Privacy-Aware Eye Tracking Using Differential Privacy

Julian Steil
Max Planck Institute for Informatics
Saarland Informatics Campus
jsteil@mpi-inf.mpg.de

Inken Hagestedt
CISPA Helmholtz Center for Information Security
Saarland Informatics Campus
inken.hagestedt@uni-saarland.de

Michael Xuelin Huang
Max Planck Institute for Informatics
Saarland Informatics Campus
mhuang@mpi-inf.mpg.de

Andreas Bulling
University of Stuttgart
Institute for Visualisation and Interactive Systems
andreas.bulling@vis.uni-stuttgart.de

ABSTRACT

With eye tracking being increasingly integrated into virtual and augmented reality (VR/AR) head-mounted displays, preserving users' privacy is an ever more important, yet under-explored, topic in the eye tracking community. We report a large-scale online survey (N=124) on privacy aspects of eye tracking that provides the first comprehensive account of with whom, for which services, and to what extent users are willing to share their gaze data. Using these insights, we design a privacy-aware VR interface that uses differential privacy, which we evaluate on a new 20-participant dataset for two privacy sensitive tasks: We show that our method can prevent user re-identification and protect gender information while maintaining high performance for gaze-based document type classification. Our results highlight the privacy challenges particular to gaze data and demonstrate that differential privacy is a potential means to address them. Thus, this paper lays important foundations for future research on privacy-aware gaze interfaces.

CCS CONCEPTS

• **Security and privacy → Human and societal aspects of security and privacy**; • **Human-centered computing → Human computer interaction (HCI)**;

KEYWORDS

Online Survey; Data Sharing; Privacy Protection; Gaze Behaviour; Eye Movements; User Modeling

ACM Reference Format:
Julian Steil, Inken Hagestedt, Michael Xuelin Huang, and Andreas Bulling. 2019. Privacy-Aware Eye Tracking Using Differential Privacy. In *2019 Symposium on Eye Tracking Research and Applications (ETRA '19), June 25–28, 2019, Denver, CO, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3314111.3319915

1 INTRODUCTION

With eye tracking becoming pervasive [Bulling and Gellersen 2010; Tonsen et al. 2017], preserving users' privacy has emerged as an important topic in the eye tracking, eye movement analysis, and gaze interaction research communities. Privacy is particularly important in this context given the rich information content available in human eye movements [Bulling et al. 2011a], on one hand, and the rapidly increasing capabilities of interactive systems to sense, analyse, and exploit this information in everyday life [Hansen et al. 2003; Stellmach and Dachselt 2012; Vertegaal et al. 2003] on the other. The eyes are more privacy-sensitive than other input modalities: They are typically not consciously controlled; they can reveal unique private information, such as personal preferences, goals, or intentions. Moreover, eye movements are difficult to remember, let alone reconstruct in detail, in retrospect, and hence do not easily allow users to "learn from their mistakes", i.e. to reflect on their past and change their future privacy-related behaviour.

These unique properties and rapid technological advances call for new research on next-generation eye tracking systems that are *privacy-aware*, i.e. that preserve users' privacy in all interactions they perform with other humans or computing systems in everyday life. However, *privacy-aware eye tracking* remains under-investigated as of yet [Liebling and Preibusch 2014].

The lack of research on privacy-aware eye tracking results in two major limitations: First, there is a lack of even basic understanding

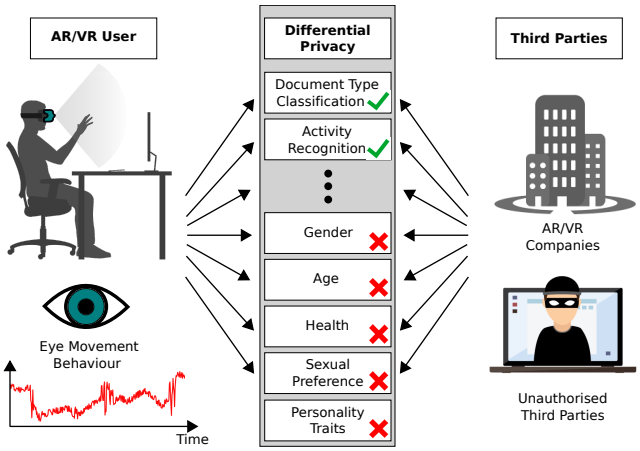

Figure 1: Using differential privacy prevents third parties, like companies or hackers, from deriving private attributes from a user's eye movement behaviour while maintaining the data utility for non-private information.

ETRA '19, June 25–28, 2019, Denver, CO, USA

ACM ISBN 978-1-4503-6709-7/19/06…$15.00
https://doi.org/10.1145/3314111.3319915

of users' privacy concerns with eye tracking in general and eye movement analysis in particular. Second, there is a lack of eye tracking methods to preserve users' privacy, corresponding systems, and user interfaces that implement (and hence permit the evaluation of) these methods with end users. Our work aims to address both limitations and, as such, make the first crucial step towards a new generation of eye tracking systems that respect and actively protect private information that can be inferred from the eyes.

Our work first contributes a large-scale online survey on privacy aspects of eye tracking and eye movement analysis. The survey provides the first comprehensive account of with whom, for which services, and to what extent users are willing to share their eye movement data. The survey data is available at https://www.mpi-inf.mpg.de/MPIIDPEye/. Informed by the survey, *we further contribute the first method to protect users' privacy in eye tracking based on differential privacy (DP)*, a well-studied framework in the privacy research community. In a nutshell, DP adds noise to the data so as to minimise chances to infer privacy-sensitive information or to (re-)identify a user while, at the same time, still allow use of the data for desired applications (the so-called utility task), such as activity recognition or document type classification (see Figure 1). We illustrate the use of differential privacy for a sample virtual reality (VR) gaze interface. We opted for a VR interface given that eye tracking will be readily integrated into upcoming VR head-mounted displays, and hence, given the significant and imminent threat potential [Adams et al. 2018]: Eye movement data may soon be collected at scale on these devices, recorded in the background without the user noticing, or even transferred to hardware manufacturers.

2 RELATED WORK

We discuss previous works on 1) information available in eye movements, 2) eye movements as a biometric, and 3) differential privacy.

2.1 Information Available in Eye Movements

A large body of work across different research fields has demonstrated the rich information content available in human eye movements. Pupil size is related to a person's interest in a scene [Hess and Polt 1960] and can be used to measure cognitive load [Matthews et al. 1991]. Other works have shown that eye movements are closely linked to mental disorders, such as Alzheimer's [Hutton et al. 1984], Parkinson's [Kuechenmeister et al. 1977], or schizophrenia [Holzman et al. 1974]. More recent work in HCI has demonstrated the use of eye movement analysis for human activity recognition [Bulling et al. 2013; Steil and Bulling 2015] as well as to infer a user's cognitive state [Bulling and Zander 2014; Faber et al. 2017] or personality traits [Hoppe et al. 2018]. More closely related to our work, several researchers have shown that gender and age can be inferred from eye movements, e.g. by analysing the spatial distribution of gaze on images like faces [Cantoni et al. 2015; Sammaknejad et al. 2017].

All of these works underline the significant potential of eye movement analysis for a range of future applications, some of which may soon become a reality, for example, with the advent of eye tracking-equipped virtual and augmented reality head-mounted displays. Despite the benefits of these future applications, the wide availability of eye tracking will also pose significant privacy risks that remain under-explored in the eye tracking community.

2.2 Eye Movements as a Biometric

Eye movement biometrics has emerged as a promising approach to user authentication [Kasprowski and Ober 2003]. While first works required a point stimulus that users were instructed to follow with their eyes [Kasprowski 2004; Kasprowski and Ober 2005], later ones explored static points [Bednarik et al. 2005] or images [Maeder and Fookes 2003]. Kinnunen et al. presented the first method for "task-independent" person authentication using eye movements [Kinnunen et al. 2010]. Komogortsev et al. proposed the first attempt to model eye movements for authentication using an Oculomotor Plant Mathematical Model [Komogortsev and Holland 2013; Komogortsev et al. 2010]. Eberz et al. presented a biometric based on eye movement patterns. They used 20 features that allowed them to reliably distinguish and authenticate users across a variety of real-world tasks, including reading, writing, web browsing, and watching videos on a desktop screen [Eberz et al. 2016]. Zhang et al. used eye movements to continuously authenticate the wearer of a VR headset by showing different visual stimuli [Zhang et al. 2018].

While an ever-growing body of research explores eye movements as a promising modality for privacy applications and user authentication, we are the first to practically explore eye movements recorded using eye tracking as a potential threat to users' privacy.

2.3 Differential Privacy

Differential privacy has been studied in privacy research for more than a decade in terms of its theoretical foundations and its practical applications to different data types, such as location [Pyrgelis et al. 2017], biomedical data [Saleheen et al. 2016], or continuous time series data [Fan and Xiong 2012]. We refer the reader to [Zhu et al. 2017] for a survey. A key challenge in differential privacy is to find the right trade-off between privacy and utility, that is, the right amount of random noise to "hide" an individual without hampering data utility. Fredrikson et al. demonstrated how important it is to balance privacy and utility [Fredrikson et al. 2014]. They observed that either privacy was not preserved or that utility suffered, leading to increased health risks for the patients from unsuitable drug dosage. A good privacy-utility trade-off is possible if privacy mechanisms are tailored towards a specific use case [Fan and Xiong 2012; Pyrgelis et al. 2017]. While differential privacy has a long history in privacy research, to the best of our knowledge, we are the first to apply this framework to eye tracking data.

3 PRIVACY CONCERNS IN EYE TRACKING

We conducted a large-scale online survey to shed light on users' privacy concerns related to eye tracking technology and the information that can be inferred from eye movement data. We advertised our survey on social platforms (Facebook, WeChat) and local mailing lists for study announcements. The survey opened with general questions about eye tracking and VR technologies; continued with questions about future use and applications, data sharing and privacy (especially regarding with whom users are willing to share their data); and concluded with questions about the participants' willingness to share different eye movement representations. Participants answered each question on a 7-point Likert scale (1: Strongly disagree to 7: Strongly agree). To simplify the analysis, we merged scores 1 to 3 to "Disagree" and 5 to 7 to "Agree".

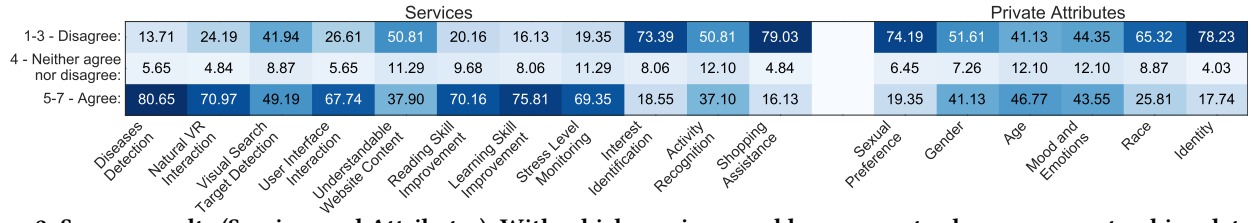

Figure 2: Survey results (Services and Attributes): With which services would you agree to share your eye tracking data (Services)?; Would you agree to private attributes being inferred by these services (Private Attributes)?

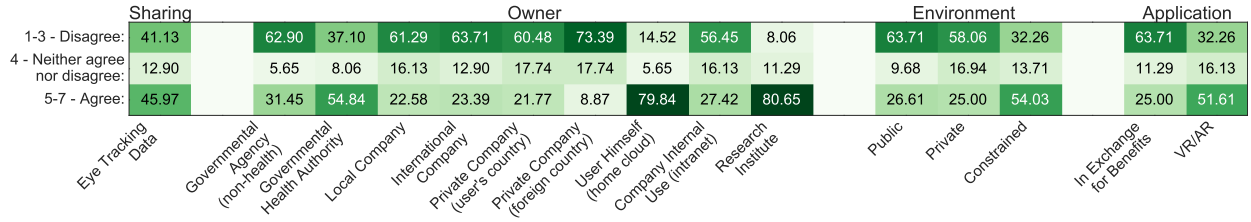

Figure 3: Survey results (Whom and Where): Would you agree to share your eye tracking data in general (Sharing); with whom (Owner); where (Environment); in exchange for benefits or for VR/AR usage (Application)?

The survey took about 20 minutes to complete, was set up as a Google Form, and was split into the parts described above. Our design ensured that participants without pre-knowledge of eye tracking and VR technology could participate as well: We provided a slide show containing information about eye tracking in general, and in VR devices specifically, and introduced the different forms of data representation, showing example images or explanatory texts.

In our survey, 124 people (81 male, 39 female, 4 did not tick the gender box) participated, aged 21 to 66 (mean = 28.07, std = 5.89). The participants were from all over the world, coming from 29 different countries (Germany: 39%, India: 12%, Pakistan: 6%, Italy: 6%, China: 5%, USA: 3%). Sixty-seven percent of them had a graduate university degree (master's or PhD), and 22% had an undergraduate university degree (bachelor's). Fifty-one percent were students of a variety of subjects (law, language science, computer science, psychology, etc.); 34% were scientists and researchers, IT professionals (7%), or had business administration jobs (2%). Since the topic of the survey was in the title of posts and emails, most likely people inherently interested in the topic participated. The majority were young, educated people with a technical background the exact group of people most likely to experience AR/VR technology (73%) in contrast to, for example, older generations.

Given the breadth of results, we highlight key insights most relevant for the current paper. We found nearly all answers for the provided questions to be significantly different from an equal distribution tested with Pearson's chi-squared test ($p < 0.001$, dof = 6). Additionally, we calculated the skewness and observed that the majority of questions show a significant difference to the corresponding normal distribution ($p < 0.1$). Detailed numbers, plots, significance and skewness test results can be found in the supplementary material (see https://www.mpi-inf.mpg.de/MPIIDPEye/).

Services and Attributes: In the first part of our survey, we asked participants for which services they would share their eye tracking data and presented both currently available and potential future services as answer options. As we can see from Figure 2, more than 80% of all participants agreed to share their eye tracking data for (early) detection of diseases like Alzheimer's or Parkinson's. Likewise, the majority agreed to share their data for hands-free VR and user interface interaction. Similar results can be observed for learning and reading skill detection as well as for stress level monitoring. However, for improved gaze target recognition, website content, and activity recognition, we observe two peaks. A clear majority is unwilling to share data with shopping assistance and interest detection services.

Our next set of questions indicated the fact that services could be able to infer private attributes from their data, and we asked whether participants would still want to share their eye tracking data. We clearly observed that if the attributes of sexual preference, gender, race, and identity can be inferred, a majority do not want to share their data. It was only for age and emotion detection that we identified two different interest groups that either agree with or object to sharing their data.

Whom and Where: In the second part, of our survey we asked participants whether they would share eye tracking data in general, and with whom. Moreover, we were interested in whether the environment has an influence on their sharing behaviour (see Figure 3). Finally, we wanted to know whether the sharing behaviour is different if participants get benefits (not specified) in exchange for their data or if the data is collected during VR/AR usage in general.

The answers as to whether participants would share their eye tracking data in general do not show a clear tendency; the participants' opinions are split in two groups ($\chi^2(\text{dof} = 6) = 32.25$, $p = 1.46 \times 10^{-6}$). Next, we asked more specifically whether participants would share their data if it were later owned and operated by one of the given "owner" options in Figure 3. According to their answers, participants would only share their data if the co-owner is a governmental health-agency; they do not trust local and international companies, or company internal use. However, participants would also share their data for research purposes, which is not surprising given that 67% of participants have a graduate university degree and trust in research institutes. Participants would not agree to share their data in public, nor in private environments, but they would agree to constrained environments. Furthermore, the participants object to sharing their data for any kind of benefit, but would agree when their eye tracking data was collected in VR/AR ($\chi^2(\text{dof} = 6) = 26.72$, $p = 0.00016$).

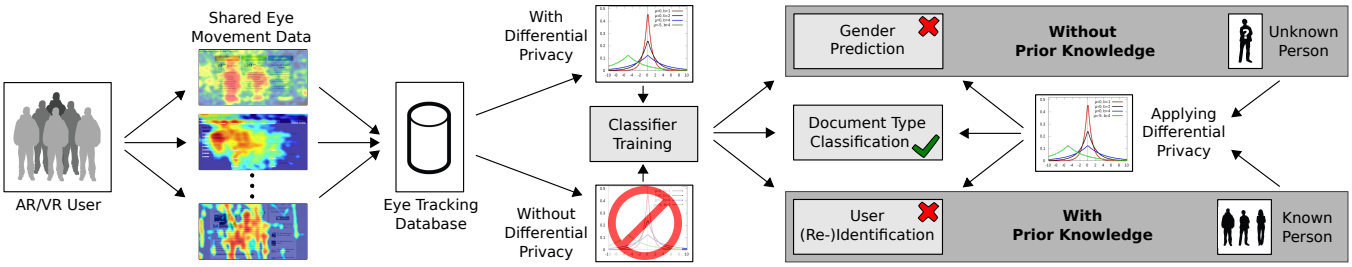

Figure 4: (Left) Our method assumes that AR/VR users share their eye tracking data and privacy-sensitive information with a third party, which is able to train classifiers with or without differentially private data to infer private attributes of an unknown (without prior knowledge) or a known (with prior knowledge) person; (Right) Applying differential privacy to test data prevents private information inference (gender, user (re-)identification) but maintains data utility (document type classification).

Data Representation: In the final part of the survey, we asked participants in what form they would agree to share their data. We discriminate 12 different representations, ranging from raw eye tracking, to heatmaps, to aggregated features (see Figure 3 in the supplementary material). Additionally, we were interested in whether their sharing behaviour changes if the data is first anonymised. Information which provides gaze information, like fixations, or scan path information on a surface would mostly not be shared. Participants largely agree to share their eye tracking data as statistical features, and especially aggregated features. This is why we focus in our study on the aggregated feature representation to apply differential privacy. Our survey shows a clear increase in participants willing to share their data in anonymised form.

4 PRIVACY-PRESERVING EYE TRACKING

The findings from our survey underline the urgent need to develop *privacy-aware eye tracking systems* – systems that provide a formal guarantee to protect the privacy of their users. Additionally, it is important not to forget that eye movement data typically also serves a desired task – a so-called *utility*. For example, eye movement data may be used in a reading assistant to detect the documents a user is reading [Kunze et al. 2013b] or to automatically estimate how many words a user reads per day [Kunze et al. 2013a, 2015]. Therefore, it is important to ensure that any privacy-preserving method does not render the utility dysfunctional, i.e. that the performance on the utility task will not drop too far. The key challenge can thus be described as *ensuring privacy without impeding utility*.

We assume in the following that multiple users share their eye tracking data in the form of aggregated features. The resulting eye tracking database is visualised in the left part of Figure 4. This database can be downloaded both for legitimate use cases as well as for infringing on users' privacy, for example, to train classifiers for various tasks. Therefore, our proposed privacy mechanism is applied prior to the release by a trusted curator.

4.1 Threat Models

We have identified two attack vectors on users' privacy in the context of eye tracking that we formalise in two threat models. They differ in their assumption about the attackers' prior knowledge about their target (see the right part of Figure 4).

Without Prior Knowledge. In the first threat model, we assume that an attacker has no prior knowledge about the target and wants to infer a private attribute; we focus on gender in our example study. The attacker can only rely on a training data set from multiple participants different from the target. This data can be gathered by companies or game developers we share our data with in exchange for a specific service. Some users might opt in to share their data with a third party to receive personalised advertisements, or they might create a user account to remove advertisements. These companies with eye tracking data can misuse the data, forward it to third parties or get hacked by external attackers. Another source for attackers to get eye tracking datasets is publicly available datasets generated for research purposes. Concretely, VR glasses are offered in gaming centres and used by multiple visitors, which we refer to as the one-device-multiple-users scenario. An attacker with access to the eye tracking data might be interested in inferring the gender of the current user to show gender-specific advertisements.

With Prior Knowledge. The second threat model assumes that the attacker has already gathered prior knowledge about the target. Observing further eye tracking data, the attacker wants to re-identify the target to inspect the target's habits. Concretely, the target might be using different user accounts or even different devices for work and leisure time (a one-user-multiple-devices scenario). We assume the attacker is able to link the target's work data to the target's identity and now wants to identify the target's data from his/her leisure activities. Again, the attacker could be a VR/AR company exploiting their data to check whether a device is only used by one person, or re-identify a user automatically to adapt device settings. Moreover, data could be released intentionally to a third party for money or unintentionally through a hack.

4.2 Differential Privacy for Eye Tracking

We propose to mitigate the privacy threats emerging from our two threat models using *differential privacy*, a well-known framework from privacy research [Dwork et al. 2014]. Differential privacy guarantees that the answer of the privacy-preserving mechanism does not depend on whether a single user contributed his/her data or not; hence, there is no way to infer further information about this user. Concretely, the answer to the question "What is the average fixation rate when reading a text?" should be almost the same, whether or not a specific user, say, Alice, has contributed her data to our database of fixation rates. We denote a differentially private mechanism by $\mathcal{M}$ and refer to Alice's data as a single data element in the database D. Typically, $\mathcal{M}$ adds random noise to "hide" each data element, which we will formalise in the following.

Definition 1 (ϵ-Differential Privacy [Dwork et al. 2006]). *A mechanism $\mathcal{M}$ provides ϵ-differential privacy if for all databases D, D' that differ in at most one element and for every $S \subseteq Range(\mathcal{M})$, we have*

$$Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon} \cdot Pr[\mathcal{M}(D') \in S]. \quad (1)$$

Differential privacy allows computing an arbitrary function g over the database, i.e. $g : \mathcal{R}^* \mapsto \mathcal{R}^d$, where d denotes the dimensionality of the output of g. For our running example, g would compute the average and output one number, hence $d = 1$. Similarly, we could define g to average over 30-second windows of fixation data and then output a vector of length d.

How much noise we have to add depends on the variance of the data between two arbitrary elements. Formally:

Definition 2 (L_1 Sensitivity [Dwork et al. 2006]). *For all functions $g : \mathcal{R}^* \mapsto \mathcal{R}^d$, the L_1 sensitivity is the smallest number Δ_g s.th. for all databases D, D' differing in one element, we have*

$$||g(D) - g(D')||_{L_1} \leq \Delta_g. \quad (2)$$

Intuitively, the sensitivity captures the maximal influence Alice's data could have on the answer to our query. In the worst case, for her privacy, Alice's data is an outlier, e.g. Alice is a very slow reader compared to all other participants. Even in this case, the difference between Alice's data and any other entry in the database must be smaller than or equal to the sensitivity. The noise to "hide" Alice's contribution is scaled to this worst case, ensuring Alice's privacy.

Next, we formalise the exponential mechanism that is one way to generate differentially private data:

Definition 3 (Exponential Mechanism [Dwork et al. 2014]). *The exponential mechanism selects and outputs an element $r \in \mathcal{R}$ in the range of permissible output elements with probability equal to (written: $r \sim$)*

$$r \sim exp(\frac{\epsilon \cdot u(x, r)}{2\Delta_u}) \quad (3)$$

where u is a utility function judging the quality of r with respect to the original data element x.

In order to apply the exponential mechanism to our example database of fixation durations, we would first need to define a utility function u and the set of permissible outputs. Valid answers to the query "What are the average fixation rates when reading a text, sampled at 30 second windows?" are vectors of length d containing real-numbered entries; thus, $\mathcal{R} = \mathbb{R}^d_{\geq 0}$. The utility function u is a measure of quality for the output r with respect to the original data entry x. The exponential mechanism ensures that high-quality outputs r are generated exponentially more often than low-quality r.

Finally, we state one theorem that allows combining several differentially private mechanisms into one.

Theorem 1 (Composition Theorem [Dwork et al. 2006]). *Let $\mathcal{M}_1, ..., \mathcal{M}_k$ be a fixed sequence of mechanisms, where each mechanism $\mathcal{M}_i$ is ϵ_i-differentially private. Then, their joint output $\mathcal{M}(D) = (\mathcal{M}_1(D), ..., \mathcal{M}_k(D))$ is ϵ-differentially private for $\epsilon = \sum_{i=1}^{k} \epsilon_i$.*

4.3 Implementing Differential Privacy

Our dataset contains data from n participants, which we refer to as $p_1, ..., p_n$. For each participant, we measure m features, $f_1, ..., f_m$ at different points in time. In summary, p_{1,f_7,t_5} denotes the value of the 7th feature at time point 5 of participant 1, and the vector $(p_{1,f_7,t_0}, ..., p_{1,f_7,t_{max,1}})$ contains all measurements of feature 7 for participant 1. Notice that the data entries available may have different lengths, i.e. $t_{max,1}$, the last time point of participant 1, may be different from another participant's last time point, e.g. $t_{max,2}$.

The sensitivity for our mechanism then depends on the range of the features, which is different across our m features. For example, feature f_{15} is the fixation duration in our dataset, and it has an estimated range of [0.11, 2.75] seconds, while f_{22}, which describes the pupil diameter size, has an estimated range of [21.9, 133.9] pixels. Therefore, we derive one privacy mechanism $\mathcal{M}_{f_i}$ for each feature separately and use the composition theorem (Theorem 1) to combine the m mechanisms into our final mechanism. The exponential mechanism requires a utility function u. We choose the L_1 distance for simplicity of the derivation:

$$u(p_{f_i}, r) = \sum_{j=1}^{t_{max,p}} |p_{f_i,j} - r_j| \quad (4)$$

According to Definition 2, the sensitivity Δ_{u,f_i} is

$$\Delta_{u,f_i} = \max_{p_{f_i}, q_{f_i}} ||(p_{f_i,t_0}, ..., p_{f_i,t_{max,p}}) - (q_{f_i,t_0}, ..., q_{f_i,t_{max,q}})||_{L_1} \quad (5)$$

i.e. the maximal difference between the data vectors of two arbitrary participants p and q for the i-th feature. Next, we unify the length by padding the data vector with the shorter length. Let $tmax$ be the maximal length: $tmax = max(t_{max,p}, t_{max,q})$. Using this and the definition of the L_1 norm:

$$\Delta_{u,f_i} \leq \max_{p_{f_i}, q_{f_i}} \sum_{j=1}^{tmax} |p_{f_i,t_j} - q_{f_i,t_j}| = tmax \cdot \delta_i \quad (6)$$

In the last step, we used the fact that we can derive the range δ_i of feature f_i, either estimated from the data or by theoretic constraints.

We rely on the exponential mechanism (see Definition 3) to obtain a vector r that is differentially private for each participant p and feature f_i:

$$r \sim exp(\frac{\epsilon_i u(p_{f_i}, r)}{2\Delta_{u,f_i}}) \overset{\text{Eq. 4}}{=} exp(\frac{\epsilon_i \sum_{j=1}^{t_{max,p}} |p_{f_i,j} - r_j|}{2 \cdot tmax \cdot \delta_i}) \quad (7)$$

To increase readability, we define $\lambda_i = \frac{\epsilon_i}{2 \cdot tmax \cdot \delta_i}$, which is constant once i and ϵ_i are fixed. We generate such a vector r from the exponential distribution by first sampling a random scalar y from the exponential distribution with location 0 and scale parameter $\frac{1}{\lambda_i}$. We derive our differentially private vector r from y as follows:

$$y = exp(\lambda_i \cdot \sum_{j=1}^{t_{max,p}} |p_{f_i,j} - r_j|) \Leftrightarrow \frac{log_e(y)}{\lambda_i} = \sum_{j=1}^{t_{max,p}} |p_{f_i,j} - r_j| \quad (8)$$

Selecting $r_j = \pm \frac{log_e(y)}{\lambda_i \times tmax} + p_{f_i,j}$ fulfils the above constraint with randomly sampled sign.

The privacy guarantee of the combined mechanism $\mathcal{M}$ is, by the composition theorem (Theorem 1), $\sum_{i=1}^{m} \epsilon_i$.

Subsampling. In order to achieve a higher privacy guarantee, we propose to subsample the data. Given a window size w, we draw one sample from $(p_{k,i,n \cdot w}, ..., p_{k,i,(n+1) \cdot w})$ for each participant k and feature i independently where $n \in \mathbb{N}$, such that the sampling windows are non-overlapping. Notice that this subsampling approach and the corresponding window size are independent of the feature generation process. This method decreases the sensitivity further by a factor of w: $\Delta_{u,f_i,w} \leq \frac{tmax}{w} \cdot \delta_i$.

5 DATA COLLECTION

Given the lack of a suitable dataset for evaluating privacy-preserving eye tracking using differential privacy, we recorded our own dataset. As a utility task, we opted to detect different document types the users read, similar to a reading assistant [Kunze et al. 2013b]. Instead of printed documents, participants read in VR, wearing a corresponding headset. The recording of a single participant consists of three separate recording sessions, in which a participant reads one out of three different documents: a comic, online newspaper, or textbook (see Figure 5). All documents include a varying proportion of text and images. Each of these documents was about a 10-minute read, depending on a user's reading skill (about 30 minutes in total).

Participants. We recruited 20 participants (10 male, 10 female) aged 21 to 45 years through university mailing lists and adverts in different university buildings on campus. Most participants were BSc and MSc students from a large range of subjects (e.g. language science, psychology, business administration, computer science) and different countries (e.g. India, Pakistan, Germany, Italy). All participants had little or no experience, with eye tracking studies and had normal or corrected-to-normal vision (contact lenses).

Apparatus. The recording system consisted of a desktop computer running Windows 10, a 24" computer screen, and an Oculus DK2 virtual reality headset connected to the computer via USB. We fitted the headset with a Pupil eye tracking add-on [Kassner et al. 2014] that provides state-of-the-art eye tracking capabilities. To have more flexibility in the applications used by the participants in the study, we opted for the Oculus "Virtual Desktop" that shows arbitrary application windows in the virtual environment. To record a user's eye movement data, we used the capture software provided by Pupil. We recorded a separate video from each eye and each document. Participants used the mouse to start and stop the document interaction and were free to read the documents in arbitrary order. We encouraged participants to read at their usual speed and did not tell them what exactly we were measuring.

Recording Procedure. After arriving at the lab, participants were given time to familiarise themselves with the VR system. We showed each participant how to behave in the VR environment, given that most of them had never worn a VR headset before. We did not calibrate the eye tracker but only analysed users' eye movements from the eye videos post hoc. This was so as not to make participants feel observed, and to be able to record natural eye movement behaviour. Before starting the actual recording, we asked participants to sign a consent form. Participants then started to interact with the VR interface, in which they were asked to read three documents floating in front of them (see Figure 5). After finishing reading a document, the experimental assistant stopped and saved the recording and asked participants questions on their current level of fatigue, whether they liked and understood the document, and whether they found the document difficult using a 5-point Likert scale (1: Strongly disagree to 5: Strongly agree). Participants were further asked five questions about each document to measure their text understanding. The VR headset was kept on throughout the recording.

After the recording, we asked participants to complete a questionnaire on demographics and any vision impairments. We also assessed their Big Five personality traits [John and Srivastava 1999] using established questionnaires from psychology. In this work we only use the given ground truth information of a user's gender from all collected (private) information, the document type, and IDs we assigned to each participant, respectively.

(a) Comic **(b) Newspaper** **(c) Textbook**

Figure 5: Each participant read three different documents: (a) comic, (b) online newspaper, and (c) textbook.

Eye Movement Feature Extraction. We extracted a total of 52 eye movement features, covering fixations, saccades, blinks, and pupil diameter (see Table 1 in the supplementary material). Similar to [Bulling et al. 2011b], we also computed wordbook features that encode sequences of n saccades. We extracted these features using a sliding window of 30 seconds (step size of 0.5 seconds).

6 EVALUATION

The overall goal of our evaluations was to study the effectiveness of the proposed differential privacy method and its potential as a building block for privacy-aware eye tracking. In these evaluations, gaze-based document type classification served as the utility task, while gender prediction exemplified an attacker without prior knowledge about the target, and user re-identification an attacker with prior knowledge.

6.1 Classifier Training

For each task, we trained a support vector machine (SVM) classifier with radial basis function (RBF) kernel and bias parameter $C = 1$ on the extracted eye movement features. We opted for an SVM due to the good performance demonstrated in a large body of work for eye-based activity recognition [Bulling et al. 2011b; Steil and Bulling 2015]. As the first paper of its kind, one goal was to enable readers to compare our results to the state of the art. We standardised the training data (zero mean, unit variance) before training the classifiers; the test data was standardised with the same parameters. Majority voting was used to summarise all classifications from different time points for the respective participant. We randomly sampled training and test sets with an equal distribution of samples for each of the respective classes, i.e. for the three document classes, two gender classes and 20 classes for user identification.

Document Type Classification. We trained a multi-class SVM for document type classification and used leave-one-person-out cross-validation, i.e. we trained on the data of 19 participants and tested on the remaining one – iteratively over all combinations – and averaged the performance results in the end. We envision that in the future, only differentially private data will be available; therefore, we applied our privacy-preserving mechanism to the training and test sets. However, currently there is non-noised data available as well: thus, we set up an additional experiment using clean data for training and noised data for testing.

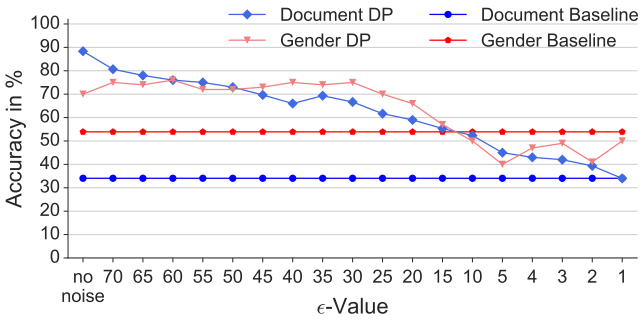

Figure 6: Performance for the threat model without prior knowledge trained on differentially private data.

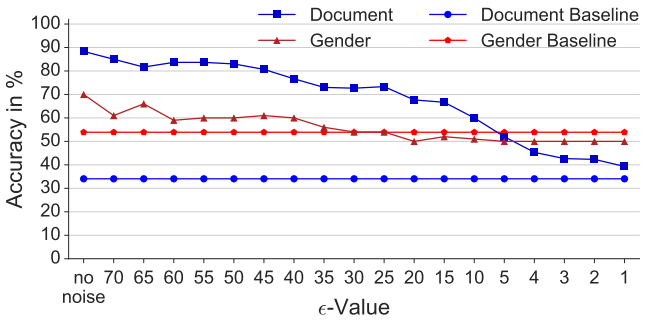

Figure 7: Performance for the threat model without prior knowledge trained on clean data.

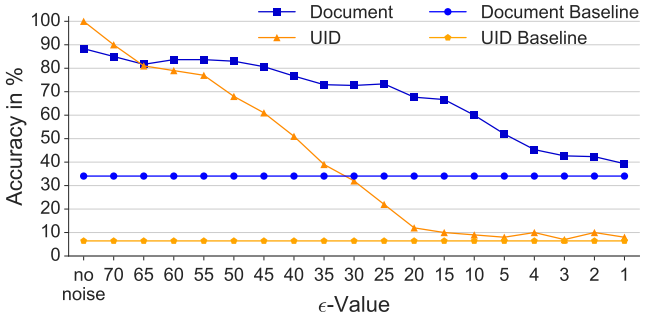

Figure 8: Performance for the threat model with prior knowledge trained on clean data.

Gender Prediction. We trained a binary SVM for gender prediction, using reported demographics as ground truth, and applied it again with a person-independent (leave-one-person-out) cross-validation. Since we are in the *without prior knowledge* threat model, we trained on differentially private and non-noised data to model both the future and current situation, as for document type classification.

User (Re-)Identification. We trained a multi-class SVM for user (re-)identification but without a leave-one-person-out evaluation scheme. Instead, we used the first half of the extracted aggregated feature vectors from each document and each participant for training. We tested on the remaining half, since here we are in the *with prior knowledge* threat model. In this scenario, we assumed a powerful attacker that was able to obtain training data from multiple people without noise and was able to map their samples to their identities. The attacker's goal was to re-identify these people when given noised samples without identity labels.

Implementing the Differential Privacy Mechanism. We applied the exponential mechanism for each of our $n = 20$ participants and for each of the $m = 52$ features, using a subsampling window size $w = 10$ to reduce sensitivity. In preliminary evaluations, we observed that subsampling alone had no negative effect on the performance of the SVM. The sensitivity for our differentially private mechanism was generated by data-driven constraints: For each feature i, we estimated δ_i by calculating the global minimum min_i and maximum max_i over all participants and time points and set $\delta_i = max_i - min_i$. This way, the sensitivity ensures privacy protection even of outliers. The noise we added in our study can be understood as reading-task-specific noise. For all f_i, we used the same ϵ_i so that the released data of the whole dataset is $\sum_{i=1}^{52} \epsilon_i$-private.

We repeated our experiments five times each and report averaged results to account for random subsampling and noise generation effects. As a performance metric, we report $Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$, where TP, FP, TN, and FN represent sample-based true positive, false positive, true negative, and false negative counts.

6.2 Without Prior Knowledge

In Figure 6, we first evaluated the gender prediction task, our example for the attacker *without prior knowledge*, trained on differentially private (noised) data (Gender DP) for decreasing ϵ values. As one might expect, decreasing ϵ, and thereby increasing the noise, negatively influences the testing performance when trained on differentially private data with $\epsilon < 30$. For $\epsilon = 15$, the performance almost drops to the chance level of 54% (random guessing in a slightly imbalanced case due to the leave-one-person-out cross-validation). We conclude that on our dataset, privacy of the participants' gender information is preserved for $\epsilon \leq 15$.

We then evaluated the impact of the noise level for this ϵ-value on utility (see Figure 6) using the SVMs trained for document type classification on noised data. As expected, noise negatively influences document type classification as well, but to a lesser extent compared to gender prediction. For privacy preservation, it is sufficient to set $\epsilon = 15$, resulting in an accuracy of about 55% for document type classification, which is still about 22% over chance level.

So far, we have assumed the SVMs were trained on noised data (Document DP). At present, to the best of our knowledge, all available eye movement datasets are not noised. To study this current situation, we trained both the gender prediction SVM and the document type classification SVM without noise and tested at various noise levels. Figure 7 shows the results of this evaluation. As can be seen, also in this scenario, privacy can be preserved: For $\epsilon = 20$, the accuracy of the gender prediction has dropped below chance level, while document type classification is still around 70%. We observed that even $\epsilon = 30$ would already preserve privacy, since training with noise seems to balance out some negative noise effects. Thus, we conclude that for both current and future situations, privacy preservation is possible while preserving most of the utility.

6.3 With Prior Knowledge

Finally, we evaluated in Figure 8 the *with prior knowledge* threat model, in which we assumed the attacker trained a SVM on the data of multiple users without noise and wanted to re-identify which

person a set of noised samples belongs to. We again added the document type classification performance to be able to judge the effects on utility. As expected, the noise on the test data disturbed the attacker's classification ability: for $\epsilon = 40$, the attacker's accuracy dropped to 50%. For $\epsilon = 15$, it dropped down almost to chance level (6.4%) while the utility preserved an accuracy of about 70%. We conclude that, in this scenario as well, it is possible to preserve a user's privacy with acceptable costs on utility.

7 DISCUSSION

7.1 Privacy Concerns in Eye Tracking

The ever-increasing availability of eye tracking to end users, e.g. in recent VR/AR headsets, in combination with the rich and sensitive information available in the eyes (e.g. on personality [Hoppe et al. 2018]), creates significant challenges for protecting users' privacy. Our large-scale online survey on privacy implications of pervasive eye tracking, the first of its kind, yielded a number of interesting insights on this important, yet so far largely unexplored, topic (see the supplementary material for the full results). For example, we found that users are willing to share their eye tracking data for medical applications, such as (early) disease detection or stress level monitoring (see Figure 2), or for services, if these improve user experience, e.g. in VR or AR (see Figure 3). On the other hand, participants refused services that use eye movement data for interest identification or shopping assistance, and a majority did not like the idea of services inferring their identity, gender, sexual preference, or race. These findings are interesting, as they suggest that users are indeed willing to relinquish privacy in return for service use. They also suggest, however, that users may not be fully aware of the fact that, and to what extent, these services could also infer privacy-sensitive information from their eyes. Our proposed differential privacy approach addresses this challenge by allowing sharing of eye movement data while protecting individual privacy.

To prevent inference of users' private attributes from eye tracking data, not every data representation is suitable. Nonetheless, we identified a clear information gap on the user side, since a majority of participants agreed to share their eye tracking data in almost every data representation (see Figure 3 in the supplementary material). Participants seemed unaware of the fact that, in particular, raw eye movement data representation is inappropriate to protect their privacy. Adding noise to this data representation would not protect their private attributes either: the added noise could easily be removed by smoothing. Instead, we recommend using statistical or aggregated feature representations that summarise temporal and appearance statistics of a variety of eye movements, such as fixation, saccades, and blinks. We are the first to propose a practical solution to this challenge by using differential privacy that effectively protects private information, while at the same time maintaining data utility.

7.2 Privacy-Preserving Eye Tracking

Informed by our survey results, we presented a privacy-aware eye tracking method in a VR setting. This is the first of its kind to quantitatively evaluate the practicability and effectiveness of privacy-aware eye tracking. For that purpose, we study 1) two realistic threat models (*with* and *without prior knowledge* about the target user), and 2) different scenarios in training with and without clean/non-noised data. We conducted an extensive evaluation on a novel 20-participant dataset and 3) demonstrated the effectiveness of the trained threat models on two example privacy-infringing tasks, namely gender inference and user identification.

Applying differential privacy mitigates these privacy threats. The fundamental principle of differential privacy is to apply appropriate noise on the data to deteriorate the accuracy of a privacy-infringing task while maintaining that of a utility task. As such, the level of noise should be smaller than the inter-class difference in the utility task but larger than that of the privacy-infringing task.

We showed in our practical evaluations that users' privacy can be preserved with acceptable accuracy of the utility task by applying differential privacy. This conclusion was consistent across different evaluation paradigms in our example study, which aimed to perform gaze-based document type classification while preserving the privacy of users' gender and identity.

Our mechanism can be used to sanitise data not only before releasing it to the public, but also in VR/AR devices themselves, since it sanitises one user at a time. Although our example study focuses only on reading, we expect our method to generalise to any other activity involving eye tracking. Due to our data-driven approach, sensitivity can be adapted so that a similar trade-off can be found. Depending on sensitivity and data vector length, the privacy level ϵ of this trade-off may differ from the presented results. Similarly, our study was evaluated on a typical HCI dataset size, and we expect our approach to generalise to larger datasets that will be available in the future, given the rapid emergence of VR and eye tracking technology.

To conclude, the proposed method is an effective and low-cost solution to preserve users' privacy while maintaining the utility task performance.

8 CONCLUSION

In this work we reported the first large-scale online survey to understand users' privacy concerns about eye tracking and eye movement analysis. Motivated by the findings from this survey, we also presented the first privacy-aware gaze interface that uses differential privacy. We opted for a virtual reality gaze interface, given the significant and imminent threat potential created by upcoming eye tracking technology equipped VR headsets. Our experimental evaluations on a new 20-participant dataset demonstrated the effectiveness of the proposed approach to preserve private information while maintaining performance on a utility task – hence, implementing the principle *ensure privacy without impeding utility*.

ACKNOWLEDGMENTS

This work was funded, in part, by the Cluster of Excellence on Multimodal Computing and Interaction (MMCI) at Saarland University, Germany, by a JST CREST research grant under Grant No.: JPMJCR14E1, Japan, as well as by the German Federal Ministry of Education and Research (BMBF) for the Center for IT-Security, Privacy and Accountability (CISPA) (FKZ: 16KIS0656).

REFERENCES

Devon Adams, Alseny Bah, Catherine Barwulor, Nureli Musaby, Kadeem Pitkin, and Elissa M Redmiles. 2018. Ethics emerging: the story of privacy and security perceptions in virtual reality. In *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*. 427–442. https://doi.org/10.13016/M2B853K5P

Roman Bednarik, Tomi Kinnunen, Andrei Mihaila, and Pasi Fränti. 2005. Eye-movements as a biometric. In *Scandinavian conference on image analysis*. Springer, 780–789. https://doi.org/10.1007/11499145_79

Andreas Bulling and Hans Gellersen. 2010. Toward Mobile Eye-Based Human-Computer Interaction. *IEEE Pervasive Computing* 9, 4 (2010), 8–12. https://doi.org/10.1109/MPRV.2010.86

Andreas Bulling, Daniel Roggen, and Gerhard Tröster. 2011a. What's in the Eyes for Context-Awareness? *IEEE Pervasive Computing* 10, 2 (2011), 48–57. https://doi.org/10.1109/MPRV.2010.49

Andreas Bulling, Jamie A. Ward, Hans Gellersen, and Gerhard Tröster. 2011b. Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 4 (2011), 741–753. https://doi.org/10.1109/TPAMI.2010.86

Andreas Bulling, Christian Weichel, and Hans Gellersen. 2013. EyeContext: Recognition of High-level Contextual Cues from Human Visual Behaviour. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. 305–308. https://doi.org/10.1145/2470654.2470697

Andreas Bulling and Thorsten O. Zander. 2014. Cognition-Aware Computing. *IEEE Pervasive Computing* 13, 3 (2014), 80–83. https://doi.org/10.1109/MPRV.2014.42

Virginio Cantoni, Chiara Galdi, Michele Nappi, Marco Porta, and Daniel Riccio. 2015. GANT: Gaze analysis technique for human identification. *Pattern Recognition* 48, 4 (2015), 1027–1038. https://doi.org/10.1016/j.patcog.2014.02.017

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*. Springer, 265–284. https://doi.org/10.1007/978-3-540-32732-5_32

Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (2014), 211–407. https://doi.org/10.1561/0400000042

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 (TOPS)* 19, 1 (2016), 1. https://doi.org/10.1145/2904018

Myrthe Faber, Robert Bixler, and Sidney K D'Mello. 2017. An automated behavioral measure of mind wandering during computerized reading. *Behavior Research Methods* (2017), 1–17. https://doi.org/10.3758/s13428-017-0857-y.

Liyue Fan and Li Xiong. 2012. Adaptively sharing time-series with differential privacy. *arXiv preprint arXiv:1202.3461* (2012).

Matthew Fredrikson, Eric Lantz, Somesh Jha, Simon Lin, David Page, and Thomas Ristenpart. 2014. Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing. In *USENIX Security Symposium*. 17–32. https://doi.org/10.1.1.469.4356

John Paulin Hansen, Anders Sewerin Johansen, Dan Witzner Hansen, Kenji Itoh, and Satoru Mashino. 2003. Command without a click: Dwell time typing by mouse and gaze selections. In *Proceedings of Human-Computer Interaction–INTERACT*. 121–128. https://doi.org/10.1.1.535.168

Eckhard H Hess and James M Polt. 1960. Pupil size as related to interest value of visual stimuli. *Science* 132, 3423 (1960), 349–350. https://doi.org/10.1126/science.132.3423.349

Philip S Holzman, Leonard R Proctor, Deborah L Levy, Nicholas J Yasillo, Herbert Y Meltzer, and Stephen W Hurt. 1974. Eye-tracking dysfunctions in schizophrenic patients and their relatives. *Archives of general psychiatry* 31, 2 (1974), 143–151. https://doi.org/10.1001/archpsyc.1974.01760140005001

Sabrina Hoppe, Tobias Loetscher, Stephanie A Morey, and Andreas Bulling. 2018. Eye movements during everyday behavior predict personality traits. *Frontiers in Human Neuroscience* 12 (2018), 105. https://doi.org/10.3389/fnhum.2018.00105

J Thomas Hutton, JA Nagel, and Ruth B Loewenson. 1984. Eye tracking dysfunction in Alzheimer-type dementia. *Neurology* 34, 1 (1984), 99–99. https://doi.org/10.1212/WNL.34.1.99

Oliver P John and Sanjay Srivastava. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research* 2, 1999 (1999), 102–138.

Paweł Kasprowski. 2004. Human identification using eye movements. *Praca doktorska, Politechnika Œląska* (2004). https://doi.org/10.13140/RG.2.1.3466.3924

Pawel Kasprowski and J Ober. 2003. Eye movement tracking for human identification. In *6th World Conference BIOMETRICS*.

Pawel Kasprowski and Józef Ober. 2005. Enhancing eye-movement-based biometric identification method by using voting classifiers. In *Biometric Technology for Human Identification II*, Vol. 5779. International Society for Optics and Photonics, 314–324. https://doi.org/10.1117/12.603321

Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Adj. Proc. ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. 1151–1160. https://doi.org/10.1145/2638728.2641695

Tomi Kinnunen, Filip Sedlak, and Roman Bednarik. 2010. Towards task-independent person authentication using eye movement signals. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ACM, 187–190. https://doi.org/10.1145/1743666.1743712

Oleg V Komogortsev and Corey D Holland. 2013. Biometric authentication via complex oculomotor behavior. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. IEEE, 1–8. https://doi.org/10.1109/BTAS.2013.6712725

Oleg V Komogortsev, Sampath Jayarathna, Cecilia R Aragon, and Mechehoul Mahmoud. 2010. Biometric identification via an oculomotor plant mathematical model. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ACM, 57–60. https://doi.org/10.1145/1743666.1743679

Craig A Kuechenmeister, Patrick H Linton, Thelma V Mueller, and Hilton B White. 1977. Eye tracking in relation to age, sex, and illness. *Archives of General Psychiatry* 34, 5 (1977), 578–579. https://doi.org/10.1001/archpsyc.1977.01770170088008

Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura, and Koichi Kise. 2013a. The Wordometer–Estimating the Number of Words Read Using Document Image Retrieval and Mobile Eye Tracking. In *12th International Conference on Document Analysis and Recognition (ICDAR)*. 25–29. https://doi.org/10.1109/ICDAR.2013.14

Kai Kunze, Katsutoshi Masai, Masahiko Inami, Ömer Sacakli, Marcus Liwicki, Andreas Dengel, Shoya Ishimaru, and Koichi Kise. 2015. Quantifying reading habits: counting how many words you read. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 87–96. https://doi.org/10.1145/2750858.2804278

Kai Kunze, Yuzuko Utsumi, Yuki Shiga, Koichi Kise, and Andreas Bulling. 2013b. I know what you are reading: recognition of document types using mobile eye tracking. In *Proceedings of the 2013 International Symposium on Wearable Computers*. ACM, 113–116. https://doi.org/10.1145/2493988.2494354

Daniel J Liebling and Sören Preibusch. 2014. Privacy considerations for a pervasive eye tracking world. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 1169–1177. https://doi.org/10.1145/2638728.2641688

Anthony J Maeder and Clinton B Fookes. 2003. A visual attention approach to personal identification. (2003).

G Matthews, W Middleton, B Gilmartin, and MA Bullimore. 1991. Pupillary diameter and cognitive load. *Journal of Psychophysiology* (1991).

Apostolos Pyrgelis, Carmela Troncoso, and Emiliano De Cristofaro. 2017. Knock Knock, Who's There? Membership Inference on Aggregate Location Data. *arXiv preprint arXiv:1708.06145* (2017).

Nazir Saleheen, Supriyo Chakraborty, Nasir Ali, Md Mahbubur Rahman, Syed Monowar Hossain, Rummana Bari, Eugene Buder, Mani Srivastava, and Santosh Kumar. 2016. mSieve: differential behavioral privacy in time series of mobile sensor data. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 706–717. https://doi.org/10.1145/2971648.2971753

Negar Sammaknejad, Hamidreza Pouretemad, Changiz Eslahchi, Alireza Salahirad, and Ashkan Alinejad. 2017. Gender classification based on eye movements: A processing effect during passive face viewing. *Advances in cognitive psychology* 13, 3 (2017), 232. https://doi.org/10.5709/acp-0223-1

Julian Steil and Andreas Bulling. 2015. Discovery of everyday human activities from long-term visual behaviour using topic models. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 75–85. https://doi.org/10.1145/2750858.2807520

Sophie Stellmach and Raimund Dachselt. 2012. Look & touch: gaze-supported target acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2981–2990. https://doi.org/10.1145/2208636.2208709

Marc Tonsen, Julian Steil, Yusuke Sugano, and Andreas Bulling. 2017. InvisibleEye: Mobile Eye Tracking Using Multiple Low-Resolution Cameras and Learning-Based Gaze Estimation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 106. https://doi.org/10.1145/3130971

Roel Vertegaal et al. 2003. Attentive user interfaces. *Commun. ACM* 46, 3 (2003), 30–33. https://doi.org/10.1145/636772.636794

Yongtuo Zhang, Wen Hu, Weitao Xu, Chun Tung Chou, and Jiankun Hu. 2018. Continuous Authentication Using Eye Movement Response of Implicit Visual Stimuli. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 177. https://doi.org/10.1145/3161410

Tianqing Zhu, Gang Li, Wanlei Zhou, and S Yu Philip. 2017. Differentially private data publishing and analysis: a survey. *IEEE Transactions on Knowledge and Data Engineering* 29, 8 (2017), 1619–1638. https://doi.org/10.1109/TKDE.2017.2697856