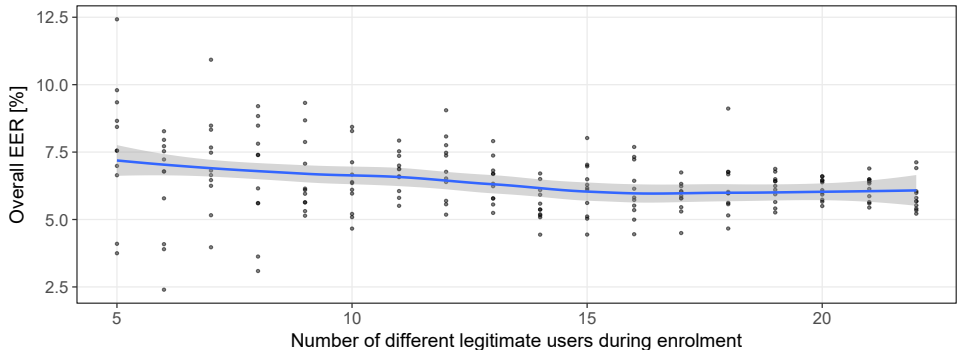


Fig. 13. The impact of the size of the negative class on the overall EER, computed by varying the number of different users' samples that the classifier is exposed to during enrollment. For each negative class size, we repeat the measurement 10 times to show the variability in performance. Increasing the negative class size decreases the variance in system performance, while at the same time slightly reducing the overall error rates. However, once the size of the negative class reaches about 10, the overall performance stabilizes at 5-7% EER.

which are supported by a range of affordable eye tracking devices, is around 2%, remaining well below 10%.

Consequently, we compute and show in Figure 12 the RMI_{ID} values for each feature as they are subsampled with 250 Hz, 125 Hz, and 62 Hz. The figure shows that the RMI_{ID} of several features, mostly those related to the velocity of eye movements during fixations or saccades, actually does increase as the sampling rate goes from 500 Hz to 250 Hz, while the informativeness of the majority of the other features remains relatively unchanged. As a result, the classification performance remains very similar despite the halving of the sampling frequency.

Overall, these results show that, while using a high-speed eye tracking device does indeed improve classification performance and the quality of the extracted features, even when low- and mid-range eye tracking devices are used, the overall authentication success rates are expected to remain high.

9.2 Size of the Negative Class During Enrollment

One of the crucial factors that impact the performance of classifiers is the variability of data seen during training. This is especially true for binary classification performed during biometric authentication, in which a specific classifier is trained for each enrolled user, with the purpose of deciding whether a new biometric sample does indeed belong to the claimed identity or not. During training, if the classifier is supplied with negative samples of limited variability (i.e. other legitimately enrolled users' gaze samples), it is likely to overfit, failing to generalize when required to classify biometric samples of new, never seen individuals. In such scenario, the classifier learns to reject specific characteristics of the seen negative samples, rather than recognizing the characteristics of positive samples and rejecting the rest.

Authentication vs Identification. When discussing classification performance in relation to the number of different classes, it is important to distinguish biometric authentication (1-1 classification) from biometric identification, as the latter requires 1-N classification to determine the identity of the biometric sample. In the identification scenario, introducing each additional class (i.e., increasing

the number of users) results in making the problem distinctly harder, the probability of successfully identifying the correct class out of (N+1) is smaller than out of N classes.

However, given that in the authentication scenario the classifier always makes a binary decision, increasing the number of different users seen during training or testing introduces significant variability only up to some level, after which the performance usually stabilizes, assuming that the samples seen during training and the samples on which the system is tested are representative of the true distribution. As a result, while it is not straightforward to generalize the identification performance of a classifier beyond the number of tested classes, this is less true for authentication performance, assuming that the classifier is exposed to a sufficient amount of variability during training and that the testing samples correspond to the actual real distribution.

In order to estimate the required variability of the negative class required to achieve stable biometric authentication performance, we now evaluate the error rates of our system, depending on the number of different users that are available during enrollment of each classifier.

Setup. We simulate the scenario in which only a random subset of legitimate users' gaze tracking measurements are available during enrollment as the negative class. For each size of the negative class, we randomly choose 10 different subsets of other legitimate users, and we repeat the training process three times for each user, computing and reporting the equal error rates using the procedure outlined in Section 8.2.

Results. The results of the evaluation are shown in Figure 13. As the number of users U increases from the initial scenario of $U = 5$ towards $U = 15$, a decrease in the overall equal error rates, shown as the blue line, is visible, resulting in about 1% stronger performance. However, when $U > 15$, the measurements show no significant difference between the average classification error rates as the number of users increases.

Given that each classifier makes a binary decision "legitimate user or attacker", the amount of additional variability does not increase significantly after the number of users seen during enrollment reaches 15, resulting in comparable classification performance. Additionally, as the size of the negative class seen during training increases, so does the variability of the measured equal error rates decrease, stabilizing at 5-7% EER for U over 19.

9.3 Dwell-time Threshold D

One of the main characteristics of the proposed authentication system based on reflexive eye movements is the interactivity of the stimulus in response to user's gaze. As discussed in Section 5, interactively moving the position of the stimulus dot allows the system to extract the maximal number of saccades in a given time, while at the same time reducing voluntary saccades that happen while the user is waiting for the stimulus to change. However, due to head and body movements as well as imperfections in the gaze tracking devices, the location of users' gaze is not always captured perfectly, which results in users sometimes not being able to successfully gaze at the stimulus position for a period of time.

The prototype that we use for experimental data acquisition was built such that the visual stimulus remains at the same position until it is either successfully gazed, or for the duration of the maximum dwell-time threshold ($D = 1000\ ms$). This ensures that the authentication process continues even if the user is unable to gaze at a particular stimulus location, or, e.g., does not pay attention to the screen - in which case the authentication will expectedly fail.

Given that the majority of reflexive eye movements are considered to have latencies below 250 ms [47], and taking into account the distribution of saccade latencies from our experiments (shown in Figure 4), it is likely that the presented system could use a lower dwell-time threshold

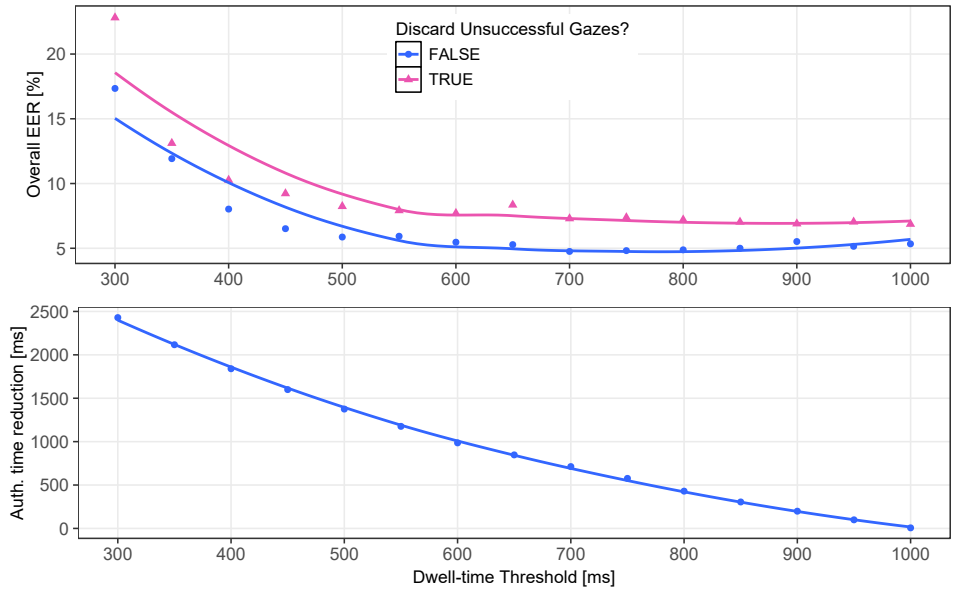


Fig. 14. The impact of different dwell-time thresholds (D) on (a) the mean equal error rates and (b) the reduction in median authentication time. In (a), the blue line approximates the performance of completely discarding those parts of user's eye movements which do not correspond in a successful gaze at a certain stimulus location, while the purple line corresponds to the case in which only the parts of the gaze after the threshold are discarded. Considering the physical limitations of eye movements, low dwell-time thresholds expectedly result in high error rates. However, as the threshold reaches 500 ms, error rates stabilize while at the same time reducing median authentication times by about 1500 ms.

(D) without significant loss in authentication performance. Since reducing D is expected to result in a decrease of the authentication times for the users, we now analyze the impact of different values of dwell-time threshold on the system performance.

Setup. Considering that our data was recorded with $D = 1000$ *ms*, we can evaluate any $D < 1000$ *ms* by simply discarding those parts of gaze tracking measurements that happen more than D milliseconds after the stimulus last changed position. For each analyzed value D, we create a separate dataset, train a classifier according to the procedure described in Section 6.3 and evaluate it according to the description in Section 8.2.

Additionally, by *"Discard Unsuccessful Gazes"* we consider the option of completely discarding all gaze samples that were measured for the whole duration while the stimulus dot was shown at a certain location if the user was not ultimately successful at gazing it in the period of D ms. We run this variant of the analysis expecting that measurements from unsuccessful gazing might be more noisy than the measurements which result in successful gazes at the stimulus dot, and that they could consequently impair classification performance, rather than improve it.

Results. The average error rates and the reductions in median authentication times for values of D ranging from 300 ms to 1000 ms are shown in Figure 14. In the upper graph, the purple color indicates the scenario in which unsuccessful gazes are completely discarded when computing the

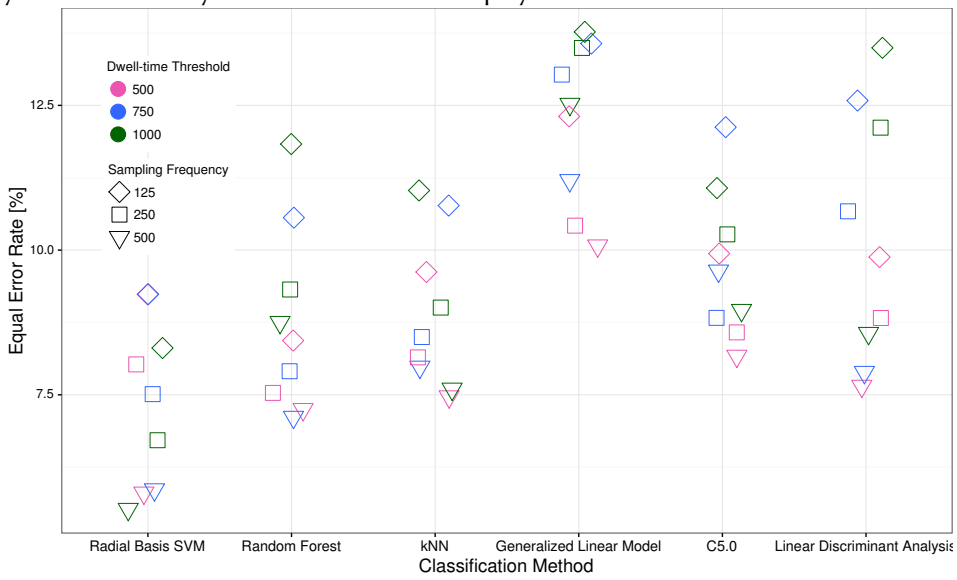


Fig. 15. Comparison of equal error rates for 9 different configurations of several classification methods: Radial basis SVMs, Random Forests, k-Nearest Neighbours, Generalized Linear Model, C5.0, and Linear Discriminant Analysis. The used sampling frequencies and dwell-time thresholds are depicted by the shape and color, respectively. Radial Basis Support Vector Machine achieves the lowest error rate in each specific configuration.

features for classification, while the blue color shows the results of discarding only the samples after the threshold D.

As expected, as D decreases from 1000 ms towards 300 ms, the overall equal error rates increase as well, since some of the useful distinctive information will be discarded. Contrary to our expectation, fully discarding all measurements that did not result in a successful gaze actually increases the error rates, indicating that such gaze data still carries valuable information.

While the error rates for thresholds below 450 ms quickly grow above 15%, it is important to note that the error rates for thresholds above 500 ms are almost identical. This confirms the hypothesis that in most cases, users reflexively gaze a specific stimulus position in less than 500 ms, and that the subsequent behavior carries significantly less useful information.

As an important consequence of this analysis, we see that reducing the dwell-time threshold to 500 ms results in the reduction of median authentication times by as much as 1500 ms, which in turn increases the usability of the proposed system, while not impacting its overall classification performance.

9.4 Choice of the Classifier

We continue the analysis of system parameters by evaluating different classification methods and models that we could use for identity verification. Besides the radial-basis SVM (`radialSVM`), which showed the best overall performance on all comparison tests, we also test a set of five other commonly used classifiers that are implemented in the Caret library: Random Forest (`rf`), k-Nearest Neighbours (`kNN`), Generalized Linear Model (`glm`), C5.0 (`C5.0`), and Linear Discriminant Analysis (`lda`).

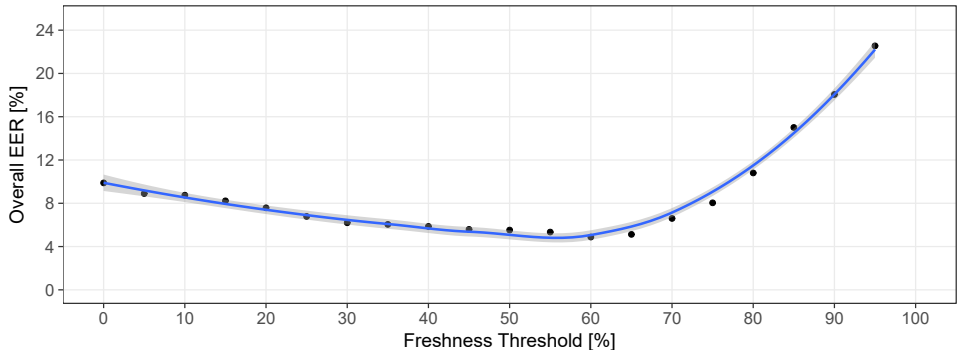


Fig. 16. Comparison of equal error rates for different freshness verification thresholds T. The minimal EER is achieved when T is close to 50%. T equal to 0 provides the error rates computed only based on eye movement data classification. While the error rates do increase by about 4%, they still remain at 10.47%.

Setup. We compare the performance of different classifiers by running a battery of tests, in which we vary the sampling rate (125 Hz, 250 Hz, 500 Hz), as well as the maximum dwell-time threshold D (500 ms, 750 ms, 1000 ms). Besides specifying the exact classification method to the Caret library, all other parameters are kept the same across different classification methods, consistent with the other computed EER rates: we run a 5-fold, 3-times repeated cross-validation, and repeatedly compute and average the results for each user three times.

Results. The results of each of the 9 evaluations that were run for each of the 6 classifiers are shown in Figure 15, with Equal Error Rate being the measure of classification performance. The color of each of the point indicates the cut-off threshold D, while the shape of each point indicates the sampling rate. The radial-basis SVM, which we use in all other analysis, clearly achieves the lower error rates for all combinations, with the Random Forrest classifier trailing a few percentage points behind. It is interesting to note that the results shown in previous subsections, which were all computed using the SVM classifier, indicated that the higher sampling rate and dwell-time threshold D should result in lower EER-s. However, the results in Figure 15 show that this is not always the case with other classifiers, especially in the case of low sampling rate, where the situation is reversed, with lower dwell-time thresholds resulting in lower overall equal error rates.

9.5 Impact of Freshness Verification Threshold

As described in Section 5.3, if the received gaze measurements do not match at least T percent of the randomly chosen stimuli positions, the respective sample is rejected as a potential replay attack. Such rejections, however, do not happen only as a result of replaying a sample that is not fresh. Firstly, they can be a result of changes in the eye tracking geometry as legitimate users carry out their daily work, counting as a false reject that negatively impacts the overall EER of the system. Secondly, since the system always uses the calibration profile of the victim whose identity is being impersonated, such rejection of a fresh measurement can also happen during an impersonation attempt. Such true rejections happen if the calibration profile does not fit the attacker sufficiently to precisely gaze at the necessary percentage and decrease the overall EER of the system.

Even though the attacker is unable to control the threshold T or disable the freshness verification component of the system, analyzing the impact of freshness verification on the overall error rates

provides an intuition on the stability of the combined system and the guarantees that could be achieved in eye tracking systems that do not require user calibration (e.g. eye-tracking glasses).

Setup. We vary the required T from 0% to 100% and compute the overall EER values for the whole dataset. Setting the threshold to 0% effectively computes the overall EER of the system in the case where freshness verification is not taken into account during authentication. On the other hand, setting the threshold to 100% is equivalent to requiring that users precisely gaze at the stimuli for all positions on the screen, resulting in high error rates due to the majority of legitimate attempts also being rejected.

Results. The results are shown in Figure 16. The overall EER is lowest when T is between 45% and 65%, which is in accordance with results of replay attacks evaluation (Section 5.3)[2].

As T increases, this causes more legitimate authentication attempts to be rejected on the grounds of potentially being a replay attack, and causes a sharp increase in the overall EER against impersonation attacks.

On the other hand, as T decreases, the effect of freshness verification module diminishes, ultimately providing the estimate of the system performance in the case where freshness verification would be completely ignored ($T = 0$). While the error rates increase by about 4% as T reduces from 50% to 0%, they ultimately stop increasing at 10.47%, showing that the performance of the system against impersonation attacks would not significantly decrease if all or most attack attempts passed freshness verification.

10 DISCUSSION

Advanced Attacks. A more sophisticated attacker could build a model of a legitimate user's eye movements to successfully respond to a given challenge. However, we argue that performing such attacks is not straightforward and requires a higher level of complexity than simply replaying a biometric sample.

Firstly, the adversary is likely to be solving a harder problem than the authentication system; while the system needs to build a discriminative model that allows making a binary decision about user's identity, the adversary needs to actually generate eye movements which correspond to the legitimate user. An indication of the difficulty of artificially creating eye-movements can be found in work by Komogortsev et al. [34], which evaluated the complexity of a significantly simpler problem: artificially generating 1-dimensional eye movements. The paper showed that those movements could be distinguished from natural recordings with high accuracy; creating realistic 2D eye-movements that correspond to a specific user is likely to be significantly harder.

Secondly, by using a challenge-response type of protocol, we ensure that the potential generative model of legitimate user's eye movements must be able to output results interactively and in real-time since the stimulus is not known in advance. This requires an additional level of sophistication that is not needed for replay attacks since the adversary needs to not only control the gaze tracking channel, but to also observe and analyze the visual channel.

Applications, Limitations, and Future Work. We now discuss several challenges that remain before the proposed concepts can be applied in a wider range of practical applications: namely the practicality of achieved error rates, the temporal stability of the eye movement biometrics, and the use of high-end devices that require calibration.

[2]We note that replay attacks are evaluated using only legitimate authentication attempts since we assume that an adversary would not try replaying an impersonation attempt.

Firstly, the equal error rates between 6% and 10% are not yet sufficient to be independently deployed in real-world systems. However, it is important to note that a real-world authentication system could combine the evaluated dynamic features with some of the static features often available as part of the standard eye-tracking procedures, such as pupil sizes, distances between user's eyes, or even iris images. Taking this research direction further, a potential future application of ideas proposed in this manuscript is in increasing the security of various face recognition systems. Many such systems already implement measures to prevent sophisticated spoofing attacks [44], e.g., by requiring users to smile, move their head, or to gaze at the direction of the camera [48]. Consequently, combining the stability and low error rates of face recognition with the dynamic characteristics and the freshness verification of reflexive eye movements could retain the usability of face recognition while ensuring that an adversary cannot spoof such a system using the currently available methods against face recognition.

An important requirement for potential long-term biometric use of this biometric is to evaluate and improve the temporal stability of the proposed eye movement features over extended periods of time. After having shown in this work that the proposed visual stimuli does indeed quickly extract features that allow discrimination between users while at the same time preventing replay attacks, we plan to next focus on designing an extensive larger set of potential features, capturing larger datasets with different eye tracking devices, and evaluating their long-term stability over multiple sessions [24], following a test-retest approach based on Intra-Class-Correlation as the feature selection criteria [16].

Finally, as a proof-of-concept evaluation of using reflexive eye movements for authentication, we measured gaze samples with a static, high-end eye tracker that relies on individual calibration profiles of each user. While calibration of video-based eye tracking devices can be an unpredictable and time-consuming process, this is still an active research area [19] and we expect this step to significantly improve in the future. Additionally, an important potential application area for eye tracking is in extending the capabilities of Virtual and Augmented Reality headsets by integrating technology similar to existing eye tracking glasses. Given that most such glasses do not require calibration, and that AR/VR headsets conveniently provide a display on which the proposed reflexive stimulus could be shown to the user, we look forward to applying the proposed system to such headsets in our future work.

11 CONCLUSION

Building upon the core idea of using reflexive human behavior for authentication, in this paper we designed an interactive visual stimulus for rapidly eliciting standardized reflexive eye movements, and showed how such stimulus can be used to construct a fast challenge-response biometric system. Based on a series of user experiments, we showed that our stimulus indeed elicits predominately reflexive saccades, which are automatic responses that only pose low cognitive load on the user. As a result of using reflexive behavior that is fast and stable, we show that our authentication system achieves fast authentication times (median of 5 seconds) and low error rates (6.3% EER for impersonation attacks).

Most importantly, however, our proposed authentication method shows resilience against replay attacks, a property difficult to achieve with most biometrics. Our evaluation shows that the system is able to detect the replay of recorded eye traces with a very high probability of 99.94%, thus preventing one of the most applicable attacks on biometric systems.

Considering the recent proliferation of reliable and affordable eye tracking devices, and the early applications of eye tracking in future VR and AR headsets, we believe that achieving fast and reliable gaze-based authentication is of broad interest and we consider our work to be an important step in this direction.

Acknowledgements. Ivo Sluganovic is supported by the UK EPSRC doctoral studentship, Scatcherd European Scholarship, and the Frankopan Fund. Authors wish to thank Armasuisse for support with the gaze tracking equipment.

REFERENCES

[1] William W Abbott and Aldo A Faisal. 2012. Ultra-low-cost 3D gaze estimation: an intuitive high information throughput compliment to direct brain-machine interfaces. *Journal of Neural Engineering* 9, 4 (2012).

[2] Richard A Abrams, David E Meyer, and Sylvan Kornblum. 1989. Speed and accuracy of saccadic eye movements: characteristics of impulse variability in the oculomotor system. *Journal of experimental psychology. Human perception and performance* 15, 3 (1989).

[3] Terry Bahill, Michael R Clark, and Lawrence Stark. 1975. The main sequence, a tool for studying human eye movements. *Mathematical Biosciences* 24, 3-4 (1975), 191–204.

[4] Terry Bahill and Tom Laritz. 1984. Why Can't Batters Keep Their Eyes on the Ball? *American Scientist* May - June (1984).

[5] Arman Boehm, Dongqu Chen, Mario Frank, Ling Huang, Cynthia Kuo, Tihomir Lolic, Ivan Martinovic, and Dawn Song. 2013. SAFE: Secure authentication with Face and Eyes. In *Privacy and Security in Mobile Systems (PRISMS), 2013 International Conference on.*

[6] Andreas Bulling, Florian Alt, and Albrecht Schmidt. 2012. Increasing the security of gaze-based cued-recall graphical passwords using saliency masks. In *CHI.*

[7] Virginio Cantoni, Chiara Galdi, Michele Nappi, Marco Porta, and Daniel Riccio. 2015. GANT: Gaze analysis technique for human identification. *Pattern Recognition* 48, 4 (2015).

[8] Monica S Castelhano and John M. Henderson. 2008. Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology* 62, 1 (2008), 1–14.

[9] Jennie E S Choi, Pavan a Vaswani, and Reza Shadmehr. 2014. Vigor of movements and the cost of time in decision making. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 34, 4 (2014).

[10] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning* 20 (1995), 273–297.

[11] Alexander De Luca, Martin Denzel, and Heinrich Hussmann. 2009. Look into My Eyes!: Can You Guess My Password?. In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS '09).* ACM, New York, NY, USA, Article 7.

[12] Francesco Di Russo, Sabrina Pitzalis, and Donatella Spinelli. 2003. Fixation stability and saccadic latency in elite shooters. *Vision Research* 43, 17 (2003).

[13] Simon Eberz, Kasper B Rasmussen, Vincent Lenders, and Ivan Martinovic. 2015. Preventing Lunchtime Attacks: Fighting Insider Threats With Eye Movement Biometrics. In *Proceedings of the 2015 Networked and Distributed System Security Symposium.*

[14] Simon Eberz, Kasper B Rasmussen, Vincent Lenders, and Ivan Martinovic. 2016. Looks Like Eve: Exposing Insider Threats Using Eye Movement Biometrics. *ACM Transactions on Privacy and Security* 19, 1, Article 1 (June 2016), 31 pages.

[15] Mario Frank, Ralf Biedert, Eugene Ma, Ivan Martinovic, and Dawn Song. 2013. Touchalytics: On the Applicability of Touchscreen Input As a Behavioral Biometric for Continuous Authentication. *Trans. Info. For. Sec.* 8, 1 (Jan 2013), 136–148.

[16] Lee Friedman, Mark S Nixon, and Oleg V Komogortsev. 2017. Method to assess the temporal persistence of potential biometric features: Application to oculomotor, gait, face and brain structure databases. *PloS one* 12, 6 (2017), e0178501.

[17] Chiara Galdi, Michele Nappi, Daniel Riccio, Virginio Cantoni, and Marco Porta. 2013. A new gaze analysis based soft-biometric. *Lecture Notes in Computer Science* 7914 LNCS (2013).

[18] Lawrence R Gottlob, Mark T Fillmore, and Ben D Abroms. 2007. Age-group differences in saccadic interference. *The journals of gerontology. Series B, Psychological sciences and social sciences* 62, 2 (2007), 85–89.

[19] Katarzyna Harezlak, Pawel Kasprowski, and Mateusz Stasch. 2014. Towards accurate eye tracker calibration–methods and procedures. *Procedia Computer Science* 35 (2014), 1073–1081.

[20] Corey D Holland and Oleg V Komogortsev. 2011. Biometric identification via eye movement scanpaths in reading. *2011 International Joint Conference on Biometrics, IJCB 2011* (2011).

[21] Corey D Holland and Oleg V Komogortsev. 2013. Complex eye movement pattern biometrics: Analyzing fixations and saccades. In *Biometrics (ICB), 2013 International Conference on.*

[22] Kenneth Holmqvist, Marcus Nystrom, Richard Andersson, Richard Dewhurst, Jarodzka Halszka, and Joost van de Weijer. 2011. *Eye Tracking : A Comprehensive Guide to Methods and Measures.* Oxford University Press. 560 pages.

[23] Pawel Kasprowski. 2004. Human Identification Using Eye Movements. *Institute of Computer Science* (2004).

[24] Paweł Kasprowski. 2013. The impact of temporal proximity between samples on eye movement biometric identification. In *Computer Information Systems and Industrial Management.* Springer, 77–87.

[25] Pawel Kasprowski. 2014. The Second Eye Movements Verification and Identification Competition. In *IEEE & IAPR International Joint Conference on Biometrics*.
[26] Pawel Kasprowski and Katarzyna Harezlak. 2018. Fusion of eye movement and mouse dynamics for reliable behavioral biometrics. *Pattern Analysis and Applications* 21, 1 (2018), 91–103.
[27] Pawel Kasprowski, Oleg V Komogortsev, and Alex Karpov. 2012. First eye movement verification and identification competition. *2012 IEEE 5th International Conference on Biometrics: Theory, Applications and Systems, BTAS 2012* (2012).
[28] Pawel Kasprowski and Jozef Ober. 2003. Eye Movements in Biometrics. *Biometrics* 3087 / 200 (2003).
[29] Katharine Byrne. 2015. MSI & Tobii join forces to create eye-tracking gaming laptop. (2015). http://www.expertreviews.co.uk/laptops/1403340/msi-tobii-join-forces-to-create-eye-tracking-gaming-laptop
[30] Ami Klin, Warren Jones, Robert Schultz, Fred Volkmar, and Donald Cohen. 2002. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of general psychiatry* 59 (2002), 809–816.
[31] Tomasz Kocejko and Jerzy Wtorek. 2012. *Information Technologies in Biomedicine: Third International Conference, ITIB 2012, Gliwice, Poland, June 11-13, 2012. Proceedings*. Springer Berlin Heidelberg, Berlin, Heidelberg, Chapter Gaze Pattern Lock for Elders and Disabled, 589–602.
[32] Olga V Kolesnikova, Lav V Tereshchenko, Alexander V Latanov, and Viktor V. Shulgovskii. 2010. Effects of Visual Environment Complexity on Saccade Performance in Humans with Different Functional Asymmetry Profiles. *Neuroscience and Behavioral Physiology* 40, 8 (2010), 869–876.
[33] Oleg V Komogortsev, Ukwatta K S Jayarathna, Cecilia R Aragon, and Mahmoud Mechehoul. 2010. Biometric Identification via an Oculomotor Plant Mathematical Model. *Eye Tracking Research & Applications Symposium* (2010).
[34] Oleg V Komogortsev, Alexey Karpov, and Corey D Holland. 2015. Attack of Mechanical Replicas : Liveness Detection With Eye Movements. *IEEE TIFS* 10, 4 (2015).
[35] Manu Kumar, Tal Garfinkel, Dan Boneh, and Terry Winograd. 2007. Reducing Shoulder-surfing by Using Gaze-based Password Entry. In *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS '07)*. ACM, New York, NY, USA.
[36] Michael F Land. 2011. Oculomotor behaviour in vertebrates and invertebrates. *The Oxford handbook of eye movements* 1 (2011).
[37] Emiliano Miluzzo, Tianyu Wang, and Andrew T Campbell. 2010. EyePhone: Activating Mobile Phones with Your Eyes. *Workshop on Networking, Systems, and Applications on Mobile Handhelds (MobiHeld)* (2010).
[38] Marcus Nystrom and Kenneth Holmqvist. 2010. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods* 42, 1 (2010).
[39] Tony Poitschke, Florian Laquai, Stilyan Stamboliev, and Gerhard Rigoll. Gaze-based interaction on multiple displays in an automotive environment. In *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*.
[40] Ioannis Rigas, George Economou, and Spiros Fotopoulos. 2012. Biometric identification based on the eye movements and graph matching techniques. *Pattern Recognition Letters* 33, 6 (2012).
[41] Ioannis Rigas and Oleg V Komogortsev. 2014. Biometric Recognition via Probabilistic Spatial Projection of Eye Movement Trajectories in Dynamic Visual Environments. *IEEE TIFS* 9, 10 (2014).
[42] Usman Saeed. 2015. Eye movements during scene understanding for biometric identification. *Pattern Recognition Letters* 59 (2015). Issue July.
[43] SensoMotoric Instruments GmbH. 2011. *SMI RED500 Technical Specification*. Technical Report. SensoMotoric Instruments GmbH, Teltow, Germany. 1 pages.
[44] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K. Reiter. 2016. Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16)*. ACM, New York, NY, USA, 1528–1540.
[45] Ivo Sluganovic, Marc Roeschlin, Kasper B Rasmussen, and Ivan Martinovic. 2016. Using Reflexive Eye Movements for Fast Challenge-Response Authentication. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16)*. ACM, New York, NY, USA, 1056–1067.
[46] Petroc Sumner. 2011. Determinants of saccade latency. In *Oxford handbook of eye movements*. Vol. 22. 411–424.
[47] Robin Walker, David G Walker, Masud Husain, and Christopher Kennard. 2000. Control of voluntary and reflexive saccades. *Experimental Brain Research* 130, 4 (Feb. 2000), 540–544.
[48] Yi Xu, True Price, Jan-Michael Frahm, and Fabian Monrose. 2016. Virtual U: Defeating Face Liveness Detection by Building Virtual Models from Your Public Photos. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, Austin, TX, 497–512.
[49] Yun Zhang, Zheru Chi, and Dagan Feng. 2011. An Analysis of Eye Movement Based Authentication Systems. *International Conference on Mechanical Engineering and Technology (ICMET-London 2011)* (2011).
[50] Youming Zhang, Jorma Laurikkala, and Martti Juhola. 2014. Biometric Verification of a Subject with Eye Movements. *Int. J. Biometrics* 6, 1 (March 2014), 75–94.

Received September 2017; revised May 2018; accepted September 2018