Your Eyes Tell: Leveraging Smooth Pursuit for Assessing Cognitive Workload

Thomas Kosch[1], Mariam Hassib[1], Paweł W. Woźniak[2], Daniel Buschek[1], Florian Alt[1,3]
[1]LMU Munich, Munich, Germany, {firstname.lastname}@ifi.lmu.de
[2]University of Stuttgart, Stuttgart, Germany, {firstname.lastname}@vis.uni-stuttgart.de
[3]Munich University of Applied Sciences, Munich, Germany, {firstname.lastname}@hm.edu

ABSTRACT
A common objective for context-aware computing systems is to predict how user interfaces impact user performance regarding their cognitive capabilities. Existing approaches such as questionnaires or pupil dilation measurements either only allow for subjective assessments or are susceptible to environmental influences and user physiology. We address these challenges by exploiting the fact that cognitive workload influences smooth pursuit eye movements. We compared three trajectories and two speeds under different levels of cognitive workload within a user study (N=20). We found higher deviations of gaze points during smooth pursuit eye movements for specific trajectory types at higher cognitive workload levels. Using an SVM classifier, we predict cognitive workload through smooth pursuit with an accuracy of 99.5% for distinguishing between low and high workload as well as an accuracy of 88.1% for estimating workload between three levels of difficulty. We discuss implications and present use cases of how cognition-aware systems benefit from inferring cognitive workload in real-time by smooth pursuit eye movements.

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation:
User Interfaces – Input Devices and Strategies

Author Keywords
Cognitive Workload; Mental Workload; Eye Tracking;
Smooth Pursuit; Cognition-Aware User Interfaces;
Workload-Aware Computing

INTRODUCTION
Eye gaze-based interactive systems hold a lot of promise for cognition-aware interaction [10]. The human eye is the central organ of the body when it comes to perception and information processing. Complemented by previous research, eye gaze as an input for interactive systems has been extensively explored [16]. Recently, smooth pursuit eye movements have been utilized as an alternative eye movement input modality to interactive systems [22, 61]. This interaction technique overcomes the need for precise calibration and training of the user before interaction and can be deployed without any user effort. Previous research has used this technique to enable calibration-free interaction with distant displays [42, 62], and mobile devices [17, 35]. Smooth pursuit eye movements cannot be faked by the users since they require locking onto a moving target [59]. This robustness against false positives is another benefit of using smooth pursuit eye movement as an interaction technique.

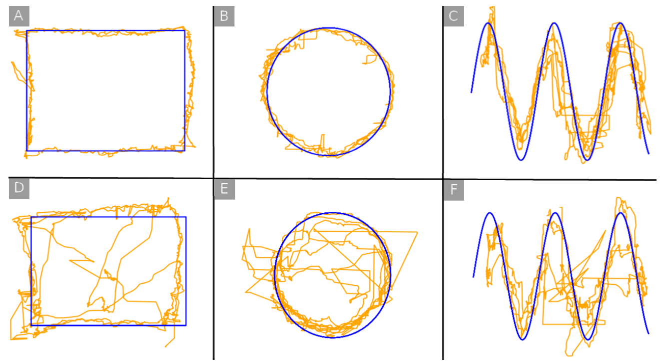

Figure 1. Smooth pursuit recordings under different levels of cognitive workload for three trajectories. The blue line shows the displayed trajectory, while the orange line visualizes the gaze path. (A), (B), and (C) show the gaze path during low cognitive workload phases, while (D), (E), and (F) shows the gaze path during perception of high workload.

Research has shown that the behavior of the eye is strongly affected by psychological [47, 48] and psychophysiological states of the human body [52, 66]. One such state is cognitive workload, which has a remarkable impact on eye movements [5, 26, 54]. Figure 1 shows how cognitive workload affects smooth pursuit eye movements. Driving in a stressful context or performing multitasking during cognitively demanding tasks under time pressure are just two examples.

Quantitative measures of cognitive workload include questionnaires, such as the NASA-TLX [19, 20] or the Driver Activity Load Index (DALI) [41], which are often used in HCI to evaluate interfaces. Yet these questionnaires are prone to the interpretation of the questions by the user and only allow for an assessment at the end of the task, hence providing rather coarse-grained insights.

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ISBN 978-1-4503-5620-6/18/04...$15.00
DOI: https://doi.org/10.1145/3173574.3174010

At the same time, physiological data can be used to assess cognitive workload [14]. In particular, previous work showed that under controlled conditions, pupil dilation provides an estimate of cognitive load [43, 45]. However, pupil dilation is highly sensitive to light conditions, which change in outdoor environments, and hence cannot easily be applied in many ubiquitous computing settings. Other touchless methods which infer cognitive workload via physiological sensing involve heart rate [31] or facial skin temperature measures [1, 53]. However, these modalities depend on several other factors, such as constant temperatures or highly controlled environments.

In this work, we propose an approach that exploits smooth pursuit eye movements to assess cognitive workload. Since the term *cognitive workload* is used in various fields and its definition varies widely, we investigate the correlation between smooth pursuit eye movements and working memory. Working memory describes the cognitive processing system that stores information temporarily and affects higher-order cognitive functions such as understanding and information processing [4]. Smooth pursuit describes eye movements which occur as the eyes closely follow a moving object. These movements are evaluated in a number of studies used for user interface element selection [17, 59] or intuitive eye tracker calibration [29, 61]. By conducting a user study, we found that smooth pursuit can be used for contactless assessment of cognitive workload with an accuracy of up to 99.5%. Thereby, the need of body-worn sensors is obviated.

CONTRIBUTION STATEMENT

The contribution of this paper is threefold: (1) We report on a user study that investigates the impact of cognitive workload on smooth pursuit eye movements. Based on the insights from the user study, (2) we build a classifier to approximate the level of cognitive workload using gaze differences during smooth pursuit eye movements. (3) Finally, we discuss how these findings can be used during the evaluations that require the assessment of cognitive workload as well as for the design of cognition-aware interactive systems.

RELATED WORK

In this section, we present related work concerning existing research on using smooth pursuit eye movements for interaction and the influence of cognitive workload on eye movements.

Interacting with Smooth Pursuit

Using eye gaze as an input for ubiquitous interactive systems has been extensively explored. Duchowski [15, 16] performed a literature survey about current eye tracking applications and technologies. Focusing on usability, related research shows how eye tracking technologies evolved until they became usable in human-computer interaction research [24, 46]. Recent research has addressed the use of eye gaze for selection tasks [40, 50], which can be used by physiologically impaired individuals to perform input on computers [22]. Text input via eye input has been researched using dwell time [6, 37], off-screen targets [23], or eye movements along the y-axis on a display [57]. The usage of such methods was also evaluated in real-world scenarios [13, 65]. Furthermore, eye movements and eye gestures can be used to interact with devices.

Recently, smooth pursuit eye movement [22, 61] has been used as an alternative gaze-based input modality to interactive systems. This interaction technique overcomes the need for precise calibration and training of the user before interaction and further supports ubiquitous interaction. Several researchers have investigated how smooth pursuits can be used in an intuitive way. People cannot pretend to perform smooth pursuit movements since they require locking onto a moving target [59]. This reduces the likeliness of detecting false positives during an interaction.

Another approach to leverage smooth pursuit as an input modality has been researched by Esteves et al. [17]. They used a mobile eye tracker to enable hands-free interaction with a smartwatch by showing moving dots on the smartwatch display. Results show robustness against false positives regarding input, no need for calibrating an eye tracker, and efficient hands-free interaction. Using smooth pursuit to interact with physical real-world devices also showed promising results. Velloso et al. [59] developed and evaluated an object-driven system, which can be operated by only performing smooth pursuit. As soon as a user approaches an object such as a fan or windmill, it begins to present moving targets which trigger actions as soon as selected by smooth pursuit.

Researchers have also used smooth pursuit to calibrate eye trackers. Pfeuffer et al. [44] investigated this approach by using animations on a display to implicitly calibrate an eye tracker. Their results state that a 5-point calibration achieved a higher detection rate and required less calibration time. They concluded that smooth pursuits calibration provides better usability and flexibility for eye tracker calibration on small displays.

Vidal et al. [61] proposed different applications for smooth pursuit eye interaction, such as selection tasks or password authentication. The implemented applications showed a fast selection and completion time for different tasks. Since smooth pursuit can be performed with almost every kind of animated trajectory, more recently Khamis et al. [29] investigated which trajectories are most suitable for smooth pursuit interaction and calibration when showing text-based content on a display. While showing a question in the top left corner of a screen, they also showed several answers to the question on the same screen displayed in different trajectories. Schenk et al. [49] used smooth pursuit as an element selection mechanism in desktop settings to avoid the Midas touch problem [25]. Similarly, Lohr and Komogortsev [36] compared smooth pursuit-based input against dwell-based input approaches. Significant faster selection of elements was achieved when using smooth pursuit at the cost of likely unwanted selections.

Impact of Cognitive Workload on Eye-based Properties

Previous research has shown that the behavior of the eye is strongly affected by the psychological [47, 48] and psychophysiological [52, 66] state of the human body. One such state is cognitive workload, which has a shown impact on eye movements [5, 54].

Researchers found a relation between pupillary responses and cognitive demand [2, 21]. During a task comprising different task complexities, the pupil diameter of participants was measured as an indicator for cognitive workload. Results show increasing pupil dilation with increasing task difficulty. Pfleging et al. [45] created a model to estimate cognitive workload under different lighting conditions based on pupil dilation. The feasibility of measuring pupil dilation using a remote eye tracker has been investigated by Klingner et al. [30]. Their findings show a lower accuracy compared to mobile eye trackers due to noise and head movements. Kruger et al. [32] investigated eye behavior when perceiving a stimulus with and without subtitles using electroencephalography and pupillary measurements. In an experiment with two groups, lower cognitive workload was measured within the group perceiving subtitles than in the group lacking subtitles.

Liang and Lee [34] compared the efficiency of different machine learning algorithms to estimate distractions during driving tasks based on saccadic eye movements. Higher frequencies of saccadic eye movements have been used as an assessment for cognitive workload. Best results were achieved using a support vector machine (SVM) for machine learning. However, their findings address saccadic eye movements only and do not examine the effects of smooth pursuit eye movements under cognitive workload.

Benedetto et al. [7] investigated the correlation between blink duration and visual and cognitive workload of a driver operating a car in a simulated environment. Participants had to perform a Lane Change Test [38] while doing an in-vehicle information system task [60] at the same time. Their findings show a lower blink duration compared to their baseline task than in the cognitively demanding task. Ahlstrom and Friedman-Berg [3] also found a significant correlation between shorter blink durations and cognitive workload.

Stuyven et al. [54] investigated the impact of cognitive workload on saccadic eye movements. Their findings show increased occurrence of saccadic eye movements when inducing cognitive workload. Tsai et al. [55] investigated how eyes behave under cognitive workload while performing a paced auditory serial addition task [18], however, their work did not investigate the impact on smooth pursuit eye movements. Recently, Zagermann et al. [64] developed a model and showed concepts to derive cognitive workload from eye behavior, such as saccades, fixations, pupil dilation, and eye blinks. Cognitive workload influences microsaccadic eye movements [51]. They found that microsaccades occur more frequently with higher perceived workload during a non-visual task.

The voluntary involvement of smooth pursuit eye movements has been researched by Barnes [5]. He showed that cognitive processes in smooth pursuit eye movements are even involved without voluntary participation. Important factors for following moving stimuli were attention and awareness, which trigger the process of smooth pursuit on a neuro-scientific level. Therefore, smooth pursuit eye movements are voluntary up to a certain degree. In contrast, Collewijn and Tamminga [11] investigated contexts in which smooth pursuit movements are voluntary. They used different targets and backgrounds to investigate smooth pursuit performances of humans. Contreras et al. [12] researched eye-target synchronization performance of people with traumatic brain injury. People suffering from traumatic brain injuries show a worse performance when performing smooth pursuit in terms of deviation points apart from a shown moving object on a display.

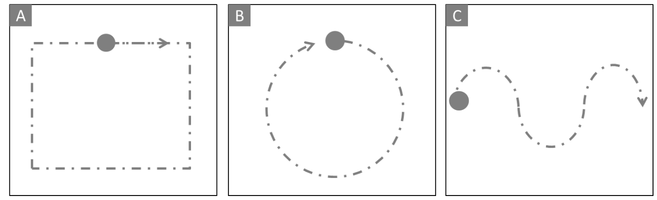

Figure 2. Three different trajectories chosen for evaluation: (A) Rectangular trajectory, (B) Circular trajectory, (C) Sinusoidal trajectory. The size of the moving object in the experiment was 10 pixels.

Summary
Previous work has investigated how smooth pursuit can be leveraged for calibration and interaction. However, to our knowledge, no prior work has proposed to use work to propose using smooth pursuit eye movements as a measurement to derive cognitive workload. This opens new opportunities for researchers and practitioners alike. In particular, using smooth pursuits as a real-time measurement for cognitive workload is valuable in the context of evaluating interactive systems as well as for developing cognitively adaptive systems.

STUDY
To understand the impact of cognitive workload on smooth pursuit eye movements, we designed a lab user study where we induced cognitive workload while participants performed smooth pursuit eye movements.

Independent Variables
In our experiment on investigating the impact of cognitive workload on eye movements, we explore the influence of three independent variables when interacting with smooth pursuit systems: (1) trajectory type [29], (2) speed of the stimulus [28], and (3) task difficulty. In the following, we describe these three independent variables in detail.

Trajectories
Since smooth pursuit eye movements can be triggered by showing a stimulus moving along a particular trajectory, we implemented a rectangular, a circular, and a sine wave animation to produce this effect. We have chosen these trajectories based on previous research [28, 29, 61]. Furthermore, the chosen trajectories may have a physiological effect paired with the current task difficulty, since the human eye has six muscles responsible for horizontal and vertical movements as well as eye rotations [33]. Rectangular trajectories demand muscles on the left and right side of the eye for horizontal movements, while vertical movements demand the upper and lower eye muscles. Circular and sinusoidal eye movements demand four muscles around the eye. Rectangular trajectories require horizontal or vertical movements only, while circular and sinusoidal include diagonal movements as well. In the context of leveraging smooth pursuit as interaction modality,

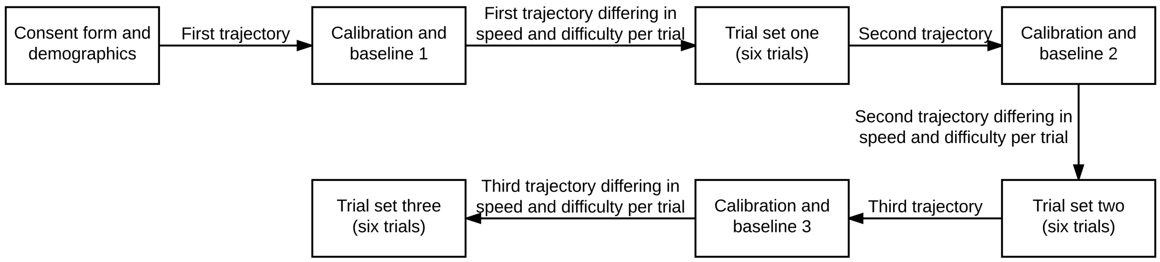

Figure 3. Study procedure. First, consent approval and demographic data were collected. Afterwards, a baseline task was conducted followed by a set of six trials, where each trial differs in trajectory speed and task complexity. This procedure is repeated for every remaining trajectory type.

the presented trajectories are used to distinguish between different user inputs. The chosen trajectories are depicted in Figure 2.

Speed of Stimulus

We compare two different speeds at which stimuli are moving based on previous research [28]: 450 $\frac{px}{s}$ (slow[1]) and 650 $\frac{px}{s}$ (fast[2]). Displaying slower or faster animations while experiencing cognitive workload can lead to different performances since eye muscles have to deal with different strains per speed and trajectory [33]. Furthermore, trajectory speeds can be used to differentiate user input, since eye movements adapt to different speeds.

Task Difficulty

To induce cognitive workload, we use an auditory delayed digit recall *N*-back task from Mehler et al. [39] with an English-spoken number set. The *N*-back task is commonly used task to artificially elicit working memory resources [4], a component of cognitive workload which strains temporal memory capacities and affects secondary task performances negatively [8, 9, 58].

Throughout the study, we use the *N*-back task to manipulate cognitive workload by demanding working memory with different difficulty levels [27]. For each trial, participants hear randomized numbers consisting of ten digits between 0 and 9. Hereby, *N* corresponds to the *N*-last digit. After hearing the number, participants have to say out loud the digit they heard *N* digits ago. In our experiment, we use a *1*-back, *2*-back, and *3*-back tasks to induce cognitive workload. By increasing *N*, more digits have to be remembered, hence increasing task difficulty. To collect baseline measures, participants were asked to follow a trajectory without performing an *N*-back task. Table 1 shows an example of the *N*-back task.

Apparatus

The study was conducted in a quiet lab with no windows, where lighting conditions were fixed. The setting was spatially divided into an experimenter area and a participant area. A separator divided both areas. While the experimenter controlled the experiment using a laptop, the participant saw an animated trajectory on a 22 inch screen with a resolution of

Heard number	5	8	3	4	3	9	1
Number to say (*1*-back task)		5	8	3	4	3	9
Number to say (*2*-back task)			5	8	3	4	3
Number to say (*3*-back task)				5	8	3	4

Table 1. Example of the auditory delayed digit recall *N*-back task. Participants have to remember the *N*-th number back of a spoken number sequence and say the number out loud.

1680×1050 pixels and a refresh rate of 60 Hertz. Eye gaze data was collected using a RED250 from SensoMotoric Instruments with a sample rate of 250 Hertz. No filter was applied to the captured gaze data, thus raw gaze data was recorded only. We used a Holosonic Audio Spotlight 24i directed speaker to provide the auditory delayed digit recall *N*-back task.

Method and Measures

We used a repeated measures design with three independent variables as described in the previous section; namely trajectory (rectangular, circular, sine wave), speed (slow, fast), and cognitive workload (no task, *1*-back task, *2*-back task, *3*-back task). Each experiment consisted of three sessions, where animated trajectories were changed for each session. The animated trajectory consisted of a white dot with a diameter of 10 pixels. The background was set to gray (RGB: [128,128,128]) to avoid eye exhaustion caused by screen brightness.

Before starting a new session, the eye tracker was calibrated to retrieve gaze points for later analysis of gaze deviations between baseline and smooth pursuit eye movements. Each session began with a 30 second baseline trial, where the animated trajectory used for the session was shown to participants without inducing cognitive workload. This allowed us to estimate eye movement differences from the displayed trajectory when no cognitive workload was present. We chose a slow speed (450 $\frac{px}{s}$) to make participants familiar with the displayed trajectory.

Cognitive workload was induced by providing a task difficulty using the auditory *N*-back task described in the prior section while showing the animated trajectory with a certain speed. Task difficulty and speed were counterbalanced during a session according to the Latin square, while each session showed the same trajectory. The order of sessions was counterbalanced participant-wise according to the Latin square. This resulted in seven trials per session, including the baseline task.

[1]This corresponds to 17.14° per second at a viewing distance of approximately 50 centimeters
[2]This corresponds to 24.76° per second at a viewing distance of approximately 50 centimeters

Paired Wilcoxon Signed-Rank Test		Significance
Circle Baseline	Circle *1*-back fast	$p = 0.002$
Circle Baseline	Circle *2*-back fast	$p = 0.002$
Circle Baseline	Circle *3*-back fast	$p = 0.003$
Circle Baseline	Circle *2*-back slow	$p = 0.004$
Circle Baseline	Circle *3*-back slow	$p = 0.005$
Sine Baseline	Sine *1*-back fast	$p < 0.001$
Sine Baseline	Sine *2*-back fast	$p < 0.001$
Sine Baseline	Sine *3*-back fast	$p < 0.001$
Sine Baseline	Sine *1*-back slow	$p = 0.003$
Sine Baseline	Sine *2*-back slow	$p = 0.002$
Sine Baseline	Sine *3*-back slow	$p = 0.002$

Table 2. Summary of significant results. Comparisons between other conditions did not result in significant differences.

Running all three sessions, the experiment comprised 21 trials per participant. Overall, the duration of each trial was 25 to 30 seconds, dependent on the length of the spoken number for tasks including task difficulty. After completing a trial, participants were asked to fill out a NASA-TLX questionnaire to assess subjectively perceived cognitive workload. Participants took a 30 second long break afterwards. Figure 3 shows an illustration of the study procedure.

Participants

We recruited 20 participants (9 female, 11 male), aged between 22 and 34 years ($M = 27.5$, $SD = 3.13$). Before the experiment, each participant signed a consent form and provided their demographic data. All participants were computer science students or researchers. All participants had normal or corrected-to-normal vision. Participants were recruited through university mailing lists. They received sweets and five Euro as compensation. The duration of the study was approximately 30 minutes. We explained the purpose of the study and tasks to the participants and informed them they could exit the study at any point. Participants signed informed-consent forms and were seated in a comfortable chair, approximately 50 centimeters in front of the display before the experimental setup. Due to technical issues, two participants were excluded from the analysis, as their gaze data was not recorded properly.

RESULTS

We analyze our data to compare the impact of trajectory type, speed, and task difficulty on smooth pursuit eye movements. We report on quantitative results by comparing measured eye gaze data with the showed trajectory. This is complemented by a subjective analysis through NASA-TLX questionnaires.

Smooth Pursuit Differences and Cognitive Workload

To evaluate the effect of different task difficulties on smooth pursuit performances, we based our analysis on pixel differences between coordinates of the displayed trajectory and measured eye gaze position at the screen. More formally, we calculated the difference between two coordinates p and q using the Euclidean distance with the formula

$$D = \sqrt{(p_{t,x} - q_{t,x})^2 + (p_{t,y} - q_{t,y})^2} \qquad (1)$$

where p and q depict a two-dimensional vector comprising the baseline coordinates and measured gaze coordinates from participants. We introduce the variable t describing the temporal dependency between eye gaze and displayed stimulus as the used 250 Hz eye tracker might introduce a temporal offset of four milliseconds, which is below the perceptual threshold for interaction. Differences were normalized with respect to the maximum gaze deviation from the shown trajectory. To enable a descriptive analysis, normalized coordinates were averaged for all participants and conditions. Lowest mean eye gaze deviations (M_{pd}) were measured for all slow trajectories, where no cognitive workload was induced (Rectangle: $M_{pd} = 9.14, SD = 6.75$, Circle: $M_{pd} = 9.51, SD = 8.03$, Sine: $M_{pd} = 4.37, SD = 1.60$). Fast rectangular trajectories using a *3*-back task ($M_{pd} = 13.25, SD = 11.06$), fast circular trajectories using a *2*-back task ($M_{pd} = 14.54, SD = 7.14$) and fast sinusoidal trajectories using a *3*-back task ($M_{pd} = 9.85, SD = 3.53$) led to highest mean eye gaze deviations.

Applying a Shapiro-Wilk test on the mean data set showed a non-normal distribution for all conditions (all $p < 0.05$). A Friedman test showed no significant differences between various levels of cognitive workload and gaze deviations of smooth pursuit eye movements for slow rectangles ($\chi^2(3) = 3.000, p = 0.392$). However, a Friedman test found significant differences within various levels of cognitive workload for fast rectangles ($\chi^2(3) = 11.867, p = 0.008$), slow circles ($\chi^2(3) = 18.867, p < 0.001$), fast circles ($\chi^2(3) = 29.400, p < 0.001$), slow sine waves ($\chi^2(3) = 14.667, p = 0.002$) and fast sine waves ($\chi^2(3) = 30.667, p < 0.001$). We conducted a Wilcoxon signed-rank post hoc test to find significant differences between pairs of task difficulties including baseline trials and normalized gaze deviations after applying a Bonferroni correction (significance level set to $p < 0.0083$). A summary of significant results can be depicted from Table 2. Further, the Cohen's d effect size values of the significant statistical comparisons ranged between $d = 0.66$ and $d = 0.88$.

To visualize averaged relative differences when performing smooth pursuit eye movements, all averaged relative eye movement differences per participant and per condition were plotted. This resulted in one data point per participant and per trial, or 21 data points per participant. The averaged plot is depicted in Figure 4, where the y-axis depicts the gaze deviations from the shown trajectory in percent. The bottom x-axis is annotated with task difficulties and the top x-axis describes trajectory velocities. The trajectory type is color coded.

Gaze deviations obtained from baseline trials do not scatter in contrast to gaze deviations measured from trials with task difficulty for circular and sinusoidal trajectories. Furthermore, gaze deviations increase between slow and fast trajectories when raising the task difficulty. However, results show important constraints when assessing cognitive workload from smooth pursuit eye movements. Rectangular trajectories show less eye gaze deviations when increasing task difficulty compared to baseline trials. Thus, evaluating the presence of cognitive workload using rectangular smooth pursuit eye movements is less accurate compared to circular and sinusoidal trajectories. Depending on the setting of a smooth pursuit driven

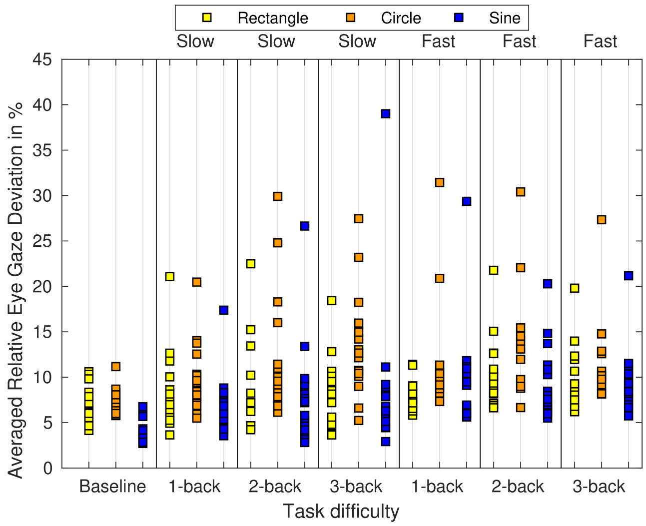

Figure 4. Scatter plot of the mean gaze deviation per participant and condition. Each dot represents the mean gaze deviations between the recorded and showed trajectory of a participant. The baseline measurements show a constant behavior and do not scatter apart from the displayed trajectory. Compared to the baseline tasks, circular and sinusoidal trajectories scatter along the y-axis with increasing task difficulty.

user interface, rectangular trajectories should be favored when accurate input is required even when the user is impacted by cognitive workload. Circular and sinusoidal trajectories show higher gaze deviations under cognitive workload. Circular and sinusoidal trajectories show clear differences between trials with and without cognitive workload. Such trajectories may be used to determine the existence of cognitive workload.

Subjective Analysis of Cognitive Workload

The lowest NASA-TLX scores per trajectory were subjectively perceived by participants when showing fast rectangles during a *1*-back task ($M = 6.27,\ SD = 4.03$), slow circles during a *1*-back task ($M = 7.13,\ SD = 3.87$), and slow sine waves during a *1*-back task ($M = 5.78,\ SD = 3.41$). Per trajectory, the highest NASA-TLX scores were measured when displaying fast rectangles during a *3*-back task ($M = 12.29,\ SD = 3.97$), fast circles during a *3*-back task ($M = 13.28,\ SD = 3.68$), and fast sine waves during a *3*-back task ($M = 12.63,\ SD = 3.42$). A repeated measures ANOVA showed statistically significant differences between different levels of cognitive workload and NASA-TLX score ($F(2,321) = 68.503, p < 0.001$). However, no statistically significant differences were found for different speeds ($F(1,322) = 2.035, p = 0.211$) and displayed trajectories ($F(2,321) = 21.461, p = 0.722$) compared to NASA-TLX scores. Figure 5 illustrates the averaged values from the obtained NASA-TLX scores.

The quantitative analysis shows significant outcomes regarding gaze deviations with different stimuli speeds and *N*-back complexities. To address the impact of different variables on the individual's subjective perception, we conducted a statistical analysis to investigate single NASA-TLX items in correlation to the stimulus speed and *N*-back difficulty. We compare the single NASA-TLX items grouped by the two different speeds. Our results show a significant difference for the *physical demand* ($p = 0.006, M_{slow} = 5.11, M_{fast} = 5.69$), *temporal*

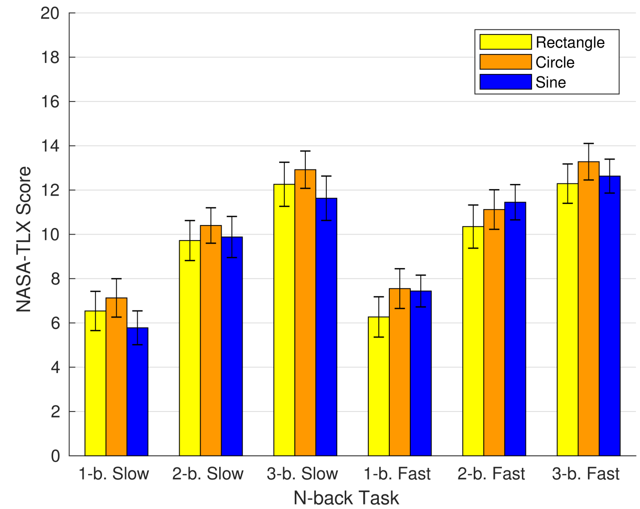

Figure 5. Mean NASA-TLX score for different trajectories, task difficulties, and speeds. The error bars depict the standard error. Increasing task difficulties led to higher NASA-TLX scores.

demand ($p = 0.003, M_{slow} = 7.29, M_{fast} = 8.27$), and *effort* ($p = 0.001, M_{slow} = 9.53, M_{fast} = 10.29$) scales. However, no significant difference was found in the mental load scale ($p > 0.05$).

To show that different *N*-back difficulties were responsible for higher perceived cognitive workload, we conducted a Wilcoxon signed-rank test to compare the mental NASA-TLX items between the different task complexities. We found significant differences between all three *N*-back difficulties (all $p < 0.001$) for the *mental demand* scales ($M_{N=1} = 5.84, M_{N=2} = 10.56, M_{N=3} = 13.84$). Further, we investigate whether the relationship between NASA-TLX scores and gaze deviations was linear using a Pearson correlation. No correlation was found ($0.11 < r < 0.15, \forall N \in \{1,2,3\}$). Thus, it appears that the relationship between the variables is more complex i.e. non-linear. This shows the need to understand how smooth pursuit affects subjectively perceived workload.

PREDICTING COGNITIVE WORKLOAD

Results from the study showed significantly increased eye movements during cognitive workload for circular and sinusoidal trajectories. The perception of cognitive workload by the individual is supported by subjective ratings of participants using NASA-TLX questionnaires. We train a classifier which predicts cognitive workload from smooth pursuit eye movements. We investigate the performance of person-dependent and person-independent classification for the different experimental conditions.

Attributes, Instances, and Classes

Data preprocessing was necessary before training a predictive model. We removed the first two seconds per trial to avoid distortions in the signal caused by participants initially searching for the stimulus. We used the collected gaze data and normalized it with respect to the coordinates of the shown trajectory. We then calculate the Euclidean distance (see Equation 1) between each coordinate of the normalized displayed trajectory and the normalized gaze points recorded from participants.

Stimulus	Binary Pers.-Indep. Accuracy	Precision	Recall	F1	Multilabel Pers.-Indep. Accuracy	Precision	Recall	F1	Multilabel Pers.-Dep. Accuracy	F1
Rectangle Slow	75.0%	37.5%	50.0%	42.9%	40.3%	26.9%	40.0%	30.5%	48.6%	52.4%
Rectangle Fast	76.4%	40.1%	52.8%	46.0%	54.2%	44.0%	54.1%	47.2%	69.4%	74.2%
Circle Slow	95.8%	89.6%	91.7%	90.5%	40.3%	26.9%	40.2%	30.9%	69.5%	60.0%
Circle Fast	97.2%	93.0%	94.4%	93.7%	55.6%	45.8%	55.6%	48.5%	85.0%	74.4%
Sine Slow	80.1%	55.6%	64.8%	59.8%	43.1%	30.6%	43.0%	34.5%	79.3%	76.3%
Sine Fast	99.5%	98.9%	99.2%	99.0%	59.7%	51.9%	59.7%	54.4%	88.1%	84.5%

Table 3. Accuracy, precision, recall, and F1 scores of the binary and multi-label person-independent (pers.-indep.) as well as person-dependent (pers.-dep.) classifications. Fast circular and sinusoidal trajectories yield higher classification performances than slower linear trajectories.

These gaze deviations are defined as the only attribute for the instances we used for classifier training and evaluation later. We smoothed the data using an averaging running window (length 250 samples) and a hop size of one sample.

The instances used for classifier training and evaluation consisting of a one-dimensional vector, where the normalized gaze deviations represent the only attribute. We define two sets of classes with different class labels. The first class has two labels consisting of *low workload*, referring to trials where no N-back task was used, or *high workload*, thus referring to trials including any N-back task. We investigate binary classification performances using this class. In contrast, the second class contains the four labels *0-back*, *1-back*, *2-back*, and *3-back*, each referring to the measurements of trials using the corresponding N-back task. Overall, we constructed 378 instances, each containing 6000 gaze values, which were used for training and evaluation.

Classifier Performance

An SVM with a linear kernel was used to investigate the prediction performance [34, 56, 63]. The previously described instances were used for training and assigned to their appropriate class label. Within our classifier learning process, we aimed to evaluate the classification efficiency for binary classification, such as detecting low cognitive workload and high cognitive workload. Furthermore, we investigated the classifier performance for determining different levels of cognitive workload with respect to the N-back task difficulty. We used scikit-learn[3] to train different prediction models.

Person-Independent Binary Classification

We performed a person-independent leave-one-person-out classification using two labels; one assigned to low cognitive workload referring to baseline trials and one assigned to high cognitive workload referring to trials comprising a *1*-back, *2*-back, or *3*-back task. A leave-one-person-out classification uses all except one participant for training, while the final participant is used for validating the trained model. The leave-one-person-out classification was carried out for each trajectory and speed separately, where the instances contain the normalized gaze deviations from the shown trajectory per participant and difficulty. The leave-one-person-out classification procedure was repeated for every participant[4] and the results were averaged. Table 3 shows the accuracies, precisions, recalls, and F1 scores of the binary classification per trajectory and speed. The classification results favor fast trajectories in combination with changing directions, such as circular or sinusoidal stimuli.

Person-Independent Multilabel Classification

To investigate the performance of classifying different levels of cognitive workload, the same leave-one-out validation procedure was conducted for four labels, with each label assigned to instances with their respective task difficulty. Again, the validation was conducted for every participant after using the other participants for training. The results are shown in Table 3.

The overall classification efficiency is lower compared to the binary classification accuracy. Different reasons can be responsible for this. First, we are aware that multiple labels may lead to a lower efficiency when the difference between values within the N-back conditions are low. Second, we combine multiple participants who may differ in their smooth pursuit behavior individually. The generalized data could, therefore, be biased by individual gaze behavior, which leads to a distorted result. Therefore, we investigate person-dependent training for cognitive workload classification purposes.

Person-Dependent Multilabel Classification

We analyzed instances within each person to investigate if higher accuracies could be achieved due to person-dependent properties. Instead of a leave-one-person-out classification, we conducted a leave-one-repetition-out classification for each participant, speed, and trajectory. However, the number of folds had to be set differently, since the animations iterated a different number of times depending on the trajectory and speed[5]. Therefore, we adjusted the number of folds of the leave-one-repetition-out classification per trajectory and speed.

The number of folds was set to $k = 2$ for slow rectangles, $k = 3$ for fast rectangles, $k = 5$ for slow circles, $k = 7$ for fast circles, $k = 2$ for slow sine waves, $k = 3$ for fast sine waves. The data per trajectory and speed and per participant was randomly partitioned into k folds, where $k-1$ folds were used for training while the last fold was used for evaluation. This procedure was conducted k times per participant and trajectory with the different N-back difficulties assigned to the

[3] `www.scikit-learn.org` - last access 2018-01-08

[4] Overall 18 runs to use every participant for validation

[5] Each trajectory started and ended at the same position

training set. The results were first averaged per participant and then over all participants. Table 3 shows the results of the cross-validation. Person-dependent classification of multiple cognitive workload levels shows a higher accuracy for fast smooth pursuit movements. Especially fast circular and fast sinusoidal trajectories show better performances compared to rectangular trajectories.

DISCUSSION

Results from our study show how cognitive workload leads to increased gaze differences of smooth pursuit eye movements during the presence of cognitive workload. We found that faster circular and sinusoidal trajectories led to higher gaze deviations, while rectangular trajectories did not show this effect. We believe, that circular and sinusoidal trajectories required more effort from users since they had to focus more on their task completion due to constantly changing directions. This may be a reason for improved classification results for this kind of trajectories. In contrast, rectangular trajectories can be deployed within smooth pursuit-based interfaces whenever reliable input, independent from the perceived cognitive workload, is required. Depending on the use case, a cognition-aware system designer can decide if reliable input through rectangular trajectories or mental workload estimation by circular and sinusoidal trajectories is desired.

The speed of the trajectories had a quantitative measurable effect on the overall smooth pursuit performances. However, as the statistical investigations of single NASA-TLX scales showed, *mental demand* was not affected compared to *physical* and *temporal demand* as well as *effort* by different speeds. Faster speeds cause therefore a different type of physical workload than cognitive workload which significantly impacts smooth pursuit eye movements. Different trajectory speeds are thus not necessarily responsible for inducing cognitive workload. By comparing different *N*-back difficulties with regard to the NASA-TLX scales, we found a significant difference between all difficulties. This supports, that subjectively perceived cognitive workload was altered by different *N*-back complexities and that both variables manipulate measurable smooth pursuit performances.

The classification results yield higher accuracies for distinguishing between low and high workload levels than for detailed levels of cognitive workload. Furthermore, binary classification can be achieved without the need for person-dependent calibration, while separating different levels of cognitive workload requires a person-dependent calibration regarding cognitive workload.

In a real deployment scenario, binary classification can be used to provide additional help for users when high cognitive workload is classified. This refers to very simple scenarios, where only the estimation of low and high workload is desired. Such places could be public places when, for example, interacting with public displays. In contrast to short interactions in the public, a classifier can be trained for multilevel classifications in private spaces where long-term interaction is conceivable.

Calibrating a classification model with multiple workload levels in the public can result in a cumbersome procedure due to external factors impacting the individual cognitive capacities, such as distracting pedestrians walking by. Our results indicate binary workload classification in the public using a pre-trained classification model, while a classification model with multiple workload levels can be used in private spaces where calibration can be done without any disruptions.

To employ sensing of cognitive workload, smooth pursuit must be elicited. In contrast to smooth pursuit-based user interfaces, other environments require explicit or implicit integration of moving elements. For example, as short waiting times occur as a result of a database query or as a new task is loaded during a user study, feedback on the system status could be presented in a way that fosters smooth pursuit movements. For example, this can be elicited by an animated progress bar. In this way, traditional methods such as the NASA-TLX or DALI questionnaires can be complemented.

Finally, a suitable workload level must be found. Looking back at the classification results of the multi-label model, we achieved reasonable accuracies for person-dependent classification. A cognition-aware system must be able to find the right difficulty for each user, as permanent support by a system may lead to boredom or no support leads to frustration for the user. Keeping the task difficulty at its highest or lowest level might not be favored. However, for deploying binary classification in public use cases minor support might be helpful compared to private settings, where user expectations on cognition-aware computing systems are higher.

Limitations

Our study is prone to certain limitations. We collected gaze data under controlled lab conditions and hence, do not know how our results generalize to other situations, where participants may be distracted for example, by the presence of other people. Still, despite the controlled conditions, participants' gaze behavior may have been influenced by physiological well-being, such as lack of sleep. Our calculation of gaze difference is based on 30 seconds of recording. However, there may be situations in which assessing workload with finer granularity is desirable. Additionally, blink frequency and blink duration were not evaluated during the course of our studies. Another limiting factor was the study execution in a calm lab. Natural distractions in real-world environments could alter our results. Furthermore, we have not investigated the effect of different eye tracking frame rates in our study. Consequently, before assessing cognitive workload through smooth pursuit, eye trackers must be tested for their suitability.

USE CASES FOR COGNITION-AWARE SYSTEMS

In the following, we present a number of different use cases to be supported within smooth pursuit scenarios.

Support in Safety-Critical Environments

Smooth pursuit can be utilized to assess cognitive workload during a monitoring task, such as air traffic controller surveying airplane flight processes, to support or warn operators for cognitive exhaustion. Alternatively, workload can be dispatched to a colleague who does not have to cope with high workload during work. Since objects of interest can

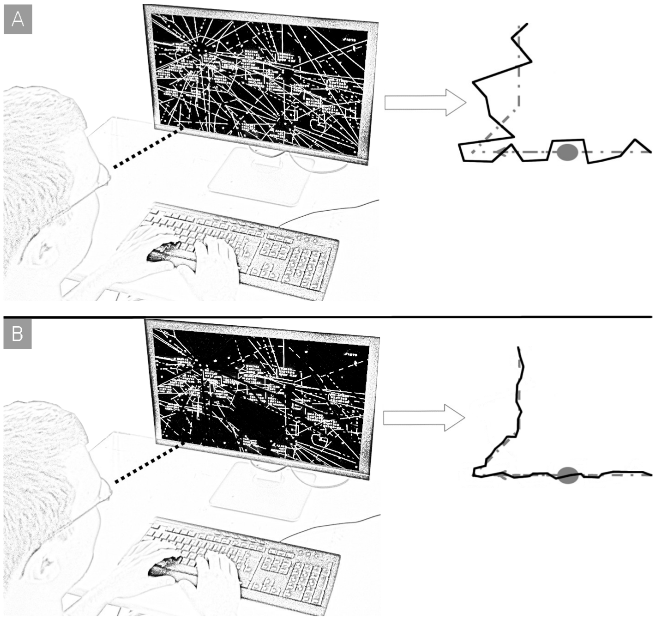

Figure 6. User working in an air control tower. Moving dots representing airplanes on the screen can cause smooth pursuit eye movements. (A) The system detects high cognitive workload from the user and dispatches some observation tasks to a colleague. (B) Alleviated cognitive workload measured after user interface adaption.

be visualized using a small moving circle, smooth pursuit eye movements can be triggered this way. As a result, accidents, which occur due to mental overload, distractions, or fatigue, can be avoided. For example, the user interface can be adapted by simplifying displayed content or dispatching a part of the observation task to another colleague. This use case is transferable to other applications, where moving objects occur naturally and require permanent attention of the user. Figure 6 shows an example of how smooth pursuit can be used in such situations.

Adaptation of Pursuit-based Interactive Systems

Prior research has introduced many applications that use smooth pursuit primarily for interaction. As we illustrated in our related work section, smooth pursuit interfaces have been used for smartwatch interaction [17] and interaction with smartphones [35]. Using mobile devices enables ubiquitous sensing of cognitive workload in outdoor settings, bypassing the disadvantages of using pupillary measures being prone to lighting conditions. Interacting with large distant displays [61], where touch and gesture interaction have emerged as state of the art to communicate input [42, 62] can use smooth pursuit to implicitly measure mental states.

Our approach enables implicit contactless assessment of cognitive workload while interacting with these devices. Pursuit-based interactive systems benefit from our classifier to dynamically predict the current cognitive workload level of the user during interaction and adapt to it accordingly. If high task load is identified during an interaction, the user interface or task objective can be modified by an easier one. Figure 7 shows how the existence of cognitive workload can be estimated to adapt a public display app.

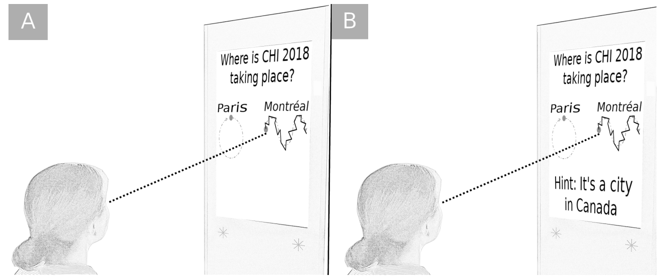

Figure 7. User playing a quiz game on a public display. (A) The system infers that the question is inducing high cognitive workload. The system is, therefore, observing if this behavior persists. (B) The system provides a hint to avoid frustration.

CONCLUSION AND FUTURE WORK

This work investigated the influence of cognitive workload on smooth pursuit eye movements using three different trajectories with two different velocities. Using an auditory delayed digit recall *N*-back task to induce cognitive workload, a higher deviation of gaze points from shown trajectories is measured compared to measurements when not inducing cognitive workload. Based on our results, we create a person-independent classifier for estimating binary workload and a person-dependent classifier for distinguishing different levels of cognitive workload. While binary cognitive workload classification can be elicited in the public using smooth pursuit interfaces, private spaces benefit from person-dependent classifier calibration to determine different levels of cognitive workload.

Having such a measurement modality without the need of body-worn devices goes a step towards real-time mental state estimation in ubiquitous computing environments. User interfaces can then provide intervention mechanisms to relax or help users based on their current context. Our classifier depends on eye gaze only and fits into a number of application scenarios. It can be deployed in real-world scenarios to estimate the presence of cognitive workload in real-time. Thereby, the assessment can be done contactless without the need for additional bodyworn sensors.

In future work, we plan to focus on specific use cases which leverage smooth pursuit as interaction modality to provide an assessment of cognitive workload in real-time. This input will be used to adapt user interfaces of applications accordingly. This includes implementations on public and head-mounted displays using smooth pursuit as input. Furthermore, we want to evaluate the efficiency of assessing cognitive workload unconsciously in user interfaces which naturally display moving elements. This comprises monitoring tasks, which can be found in air traffic and train control system. Consequently, such systems can be benchmarked and optimized regarding their usage complexity. To complement this, further research aiming to correlate objective and subjective workload measures, such as eye movement deviations and NASA-TLX questionnaires, will be conducted. Finally, we will investigate how multiple displayed moving stimuli will affect the classification performance and subjective perception of cognitive workload. To encourage research in this area, we published the data set for further analysis by the research community on our institute's homepage[6].

ACKNOWLEDGEMENTS

This work is supported by the German Federal Ministry of Education and Research as part of the project KoBeLU (Grant No. 16SV7599K). Work on this project was partially funded by the Bavarian State Ministry of Education, Science and the Arts in the framework of the Centre Digitisation, Bavaria (ZD.B). This research was supported by the Deutsche Forschungsgemeinschaft (DFG) (Grant No. AL 1899/2-1).

REFERENCES

1. Yomna Abdelrahman, Eduardo Velloso, Tilman Dingler, Albrecht Schmidt, and Frank Vetere. 2017. Cognitive Heat: Exploring the Usage of Thermal Imaging to Unobtrusively Estimate Cognitive Load. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 33 (Sept. 2017), 20 pages. DOI: http://dx.doi.org/10.1145/3130898
2. Sylvia Ahern and Jackson Beatty. 1979. Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science* 205, 4412 (1979), 1289–1292. DOI: http://dx.doi.org/10.1126/science.472746
3. Ulf Ahlstrom and Ferne J Friedman-Berg. 2006. Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics* 36, 7 (2006), 623–636. DOI: http://dx.doi.org/10.1016/j.ergon.2006.04.002
4. Alan D. Baddeley and Graham Hitch. 1974. Working Memory. Psychology of Learning and Motivation, Vol. 8. Academic Press, 47 – 89. DOI: http://dx.doi.org/https://doi.org/10.1016/S0079-7421(08)60452-1
5. Graham R Barnes. 2008. Cognitive processes involved in smooth pursuit eye movements. *Brain and cognition* 68, 3 (2008), 309–326. DOI: http://dx.doi.org/10.1016/j.bandc.2008.08.020
6. Nikolaus Bee and Elisabeth André. 2008. Writing with your eye: A dwell time free writing system adapted to the nature of human eye gaze. In *International Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems*. Springer, 111–122. DOI: http://dx.doi.org/10.1007/978-3-540-69369-7_13
7. Simone Benedetto, Marco Pedrotti, Luca Minin, Thierry Baccino, Alessandra Re, and Roberto Montanari. 2011. Driver workload and eye blink duration. *Transportation research part F: traffic psychology and behaviour* 14, 3 (2011), 199–208. DOI: http://dx.doi.org/10.1016/j.trf.2010.12.001

[6] www.hcilab.org/your_eyes_tell_data_set - last access 2018-01-08

8. Chris Berka, Daniel J Levendowski, Michelle N Lumicao, Alan Yau, Gene Davis, Vladimir T Zivkovic, Richard E Olmstead, Patrice D Tremoulet, and Patrick L Craven. 2007. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine* 78, Supplement 1 (2007), B231–B244.
9. Anne-Marie Brouwer, Maarten A Hogervorst, Jan BF Van Erp, Tobias Heffelaar, Patrick H Zimmerman, and Robert Oostenveld. 2012. Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering* 9, 4 (2012), 045008.
10. Andreas Bulling and Thorsten O Zander. 2014. Cognition-aware computing. *IEEE Pervasive Computing* 13, 3 (2014), 80–83. DOI: http://dx.doi.org/10.1109/MPRV.2014.42
11. Han Collewijn and Ernst P Tamminga. 1984. Human smooth and saccadic eye movements during voluntary pursuit of different target motions on different backgrounds. *The Journal of physiology* 351 (1984), 217.
12. R Contreras, J Ghajar, S Bahar, and M Suh. 2011. Effect of cognitive load on eye-target synchronization during smooth pursuit eye movement. *Brain research* 1398 (2011), 55–63. DOI: http://dx.doi.org/10.1016/j.brainres.2011.05.004
13. Alexander De Luca, Roman Weiss, and Heiko Drewes. 2007. Evaluation of eye-gaze interaction methods for security enhanced PIN-entry. In *Proceedings of the 19th australasian conference on computer-human interaction: Entertaining user interfaces*. ACM, 199–202. DOI: http://dx.doi.org/10.1145/1324892.1324932
14. Tilman Dingler. 2016. Cognition-aware Systems As Mobile Personal Assistants. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA, 1035–1040. DOI: http://dx.doi.org/10.1145/2968219.2968565
15. Andrew Duchowski. 2017. *Eye tracking methodology: Theory and practice*. Vol. 373. Springer Science & Business Media.
16. Andrew T Duchowski. 2002. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers* 34, 4 (2002), 455–470. DOI: http://dx.doi.org/10.3758/BF03195475
17. Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: gaze interaction for smart watches using smooth pursuit eye movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 457–466. DOI: http://dx.doi.org/10.1145/2807442.2807499
18. DMA Gronwall. 1977. Paced auditory serial-addition task: a measure of recovery from concussion. *Perceptual and motor skills* 44, 2 (1977), 367–373. DOI: http://dx.doi.org/10.2466/pms.1977.44.2.367

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

20. Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 52 (1988), 139–183.

21. Eckhard H Hess and James M Polt. 1964. Pupil size in relation to mental activity during simple problem-solving. *Science* 143, 3611 (1964), 1190–1192. DOI: http://dx.doi.org/10.1126/science.143.3611.1190

22. Thomas E Hutchinson, K Preston White, Worthy N Martin, Kelly C Reichert, and Lisa A Frey. 1989. Human-computer interaction using eye-gaze input. *IEEE Transactions on systems, man, and cybernetics* 19, 6 (1989), 1527–1534. DOI: http://dx.doi.org/10.1109/21.44068

23. Poika Isokoski. 2000. Text input methods for eye trackers using off-screen targets. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 15–21. DOI: http://dx.doi.org/10.1145/355017.355020

24. RJ Jacob and Keith S Karn. 2003. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind* 2, 3 (2003), 4.

25. Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-based Interaction Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90)*. ACM, New York, NY, USA, 11–18. DOI: http://dx.doi.org/10.1145/97243.97246

26. Marcel Adam Just and Patricia A Carpenter. 1976. Eye fixations and cognitive processes. *Cognitive Psychology* 8, 4 (1976), 441 – 480. DOI: http://dx.doi.org/10.1016/0010-0285(76)90015-3

27. Michael J Kane, Andrew RA Conway, Timothy K Miura, and Gregory JH Colflesh. 2007. Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33, 3 (2007), 615.

28. Mohamed Khamis, Florian Alt, and Andreas Bulling. 2015. A field study on spontaneous gaze-based interaction with a public display using pursuits. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM, 863–872. DOI: http://dx.doi.org/10.1145/2800835.2804335

29. Mohamed Khamis, Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, and Florian Alt. 2016. TextPursuits: Using Text for Pursuits-Based Interaction and Calibration with Public Displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM. DOI: http://dx.doi.org/10.1145/2971648.2971679

30. Jeff Klingner, Rakshit Kumar, and Pat Hanrahan. 2008. Measuring the task-evoked pupillary response with a remote eye tracker. In *Proceedings of the 2008 symposium on Eye tracking research & applications*. ACM, 69–72. DOI: http://dx.doi.org/10.1145/1344471.1344489

31. J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek. 2014. Non-contact heart rate and heart rate variability measurements: A review. *Biomedical Signal Processing and Control* 13, Supplement C (2014), 102 – 112. DOI: http://dx.doi.org/https://doi.org/10.1016/j.bspc.2014.03.004

32. Jan-Louis Kruger, Esté Hefer, and Gordon Matthew. 2013. Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts. In *Proceedings of the 2013 Conference on Eye Tracking South Africa*. ACM, 62–66. DOI: http://dx.doi.org/10.1145/2509315.2509331

33. R John Leigh and David S Zee. 2015. *The neurology of eye movements*. Vol. 90. Oxford University Press, USA.

34. Yulan Liang and John D Lee. 2008. Driver cognitive distraction detection using eye movements. In *Passive Eye Monitoring*. Springer, 285–300. DOI: http://dx.doi.org/10.1007/978-3-540-75412-1_13

35. Dachuan Liu, Bo Dong, Xing Gao, and Haining Wang. 2015. *Exploiting Eye Tracking for Smartphone Authentication*. Springer International Publishing, Cham, 457–477. DOI: http://dx.doi.org/10.1007/978-3-319-28166-7_22

36. Dillon James Lohr and Oleg V. Komogortsev. 2017. A Comparison of Smooth Pursuit- and Dwell-based Selection at Multiple Levels of Spatial Accuracy. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2760–2766. DOI: http://dx.doi.org/10.1145/3027063.3053233

37. Päivi Majaranta and Kari-Jouko Räihä. 2002. Twenty years of eye typing: systems and design issues. In *Proceedings of the 2002 symposium on Eye tracking research & applications*. ACM, 15–22. DOI: http://dx.doi.org/10.1145/507072.507076

38. Stefan Mattes and Anders Hallén. 2009. Surrogate distraction measurement techniques: The lane change test. *Driver distraction: Theory, effects, and mitigation* (2009), 107–121.

39. Bruce Mehler, Bryan Reimer, and JA Dusek. 2011. MIT AgeLab delayed digit recall task (n-back). *Cambridge, MA: Massachusetts Institute of Technology* (2011).

40. Takehiko Ohno. 1998. Features of eye gaze interface for selection tasks. In *Computer Human Interaction, 1998. Proceedings. 3rd Asia Pacific*. IEEE, 176–181. DOI: http://dx.doi.org/10.1109/APCHI.1998.704190

41. Annie Pauzié. 2008. A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems* 2, 4 (2008), 315–322.

42. Peter Peltonen, Esko Kurvinen, Antti Salovaara, Giulio Jacucci, Tommi Ilmonen, John Evans, Antti Oulasvirta, and Petri Saarikko. 2008. It's Mine, Don'T Touch!: Interactions at a Large Multi-touch Display in a City Centre. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 1285–1294. DOI: http://dx.doi.org/10.1145/1357054.1357255

43. Vsevolod Peysakhovich. 2016. *Study of pupil diameter and eye movements to enhance flight safety. Etude de diamètre pupillaire et de mouvements oculaires pour la sécurité aérienne*. Ph.D. Dissertation. Université de Toulouse, Université Toulouse III-Paul Sabatier.

44. Ken Pfeuffer, Melodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. 2013. Pursuit calibration: Making gaze calibration less tedious and more flexible. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, 261–270. DOI: http://dx.doi.org/10.1145/2501988.2501998

45. Bastian Pfleging, Drea K Fekety, Albrecht Schmidt, and Andrew L Kun. 2016. A Model Relating Pupil Diameter to Mental Workload and Lighting Conditions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5776–5788. DOI: http://dx.doi.org/10.1145/2858036.2858117

46. Alex Poole and Linden J Ball. 2006. Eye tracking in HCI and usability research. *Encyclopedia of human computer interaction* 1 (2006), 211–219.

47. Dale Purves, George J Augustine, David Fitzpatrick, Lawrence C Katz, Anthony-Samuel LaMantia, James O McNamara, and S Mark Williams. 2001. Types of eye movements and their functions. (2001).

48. Carlos H Schenck, Scott R Bundlie, Andrea L Patterson, and Mark W Mahowald. 1987. Rapid eye movement sleep behavior disorder: a treatable parasomnia affecting older adults. *Jama* 257, 13 (1987), 1786–1789. DOI: http://dx.doi.org/10.1001/jama.1987.03390130104038

49. Simon Schenk, Marc Dreiser, Gerhard Rigoll, and Michael Dorr. 2017. GazeEverywhere: Enabling Gaze-only User Interaction on an Unmodified Desktop PC in Everyday Scenarios. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3034–3044. DOI: http://dx.doi.org/10.1145/3025453.3025455

50. Linda E Sibert and Robert JK Jacob. 2000. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 281–288.

51. Eva Siegenthaler, Francisco M Costela, Michael B McCamy, Leandro L Di Stasi, Jorge Otero-Millan, Andreas Sonderegger, Rudolf Groner, Stephen Macknik, and Susana Martinez-Conde. 2014. Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *European Journal of Neuroscience* 39, 2 (2014), 287–294. DOI: http://dx.doi.org/10.1111/ejn.12395

52. David L Sparks. 2002. The brainstem control of saccadic eye movements. *Nature Reviews Neuroscience* 3, 12 (2002), 952–964. DOI: http://dx.doi.org/10.1038/nrn986

53. John Stemberger, Robert S Allison, and Thomas Schnell. 2010. Thermal imaging as a way to classify cognitive workload. In *Computer and Robot Vision (CRV), 2010 Canadian Conference on*. IEEE, 231–238. DOI: http://dx.doi.org/10.1109/CRV.2010.37

54. Els Stuyven, Koen Van der Goten, André Vandierendonck, Kristl Claeys, and Luc Crevits. 2000. The effect of cognitive load on saccadic eye movements. *Acta psychologica* 104, 1 (2000), 69–85. DOI: http://dx.doi.org/10.1016/S0001-6918(99)00054-2

55. Yi-Fang Tsai, Erik Viirre, Christopher Strychacz, Bradley Chase, and Tzyy-Ping Jung. 2007. Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviation, space, and environmental medicine* 78, Supplement 1 (2007), B176–B185.

56. Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun. 2004. Support vector machine learning for interdependent and structured output spaces. In *Proceedings of the twenty-first international conference on Machine learning*. ACM, 104. DOI: http://dx.doi.org/10.1145/1015330.1015341

57. Outi Tuisku, Päivi Majaranta, Poika Isokoski, and Kari-Jouko Räihä. 2008. Now Dasher! Dash away!: longitudinal study of fast text entry by Eye Gaze. In *Proceedings of the 2008 symposium on Eye tracking research & applications*. ACM, 19–26. DOI: http://dx.doi.org/10.1145/1344471.1344476

58. Marilyn L Turner and Randall W Engle. 1989. Is working memory capacity task dependent? *Journal of memory and language* 28, 2 (1989), 127–154.

59. Eduardo Velloso, Markus Wirth, Christian Weichel, Augusto Esteves, and Hans Gellersen. 2016. AmbiGaze: Direct Control of Ambient Devices by Gaze. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. ACM, 812–817. DOI: http://dx.doi.org/10.1145/2901790.2901867

60. Trent W. Victor, Joanne L. Harbluk, and Johan A. Engström. 2005. Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour* 8, 2 (2005), 167 – 190. DOI: http://dx.doi.org/https://doi.org/10.1016/j.trf.2005.04.014 The relationship between distraction and driving performance: towards a test regime for in-vehicle information systemsIn-vehicle information systems.

61. Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 439–448. DOI: http://dx.doi.org/10.1145/2493432.2493477

62. Daniel Vogel and Ravin Balakrishnan. 2004. Interactive Public Ambient Displays: Transitioning from Implicit to Explicit, Public to Personal, Interaction with Multiple Users. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. ACM, New York, NY, USA, 137–146. DOI: http://dx.doi.org/10.1145/1029632.1029656

63. Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. 2016. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

64. Johannes Zagermann, Ulrike Pfeil, Harald Reiterer, Yunlong Wang, Ulrike Pfeil, Harald Reiterer, Johannes Zagermann, Ulrike Pfeil, Roman Rädle, Hans-Christian Jetter, and others. 2015. Measuring Cognitive Load using Eye Tracking Technology in Visual Computing. *Proceedings of BELIV'16* (2015), 259–260. DOI: http://dx.doi.org/10.1145/2993901.2993908

65. Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2013. SideWays: a gaze interface for spontaneous interaction with situated displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 851–860. DOI: http://dx.doi.org/10.1145/2470654.2470775

66. Josef Zihl, D Von Cramon, and Norbert Mai. 1983. Selective disturbance of movement vision after bilateral brain damage. *Brain* 106, 2 (1983), 313–340. DOI: http://dx.doi.org/10.1093/brain/106.2.313