Pursuing Smooth Pursuits: Challenges and Remedies for Eye-Movement Event Classification

Luke Korthals

Master Candidate, Research Master Psychology, University of Amsterdam

luke.korthals@stud.uva.nl

Abstract

Human experts and classification algorithms often confuse fixations (fixating stationary targets) and smooth pursuits (fixating moving targets) because their feature characteristics overlap. To investigate the gazeHMM algorithm by Lüken et al. (2022) and to explore better features, I created a ground truth data set that does not rely on human annotation. The data set consists of almost four hours of eye movements. Ten participants fixated different targets designed to induce saccades, fixations, and smooth pursuits. Ground truth was established by avoiding fixations and smooth pursuits to cooccur and separating them from saccades by their velocity. Visual inspection revealed that gazeHMM confused fixations and smooth pursuits likely because their velocity, acceleration, and sample-to-sample angle were distributed similarly. In contrast, I developed two features based on findings that directions within smooth pursuits are more similar than within fixations. The *estimated direction deviation* and *estimated direction deviation spread* clearly distinguished between fixations and smooth pursuits in the current data set. Therefore, they could likely improve the automatic classification of these eye movements, which is particularly relevant for the study of cognition and neurodegenerative diseases.

Keywords: eye tracking, eye movements, event classification, event detection, fixations, smooth pursuits, eye movement features, feature engineering

Pursuing Smooth Pursuits: Challenges and Remedies for Eye-Movement Event Classification

Eye movement research relies on accurate classifications of saccades, fixations, smooth pursuits, and other eye movements. Automatic event classification algorithms like gazeHMM (Lüken et al., 2022) can't reliably distinguish fixations from smooth pursuits. Similarly, human annotators don't consistently agree on their classification (Andersson et al., 2017) and it is unclear whether expert annotations constitute a gold-standard (Andersson et al., 2017; Hooge et al., 2018). The current study aimed to address these limitations in several ways. First, by creating a benchmark data set without using human annotations. Second, by investigating how eye movement features used by gazeHMM are distributed for fixations and smooth pursuits. Third, by developing eye movement features that better differentiate fixations and smooth pursuits.

Eye Tracking Research

Eye tracking is used to study many different phenomena. Examples include user experience and human machine interaction (Khamis et al., 2018; Poole & Ball, 2006), driver safety and assisted driving (Braunagel et al., 2015; Palinko et al., 2010), visual perception and attention (Lappi, 2016), affective disorders and neurodegenerative diseases (Armstrong & Olatunji, 2012; MacAskill & Anderson, 2016), and decision making or other cognitive processes (Kucharský et al., 2020; Schulte-Mecklenbeck et al., 2017). Often research relies on accurately classifying eye movements like fixations, saccades, smooth pursuits, post-saccadic oscillations (PSO), the vestibulo-ocular reflex (VOR), the optokinetic response (OKR) or fixational micromovements (Lappi, 2016). Fixations stabilize the gaze on stationary targets; saccades are fast and jerky eye movements between fixations; smooth pursuits stabilize the gaze on moving targets using smooth eye movements of varying speeds (Lappi, 2016). PSO include overshoot, undershoot, and sudden directional or velocity changes after saccades (Larsson et al., 2013). VOR and OKR stabilize the gaze on a target by compensating for head movements and retinal image slip respectively (Lappi, 2016). Finally, fixational micromovements prevent neural adaptation through drift, micro saccades, and tremors (Martinez-Conde et al., 2004). For many applications classifying fixations, saccades, and smooth pursuits is most important (Duchowski, 2017). They guide visual attention and show what observers are interested in (Duchowski, 2017), and can even reveal cognitive strategies (Kucharský et al., 2020).

Eye Movement Classification

Manual classification is lengthy and difficult even for experts (Startsev & Zemblys, 2022) and algorithms speed up classification considerably (Startsev & Zemblys, 2022). Further, if algorithms rely on unsupervised methods and few parameters, they reduce human bias (Lüken et al., 2022). Arguably many algorithms accurately classify fixations and saccades as binary events (Andersson et al., 2017). Fewer algorithms include labels for smooth pursuits or other movements; and those that do often confuse smooth pursuits with fixations or other eye movements (Komogortsev & Karpov, 2013). This lack of accurate automatic smooth pursuit classification is a serious shortcoming because among other applications smooth pursuits are used to investigate schizophrenia (O'Driscoll & Callahan, 2008), traumatic brain injury (Hunfalvay et al., 2020), and neurodegenerative diseases (MacAskill & Anderson, 2016; Tao et al., 2020). Therefore, improving automatic classification of smooth pursuits is highly relevant and the motivation behind the current study.

Feature Engineering for Eye Movement Classification

Feature engineering means preparing and transforming raw data such that algorithms can use it to fulfil their respective tasks. Features can be based on domain knowledge, intuition, or systematic analysis (Verdonck et al., 2021). To classify eye movements, features are extracted from raw gaze which is a timeseries of x-y-coordinates (Startsev & Zemblys, 2022). For example, the Identification by Velocity Threshold (I-VT) algorithm classifies samples with velocities exceeding some threshold as saccades, and others as fixations (Salvucci & Goldberg, 2000). Rigas et al. (2018) analyzed features of fixations, saccades, and PSO. However, to the best of my knowledge no similar analysis of features for smooth pursuits exists. Additionally, classifying smooth pursuits is more difficult than separating fixations and saccades because their feature characteristics overlap with those of other eye movements (Agtzidis et al., 2016; Andersson et al., 2017; Dorr et al., 2010; Komogortsev & Karpov, 2013; Startsev et al., 2019, 2016; Vidal et al., 2011). Most commonly, slow smooth pursuits overlap with fixations because of fixational micromovements and measurement inaccuracies (Agtzidis et al., 2016; Dorr et al., 2010; Komogortsev & Karpov, 2013). Even expert annotators find their separation difficult: "Tricky distinguishing slow pursuit and fixations with drift" (Andersson et al., 2017)[1]. Additionally, fast smooth pursuits up to 90°/s – 100°/s (Meyer et al., 1985) or even 115°/s – 150°/s (Lisberger et al., 1981) can overlap with saccades that exhibit minimum velocities between 10°/s and 125°/s (Lappi, 2016). Finally, under free head movement, smooth pursuits overlap with VOR (Vidal et al., 2011). Eye-head coordination makes classifying eye movements more difficult in general (Agtzidis et al., 2019) and the following overview of classification algorithms illustrates that smooth pursuits remain difficult to classify even without considering head movements. Therefore, I focus solely on fixed-head eye tracking in the current study.

Existing Algorithms for Smooth Pursuit Classification

Berg et al. (2009) combined velocity-thresholds with principle component analysis (I-PCA). Dorr et al. (2010) used velocity- and dispersion-thresholds. Both classified fixations, saccades, and smooth pursuits but didn't evaluate the comparative performance of their algorithms.

Vidal et al. (2012) used seven features describing the shape of gaze trajectories (slope, range, mean velocity, variance, integral, energy, waveform length) and used k-nearest neighbor classification to

[1] This is a quote of annotator comments that can be found in the raw data set.

classify smooth pursuits and “other”. Their algorithm achieved relatively high accuracy, but they didn’t evaluate its comparative performance.

Komogortsev and Karpov (2013) used velocity, angle, and dispersion features and three threshold-based algorithms (I-VMP, I-VDT, I-VVT) to classify fixations, saccades, and smooth pursuits. Even their best algorithm (I-VDT) commonly misclassified fixations as smooth pursuits.

Larsson et al. (2015) used velocity, dispersion, directional consistency, positional displacement, and spatial range as threshold-based criteria. They used criteria matching to classify smooth pursuits, fixations, and “unclear events” between saccades. Their algorithm outperformed I-PCA (Berg et al., 2009) and I-VDT (Komogortsev & Karpov, 2013) on data by Larsson et al. (2013) but achieved relatively low overall agreement with expert annotations.

Agtzidis et al. (2016) used clustered gaze paths from multiple observers exposed to the same stimuli to classify smooth pursuits. Their approach outperformed I-PCA (Berg et al., 2009) and Larsson et al. (2015) but can only be used if data of multiple observers exists.

Santini et al. (2016) applied a probabilistic Bayesian approach (I-BDT) on velocity and movement ratio features to classify fixations, saccades, and smooth pursuits. Their algorithm outperformed I-VDT (Komogortsev & Karpov, 2013) on their own data but sometimes misclassified fast smooth pursuits as saccades and slow smooth pursuits as fixations.

Pekkanen and Lappi (2017) used naive segmented linear regression to filter raw gaze positions and a hidden Markov model (NSLR-HMM) to classify fixations, saccades, smooth pursuits, and PSO. As features they used velocity, and the cosine of segment-to-segment angles. Their algorithm outperformed ten algorithms investigated by Andersson et al. (2017) on saccades and fixations, but achieved relatively low agreement with expert annotations for smooth pursuits and fixations.

Startsev et al. (2019) used deep learning with a recurrent neural network relying on 1D convolution and bidirectional-long-short-term-memory (1D CNN-BLSTM) to classify fixations, saccades, and smooth pursuits. As input features they used relative x-y-coordinates, velocity, acceleration, and gaze direction. They trained their model on GazeCom (Dorr et al., 2010; Startsev et al., 2016) and it generalized reasonably well to Andersson et al. (2017). However, it was outperformed for smooth pursuits by I-VMP (Komogortsev & Karpov, 2013) on the sample level, and by Dorr et al. (2010) and Agtzidis et al. (2016) on the event level. Sample-level evaluation refers to comparing classification labels of each individual gaze sample to their ground truth labels, while event-level evaluation concerns "[…] uninterrupted sequences of samples with the same label" (Startsev & Zemblys, 2022, p. 16).

Dar et al. (2020) used a threshold approach with adaptive velocity thresholds (REMoDNaV). REMoDNaV achieved high recall but low precision for smooth pursuits on GazeCom data (Startsev et al., 2016). Recall and precision are binary performance measures. Recall describes the proportion of true positive classifications divided by all true and false positive classifications. Precision describes the proportion of true positive classifications divided by true positive and false negative classifications (Startsev & Zemblys, 2022).

Lüken et al. (2022) used a hidden Markov model (gazeHMM) to classify fixations, saccades, smooth pursuits, and PSO. As features they used velocity, acceleration, and sample-to-sample angle. They achieved good overall agreement with expert annotations on Andersson et al. (2017) data. However, gazeHMM confused fixations with smooth pursuits and rapidly switched between them.

Goals of the Current Study

The overview above shows that eye movements are difficult to classify when feature characteristics overlap. This makes the classification of smooth pursuits primarily a feature engineering problem. Rigas et al. (2018) provide a comprehensive overview of features for fixations, saccades, and PSO but neglect smooth pursuits. Therefore, and because features of current algorithms proved insufficient, the primary goal of this study was to investigate feature characteristics of smooth pursuits to improve their automatic classification.

Rigas et al. (2018) based their feature analysis on naturalistic stimuli (i.e., text reading) which promise greater ecological validity than synthetic stimuli like moving dots (Andersson et al., 2017). However, I think that synthetic stimuli specifically crafted to induce certain eye movements are more useful for the analysis of smooth pursuit feature characteristics. This is because eye movements in response to videos or other naturalistic stimuli require expert annotations to establish ground truth (Startsev & Zemblys, 2022). However, experts agree much less on the classification of smooth pursuits than on that of saccades (Andersson et al., 2017; Lüken et al., 2022) and their annotations can be biased (Hooge et al., 2018). Forgoing expert annotations would also encourage researchers to develop bottom-up, data driven approaches with reduced human influence that have recently been advocated for (Hein & Zangemeister, 2017; Lüken et al., 2022). Therefore, the second goal of this study was to create a benchmark data set of saccades, fixations, and smooth pursuits without relying on human annotation by designing stimuli that would induce either fixations <u>or</u> smooth pursuits.

While such synthetic stimuli likely induce eye movements that are less naturalistic, their range, speed, and trajectory can be controlled directly. Additionally, they will induce very clear eye movements as along as participants focus targets diligently. As existing algorithms commonly confuse slow smooth pursuits with fixations, systematically varying target velocities and trajectories will help to identify specific conditions under which feature characteristics overlap and misclassifications occur.

To clarify, I agree with (Startsev & Zemblys, 2022) classification algorithms should be evaluated at least partly based on natural eye movements and expert annotations to compare them with existing solutions. Nevertheless, because feature characteristics of smooth pursuits overlap with those of other eye movements (Agtzidis et al., 2016; Dorr et al., 2010; Komogortsev & Karpov, 2013) and expert annotations can’t be blindly trusted (Hooge et al., 2018), I think the current approach will make it easier to develop new features to classify smooth pursuits.

gazeHMM

gazeHMM classifies eye movements according to their velocity, acceleration, and sample-to-sample angle based on the following assumptions: First, saccades have high velocity, high acceleration, and are directed movements with sample-to-sample angles distributed around zero degrees. Second, Smooth pursuits have moderate velocity, low acceleration, and are also directed with sample-to-sample angles distributed around zero degrees. Third, Fixations have low velocity, low acceleration, and follow a random walk with uniformly distributed sample-to-sample angles (Lüken et al., 2022). According to these assumptions, the velocity and acceleration should distinguish saccades from both fixations and smooth pursuits, while the sample-to-sample angle should distinguish fixations from both saccades and smooth pursuits. Instead, gazeHMM confused fixations and smooth pursuits and rapidly switched between them on the Andersson et al. (2017) data set (Lüken et al., 2022).

A Priori Expectations

While I did not devise formalized hypotheses, I had three broad expectations based on previous findings outlined above. First, I expected gazeHMM to confuse fixations and smooth pursuits and rapidly switch between them. Second, I expected the velocity, acceleration, and sample-to-sample angle distributions to overlap for fixations and smooth pursuits. Third, I expected that these feature distributions would be particularly similar for fixations and slow smooth pursuits.

Structure of this Report

The remainder of the report is organized into four parts. Some have additional subsections, and all have individual methods, results, and discussion sections.

Benchmark data set. The first section concerns the data collection for the benchmark data set and how I created ground truth labels without relying on human annotations.

gazeHMM evaluation. The second section concerns the evaluation of gazeHMM. This section is further divided into two subsections. In the first subsection, I compared gazeHMM classifications to the ground truth of the benchmark data set to investigate whether it would fail to distinguish fixations and smooth pursuits. In the second subsection I explored the velocity, acceleration, and sample-to-sample angle features to investigated if their distributions would overlap for fixations and smooth pursuits.

Feature engineering. The third section concerns my exploratory attempt to develop new features that distinguish between fixations and smooth pursuits. This section is further divided into three subsections. In the first subsection I discuss how the deviation of sample directions from overall event directions could potentially distinguish between fixations and smooth pursuits based on a finding by Startsev et al. (2019). Additionally, I investigated the direction deviation in the current benchmark data set. In the second and third subsections I developed and investigated the distributions of two new features. Specifically, they concern the *estimated direction deviation* (EDD) and *estimated direction deviation spread* (EDD-S) features and their distributions for fixations and smooth pursuits respectively.

General discussion. I end the report with a general discussion to tie everything together and discuss future research directions.

Benchmark Data Set

Method

Sample

Twelve fellow research master students from the University of Amsterdam participated as a favor to the researcher and were not compensated. Two participants were excluded before data collection because they could not be calibrated. Eye movements of the remaining ten participants were collected and randomly assigned to training (n = 7) and test sets (n = 3). Participants were between 22 and 27 (M = 24.2) years old. Most identified as female (f = 7, m = 2, other = 1), and had brown eyes (brown = 7, blue = 3). Most were nearsighted but not all used correction during data collection (nearsighted + glasses = 1, nearsighted + contacts = 3, nearsighted + no correction = 4, no eye condition = 2).

Technical Setup

Participants sat 70cm from a widescreen monitor (61.4csm x 36.5cm; 2560px x 1440px) with 144hz refresh rate. Eye movements were collected at 1000hz by an EyeLink 1000 Plus (SR Research, n.d.)[2] that was mounted to the table below the monitor. Participants stabilized their head in a chin rest with their eyes at the center of the screen. I programmed the study in Python (Van Rossum & Drake Jr, 1995) using the PsychoPy package (Peirce et al., 2019).

Stimuli

During each trial a small grey circle (#808080; r = 0.2°) was presented on a black background (#000000). Participants were instructed to focus on these circles without moving their head. Circles moved 20° across the screen. They moved according to three different movement patterns (i.e., *moving circles*, *jumping circles*, *back-and-forth circles*), at three different speeds (i.e., 1°/s, 3°/s, 6°/s), and along eight different straight trajectories (i.e., horizontal- left/right, vertical- up/down, diagonal- left/right + up/down).

Movement Distance. The range of normal eye movements differs between movement directions and declines with age. It lies between 20° and 30° for upward movements (Lee et al., 2019). I tested different ranges myself and decided to use 20° because it covered a wide visual range without being uncomfortable.

Movement patterns. To create ground truth labels without relying on human annotation, I wanted to restrict each trial to include saccades and fixations or smooth pursuits. Saccades are fast eye movements, while fixations and smooth pursuits are slow eye movements (Lappi, 2016) and I wanted to classify them according to velocity thresholds for each trial. I tested different movement patterns and included three that seemed to prevent fixations and smooth pursuits to cooccur curing testing.

[2] The eye tracker was set to 1000hz and used monocular tracking of the right pupil with ellipse fitting and no drift correction.

Moving circles moved consistently at the target speed to induce smooth pursuits. *Jumping circles* jumped along fixation locations to induce series of fixations and saccades. Targets stood still for 1000ms at each fixation location which were evenly spaced according to the target speed. *Back-and-forth circles* moved consistently at the target speed and jumped back and forth along their trajectory at 1000ms intervals to induce series of smooth pursuits and saccades.

Target speed. To investigate my expectation that slow smooth pursuits are especially difficult to classify, I wanted to include very slow and very fast smooth pursuits. Importantly, smooth pursuits at all speeds had to be clear without intermittent fixations or catchup saccades. Additionally, because targets in *jumping circle* and *back-and-forth circle* jumped at 1000ms intervals, the shortest trials had to be at least one second long to include a single saccade. Maximum smooth pursuit speeds without catch up saccades are reported to be around 30°/s (Lappi, 2016). At this speed targets would move the distance of 20° in less than one second which would be too fast. To my knowledge no minimum smooth pursuit speed is reported in the literature. With these constraints in mind, I tested different speeds for *moving circle* trials and selected three speeds of 1°/s, 3°/s, and 6°/s. During testing I found it difficult to smoothly track targets slower than 1°/s and faster than 6°/s. This is consistent with findings by Drewes et al. (2018) who found that smooth pursuits of targets moving slower than 1°/s included fixations and targets moving faster than 6°/s included saccades. At these speeds trials took 20, 6.66, and 3.33 seconds respectively.

Target Trajectory. Eye movements in different directions are controlled by three pairs of muscles. To cover a wide range of movements, targets moved along horizontal, vertical, and 45°-diagonal trajectories which are controlled by individual or multiple pairs of muscles (Pinkhardt & Kassubek, 2011). To ensure that targets were easy to follow I avoided circular trajectories because plots reported by Drewes et al. (2018) indicate that they are more difficult to follow accurately.

Trials

Each participant saw each speed-trajectory-movement combination twice for a total of 144 trials (i.e., movement [3] x speed [3] x trajectory [8] x 2). Trials took 20s, 6.66s, and 3.33s for targets moving 1°/s, 3°/s, and 6°/s. Therefore, about 24 minutes of eye movements were collected for each participant and almost 4 hours were collected in total. Note that slow eye movements are overrepresented in the current data set because their trials took longer. I did not balance this by including additional trials at fast speeds because increasing the total duration could have been uncomfortable for participants and could have reduced data quality if they lost concentration. Additionally, I expected slow smooth pursuits to be particularly difficult to distinguish from fixations. This should also make it more difficult to uncover features that are distinct for fixations and smooth pursuits which would make any such features even more promising. Trials were organized in two blocks. Within each block trials were ordered from slow to fast speeds while movement pattern and trajectory were randomized. I did not randomize speed because during testing I found it difficult to readjust to slow targets after fast targets without performing unintended saccades.

Procedure

Upon entering the lab, I informed participants about the procedure, and they signed participation forms. Participants sat in front of the screen, rested their head in a chin rest, and calibrated the eye tracker using nine-point calibration by focusing on circles and pressing the “enter” key on a keyboard. Calibration was repeated until it was satisfactory or good. Two participants were excluded because I was unable to calibrate them. After calibration I instructed participants to focus all targets without moving their head. Participants completed a tutorial trial (*moving circle*, 2°/s, horizontal-right) and could ask questions. Afterwards, participants started each trial by pressing the space bar and could take as many breaks as they wanted. Before the first trial, I emphasized the importance of collecting clear data and encouraged them to take frequent breaks. To keep participants focused, I told them they were doing well and encouraged them to keep it up after the first trial and intermittently throughout data collection. Additionally, to ensure the quality of the collected eye movements I constantly monitored if the eye tracker was working and if participants were accurately following the targets. If a participant blinked frequently or lost track of the target, I reminded them to focus and take breaks. If participants removed their head from the chin rest or the eye tracker lost their pupil, calibration was repeated before the next trial. One participant was recalibrated because they removed their head during a break. Two participants were recalibrated because the eye tracker lost their pupil. After the last trial, I thanked and debriefed participants. Each session took between 45 and 60 minutes depending on time spent on breaks and calibration.

Preprocessing

During preprocessing the EyeLink data was combined with information about the targets collected in Psychopy (Peirce et al., 2019). For brevity I do not report these steps here, but the preprocessing protocol is fully described on Github (https://github.com/lukekorthals/pursuing-smooth-pursuits).

Additionally, preprocessing included important operations like classifying the ground truth baseline. These steps are reported below in the order of their execution. Note that I devised and finetuned the preprocessing protocol after data collection using two arbitrarily selected participants from the training set (7bb2338 and 6cde27b5). As fixations and smooth pursuits were classified indirectly by labeling slow eye movements according to the target movement I only considered saccades during finetuning.

Correcting target positions for back-and-forth circle trials. Due to a bug, the target locations collected during back-and-forth circle trials were shifted along their jumping positions (e.g., target at position A but position B collected). Consequently, the correct target positions were calculated according to trial time, target speed, and trajectory. During analysis I also found that initial target updates were skipped for 06b8d2d3. Therefore, the initial and subsequent target position of *back-and-*

forth trials for this participant were different from the other participants (e.g., started at position B instead of A). As the collected eye movements remain the same, this should not affect my analysis and I included their data in the analysis.

Removing blinks. Blinks are characterized by scattered, fast samples when the eye tracker loses the pupil (Hershman et al., 2018). These artifacts would affect the velocity threshold to distinguish saccades from fixations or smooth pursuits because of their high speed. Therefore, all samples the EyeLink classified as blinks were removed (set to NA). Additionally, all samples 50ms before and 70ms after these blinks were also removed. I created this asymmetrical window according to plotted data for participants 7bb2338 and 6cde27b5. Specifically, I started with a symmetrical 50ms window before and after each blink to match the approach by (Lüken et al., 2022). However, many artefacts after blinks remained and I gradually increased the tail of the window to 70ms. Some artefacts remained but I did not increase the window size further to prevent losing saccades, fixations, and smooth pursuits in trials where smaller windows sufficed to remove blink artefacts.

Interpolating target positions. The EyeLink collected data at 1000hz, but the monitor updated target positions at 144hz. Therefore, missing target positions were filled with previous target positions in the combined data set. Note that eye and target positions are not matched perfectly because I forgot to consider that the EyeLink collected an additional 100ms before and after each trial. This does not affect the analysis of eye movement features but means that the reported gaze paths are inaccurate. Specifically, true target positions would be shifted 100ms to the right and the first and last 100ms of eye movements should be removed.

Removing unrealistic velocities. According to (Lappi, 2016), saccades show peak velocities of up to 800°/s and (Lüken et al., 2022) removed all samples faster than 1300°/s. As I am investigating gazeHMM in the current study, I followed their approach and removed samples with velocities greater than 1300°/s. I expect that these samples represent unidentified blinks or eye tracking artifacts that could negatively affect the accuracy of the velocity threshold to classify the baseline. Note that Lüken et al. (2020) calculated the velocity as the first derivative of a Savitzky-Golay filter (see Schafer, 2011). In contrast I calculated the velocity as the Euclidean distance between consecutive samples divided by time because this is the simplest way to calculate raw velocities.

Classifying ground truth eye movements. This is the most important step of preprocessing, as it determines whether the dataset can be considered a good benchmark without relying on human coding. As indicated above, I designed movement patterns such that each trial should only include saccades, and fixations or smooth pursuits (slow eye movements). There is a lot of evidence in the literature that these fast and slow eye movements can be distinguished based on their velocity (Andersson et al., 2017; Dar et al., 2020; Dorr et al., 2010; Komogortsev et al., 2010; Komogortsev & Karpov, 2013; Larsson et al., 2013, 2015; Lüken et al., 2022; Salvucci & Goldberg, 2000; Startsev et al., 2019). This should be especially the case for the current data because the fastest target speed for smooth pursuits was 6°/s.

Moving circle trials should include smooth pursuits and (possibly) unintended saccades. *Back-and-forth circle* trials should include smooth pursuits and saccades. *Jumping circle* trials should include fixations and saccades. Based on these assumptions I developed the following protocol to create ground truth labels for the current data set as a basis for my further investigation.

Calculating dynamic velocity thresholds. First, a dynamic velocity threshold was calculated for each trial. This threshold was determined by taking the median velocity within the second 500ms of a trial and adding 1.5 times the 75-percentile velocity in the same time window. I selected this time window because targets were standing still for *jumping circle* trials or moving consistently for *moving circle* and *back-and-forth circle* trials. Velocities during this period should therefore represent the slow eye movements of a given trial.

I refined how thresholds are calculated iteratively by plotting arbitrarily selected trials for participants f7bb2338 and 6cde27b5. According to my findings, I selected the second 500ms because there appeared to be fewer blinks and saccades than during the first 500ms. Additionally, I used the median and 75-percentile rather than mean and standard deviation because they were more robust against unintended fast eye movements from saccades, blinks, and eye tracking artefacts. Finally, I selected the scaling constant of 1.5 because it seemed to classify all clearly visible saccades without misclassifying smooth pursuits and fixations.

Distinguishing fast and slow eye movements. All samples faster than the threshold were classified as saccades and the rest as fixations for *jumping circle* trials, or smooth pursuits for *moving circle* and *back-and-forth circle* trials.

Relabeling short events. After labeling eye movements according to the threshold, smooth pursuits, and fixations shorter than 10ms were reclassified as saccades. Afterwards, any remaining saccades shorter than 10ms were relabeled as fixations for *jumping circle* trials or smooth pursuits for *moving circle* and *back-and-forth circle* trials. I used a 10ms threshold because (Lappi, 2016) reports the shortest saccade duration as 10 milliseconds. Even though fixations usually last more than 100ms (Lappi, 2016), I did not use different minimum durations for fixations and smooth pursuits because the plots for participants f7bb2338 and 6cde27b5 indicated that the procedure was working well (INCLUDE FIGURE). The resulting labels were treated as the ground truth for the remaining analyses.

Results

In the end of the finetuning procedure, blinks and clearly visible saccades were reliably classified for participants f7bb2338 and 6cde27b5. This suggests that fast and slow eye movements were reliably distinguished by the velocity threshold.

Discussion

Classification algorithms fail to reliably distinguish fixations and smooth pursuits (Komogortsev & Karpov, 2013). High quality benchmark data is required to find new features and build better algorithms. However, this provides a challenge because human annotated data is no gold standard for these eye movements (Hooge et al., 2018). This results in a dilemma because automatic classification is required to avoid human bias but needs to be improved first using unbiased data. I tried to solve this problem based on the idea that fixations and smooth pursuits (slow eye movements) could be distinguished from saccades based on velocity thresholds. Based on this assumption I tried to restrict trials that fixations and smooth pursuits should not cooccur and created ground truth labels according to dynamic velocity thresholds. While the results suggest that this approach worked relatively well, it has some limitations.

First, any fixations that occurred during *moving circle* or *back-and-forth circle* trials were mislabeled as smooth pursuits. While smooth pursuits were theoretically impossible for *jumping circle* trials because they require moving targets (Lappi, 2016), it is impossible to fully prevent fixations during *moving circle* and *back-and-forth circle* trials. For example, if participants were distracted and looked away from the target, fixations likely occurred. To address this, the occurrence of unintended fixations was minimized by design. Specifically, *moving circle* and *back-and-forth* circle targets were designed to be easy to track without catchup saccades or intermittent fixations. Additionally, I told participants that it was very important to track targets accurately and frequently reminded them to focus and take breaks. Finally, I collected about 1.5mio rows (24 minutes) of data per participant. Even though it was impossible to exclude unwanted fixations, I am confident that my precautions and the amount of data limited their occurrence to a minimum.

Second, my own bias likely affected classification of ground truth labels. Even though I avoided direct human annotation, I manually finetuned the threshold to detect saccades, and the window to remove blinks. Consequently, ground truth labels are likely affected by my own bias. Nevertheless, as saccades are accurately distinguished from slow eye movements by their velocity, this procedure is based on a solid theoretical foundation. Additionally, I did not influence the classification of fixations and smooth pursuits directly, nor did I evaluate or compare their ground truth labels. Therefore, any bias of saccade detection should be systematic across fixation and smooth pursuit trials and thus not affect their investigation.

Third, blinks, saccades, or eye tracking artefacts during the second 500ms of a trial affect the classification of saccades. The velocity threshold to detect saccades is determined according to velocities within the second 500ms of each trial. Velocities during this period should represent fixations or smooth pursuits. However, if blink, saccades, or artefacts dominated this period, the threshold would likely be set too high, and saccades would be misclassified as fixations or smooth pursuits. This could seriously affect the analysis of feature distributions. Thus, I tried to reduce the likelihood of this problem as much as possible. First, I tried to remove as many artefacts before and after blinks as possible. Second, I used the second 500ms of each trial because they appeared to include fewer blinks, saccades, and artefacts than the first. Third, I used the median and 75-percentile because they were relatively robust against unintended fast samples. Finally, this limitation was also addressed by making targets easy to track, encouraging participants, and collecting many trials for each participant. In retrospect I could also have set an upper limit for the threshold based on thresholds used in the literature (e.g., Rigas et al., 2018). However, from visual inspection I am confident, that the procedure worked overall and that saccades were successfully separated from fixations and smooth pursuits.

Future studies should investigate alternative preprocessing procedures. For example, 2-state hidden Markov models could be fitted to each trial to distinguish slow and fast eye movements. This approach would completely avoid human influence. Additionally, classifications would be unaffected by unintended blinks, saccades, or artefacts during the second 500ms because it would be based on the entire trial. However, such an approach would increase preprocessing time considerably and other more lightweight alternatives should be investigated as well.

In addition to these limitations, data collection and preprocessing were not perfect. For example, some participants could not be calibrated, and I suspect mounting the EyeLink on an arm would have offered more flexibility than the current setup of a table mounted EyeLink. Additionally, some target positions collected for *back-and-forth circle* trials were incorrect and had to be recalculated during preprocessing. Most importantly, I forgot to consider the additional 100ms of eye movements collected before and after each trial which means that target and eye positions were shifted by 100ms, and the target positions were incorrectly interpolated for the final 100ms of each trial. As I only realized this during writing the report, I was unable to reanalyze the data. Nevertheless, before publishing the report and data, I propose to repeat all analyses after correction the preprocessing protocol.

In conclusion, I am confident that I addressed the limitations regarding ground truth labels adequately and that this initial attempt to create a benchmark dataset without human annotation was successful. Therefore, I considered it as a ground truth baseline to evaluate the classification performance of gazeHMM in the next section. As indicated above, this analysis and all other analyses reported here should be repeated after correcting the matching of eye and target positions.

gazeHMM Evaluation

gazeHMM Classification

Method

I expected that gazeHMM fails to distinguish between fixations and smooth pursuits because their feature distributions are too similar, especially for slow targets. An alternative explanation would be that features are too noisy in naturalistic data but can distinguish clear fixations and smooth pursuits in the current data set. To exclude this possibility, I first investigated if gazeHMM confused fixations and smooth pursuits in the current benchmark data set before investigating its features.

Each trial was classified separately with a 4-state gazeHMM. I used a 4-state gazeHMM even though I only expected two eye movements (or three including PSO) to be present for each trial. This was because gazeHMM only detects smooth pursuits if a 4-state model is selected (Lüken et al., 2022). After classifying the data with gazeHMM, I selected one random trial per movement-speed combination for each participant and visually compared gazeHMM classifications with the ground truth. In total gazeHMM classified 1008 trials and I inspected 63 trials (9 per participant). I specifically looked for rapid switching between fixations and smooth pursuits which was identified as a problem by Lüken et al. (2022). Also, I did not calculate any agreement measures (for an overview see Startsev & Zemblys, 2022) because rapid switching would be better identified through visual inspection and my primary goal was to investigate gazeHMM features and not to quantify its accuracy.

As part of my exploratory analysis, I also used a fork of gazeHMM by Simon Kucharsky (Lüken & Kucharský, 2023). This version of gazeHMM can classify multiple trials at once which ensured that all eye movements were present and which better justifies the use of a 4-state model. I used this version of gazeHMM to classify 48 trials with target speeds of 6°/s for some participants in the training set and compared its classifications with classifications by the original gazeHMM and the ground truth of the benchmark data set. Fitting each of these models took several hours and I did not repeat the procedure with all participants in the training set or lower target speeds.

The analysis was conducted in R (R Core Team, 2022) and Python (Van Rossum & Drake Jr, 1995). See Appendix A and B for a list of the most central packages used.

Results

Planned analysis. As expected, gazeHMM failed to distinguish fixations and smooth pursuits and rapidly switched between them for most trials (see Figures 1-3). For some participants, switching was less prevalent at faster target speeds (see Figures 4 and 5). However, overall switching occurred at all target speeds (see Figures 6 and 7) and there was no consistent evidence that gazeHMM is better at distinguishing fixations from faster smooth pursuits.

Exploratory analysis. Apart from the planned analysis, I want to point out additional findings from the visual inspection.

Classification of saccades. gazeHMM frequently classified saccades consistent with the baseline (see Figures 2, 4, and 5). This suggests that saccades were accurately detected during preprocessing and the ground truth is accurate for saccades without human annotation.

Missing saccades. In some *jumping circle* and *back-and-forth circle* trials gazeHMM classified no saccades. Instead, gazeHMM classified series of rapid switches between fixations and smooth pursuits (see Figures 3 and 7).

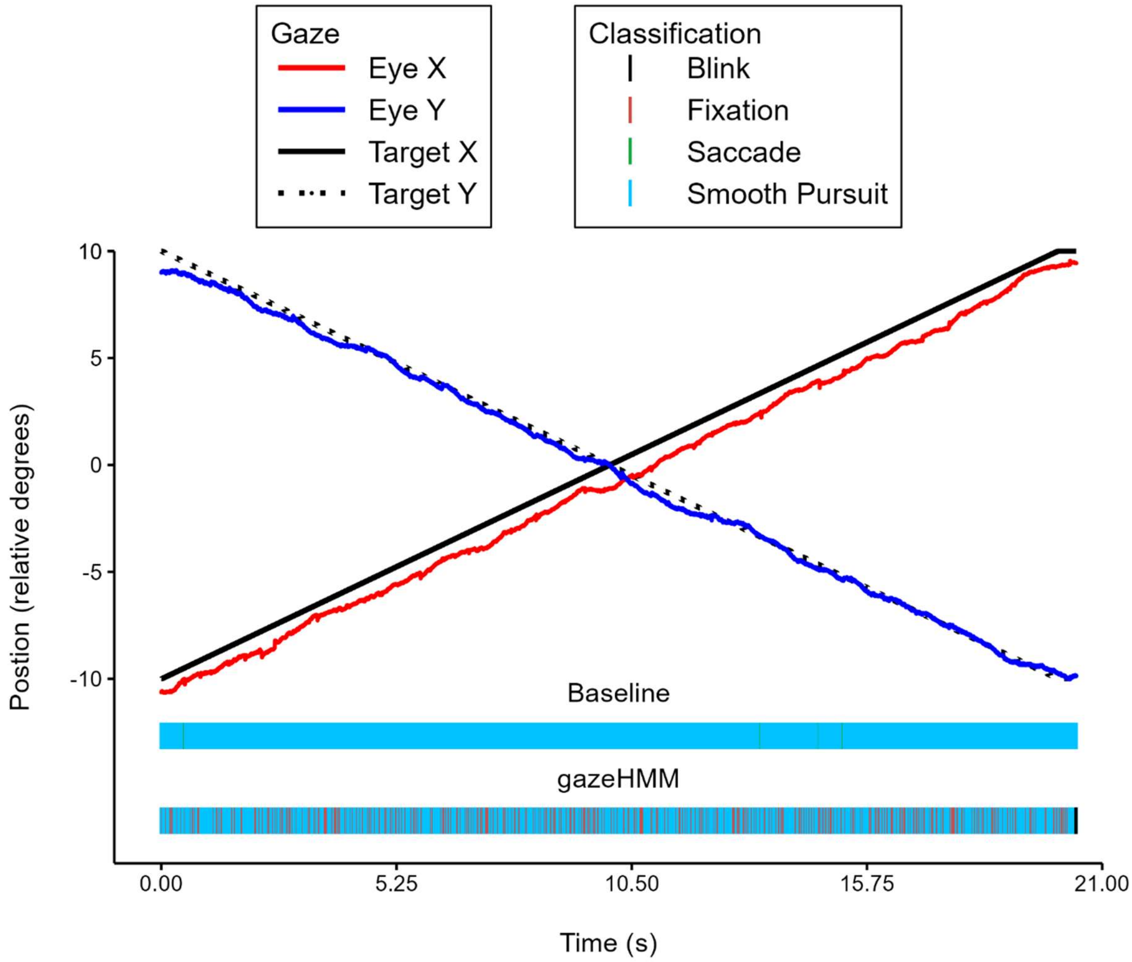
Label switching. In some trials the labels of two eye movements were switched. Fost frequently, labels of saccades and smooth pursuits were switched (see Figure 5). Hidden Markov models find solutions by maximizing a likelihood function. This function is symmetrical for the states it contains. Therefore, it can converge to a solution in which the arbitrary labels of two or more states are switched (Visser & Speekenbrink, 2022).

Inconsistency. Overall, the performance of gazeHMM was very inconsistent. I didn't identify any patterns according to movement pattern, target speed, or trajectory. This is very well illustrated by Figures 5 and 7. These figures show *jumping circle* trials at 6°/s with a vertical trajectory from bottom to top for participants 6cde27b5 (Figure 5) and f7bb2338 (Figure 7). Both participants displayed similar gaze paths. For participant 6cde27b5 gazeHMM switched smooth pursuit and saccade labels but matched the ground truth nearly perfectly. In contrast, for participant f7bb2338 gazeHMM did not classify any saccades and rapidly switched between fixations and smooth pursuits.

Fitting multiple trials. Potentially, a four-state model did not fit *jumping circle* trials because fewer actual states created its data. To address this, I fitted the gazeHMM fork (Lüken & Kucharský, 2023) as explained above. Figure 5 shows that this approach solved the problem of label switching: saccades were still detected reliably and now correctly labeled. However, Figure 5 also shows that the overall classification was much worse as gazeHMM now switched rapidly between fixations and smooth pursuits.

Figure 1

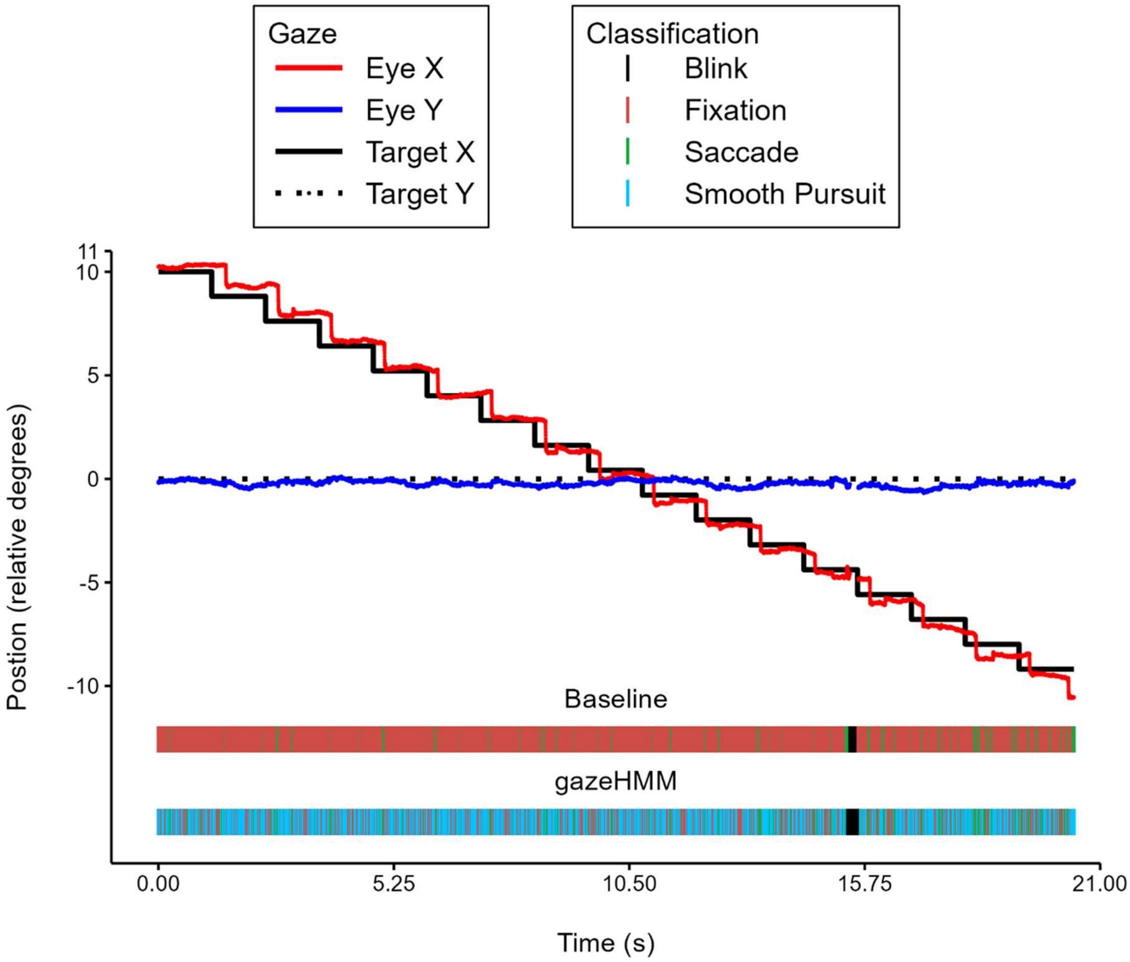
Gaze Path and Classification for Trial 84 of Participant cf910821 (Moving Circle, 1°/s, ↘)



Note. This plot shows trial 84 for participant cf910821. In this *moving circle* trial, the target moved consistently at 1°/s on a diagonal trajectory from top left to bottom right. According to the baseline this was a clear smooth pursuit with very few saccades. In contrast, gazeHMM switched rapidly between fixations and smooth pursuits.

Figure 2

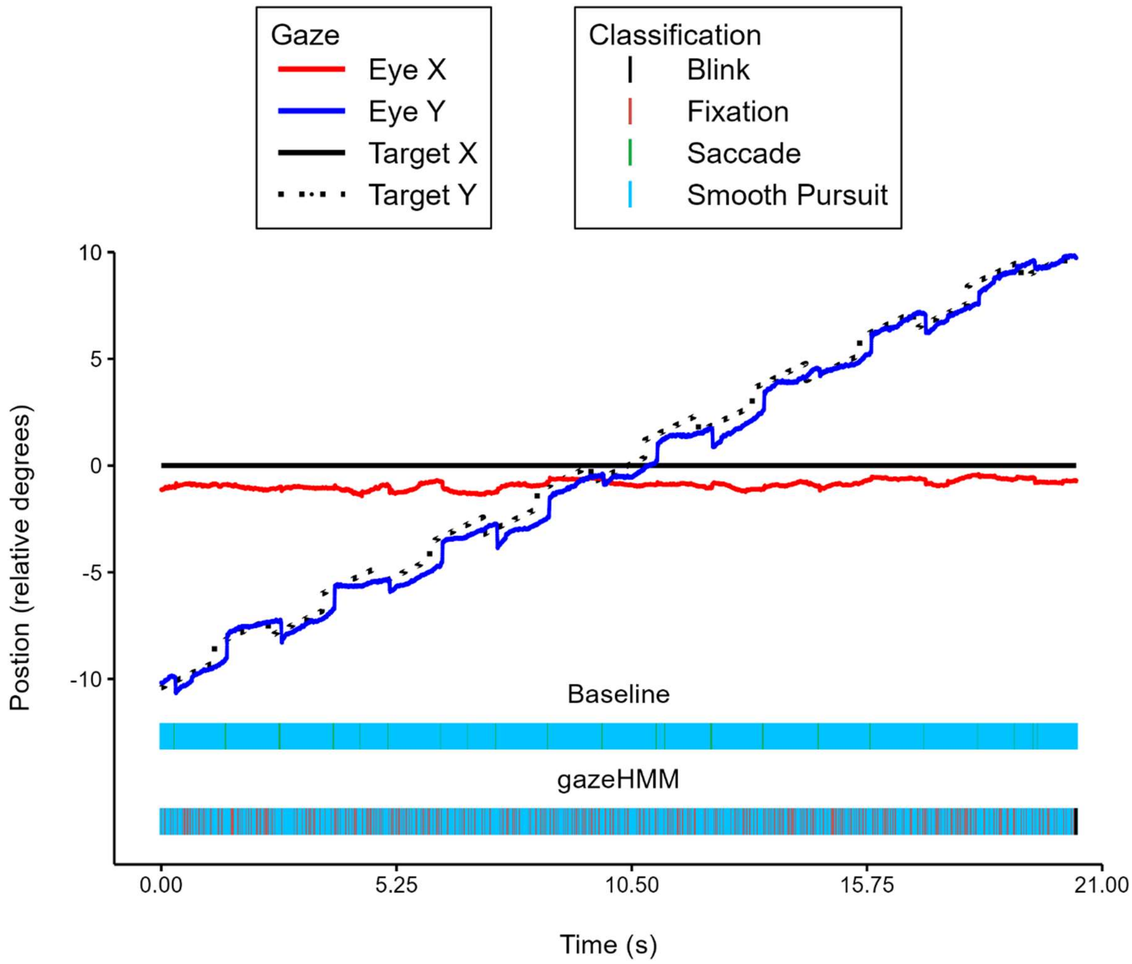
Gaze Path and Classification for Trial 3 of Participant cf910821 (Jumping Circle, 1°/s, ←)



Note. This plot shows trial 3 for participant cf910821. In this *jumping circle* trial, the target jumped 1° at 1000ms intervals along a horizontal trajectory from right to left. From the gaze path, this was a clear series of fixations and saccades. gazeHMM classified largely the same saccades as classified for the baseline but switched rapidly between fixations and smooth pursuits.

Figure 3

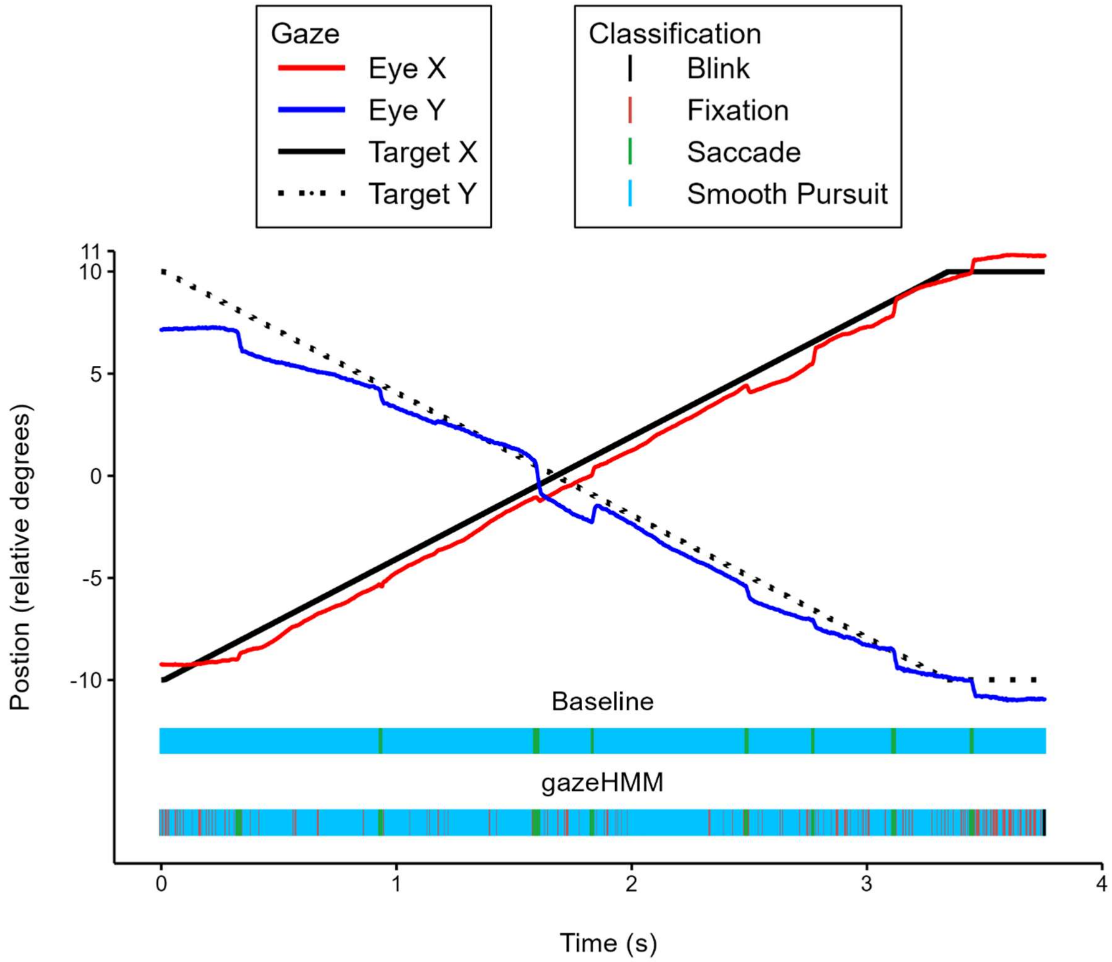
Gaze Path and Classification for Trial 17 of Participant cf910821 (BF Circle, 1°/s, ↑)



Note. This plot shows trial 17 for participant cf910821. In this *back-and-forth circle* trial, the target moved consistently at 1°/s on a vertical trajectory from the bottom to the top. Additionally, it jumped back and forth at 1000ms intervals. According to the baseline this was a clear smooth pursuit with intermittent saccades when jumps occurred. gazeHMM classified no saccades and switched rapidly between fixations and smooth pursuits.

Figure 4

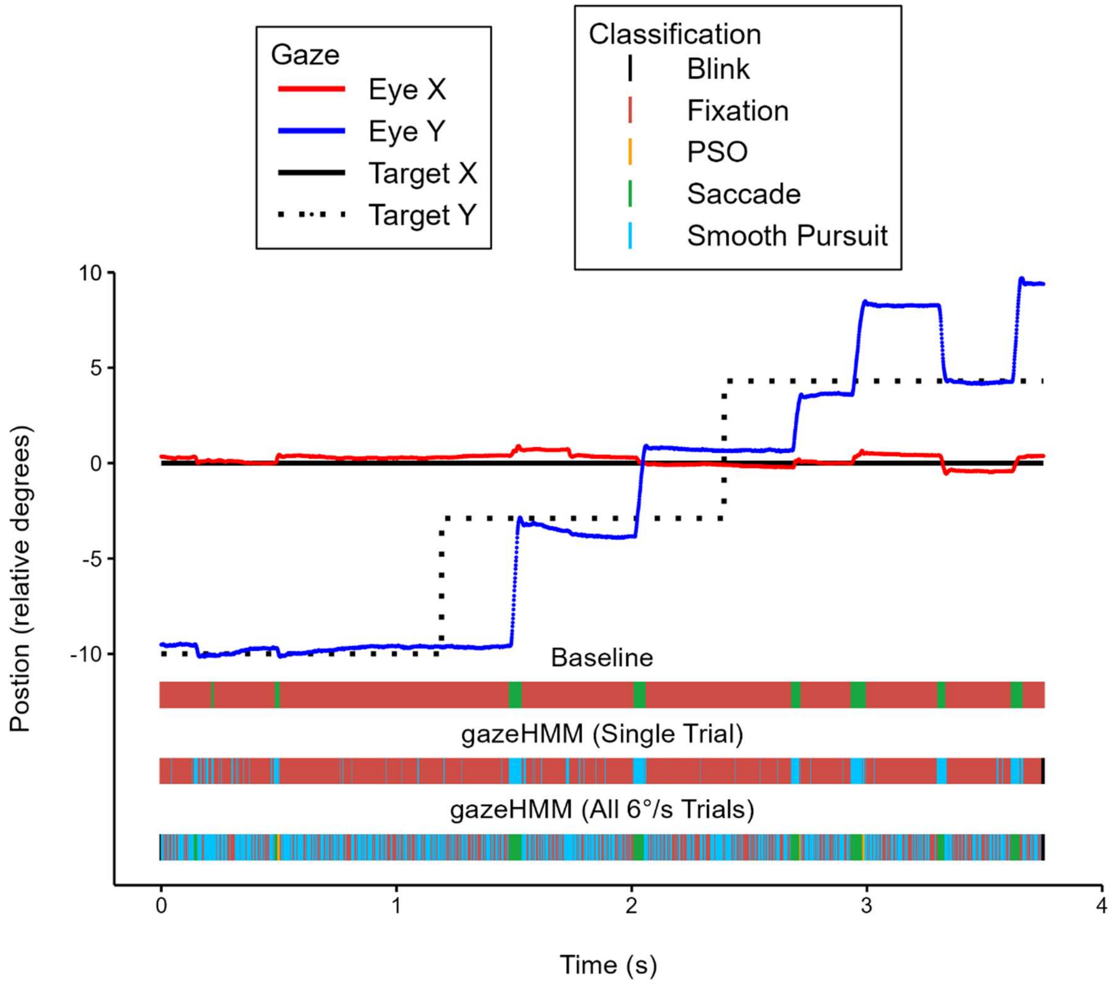
Gaze Path and Classification for Trial 55 of Participant 6cde27b5 (Moving Circle, 6°/s, ↘)



Note. This plot shows trial 55 for participant 6cde27b5. In this *moving circle* trial, the target moved consistently at 6°/s on a diagonal trajectory from top left to bottom right. According to the baseline this was a clear smooth pursuit with intermittent saccades. gazeHMM classified largely the same saccades as classified for the baseline. gazeHMM showed relatively little switching between fixations and smooth pursuits for this trial.

Figure 5

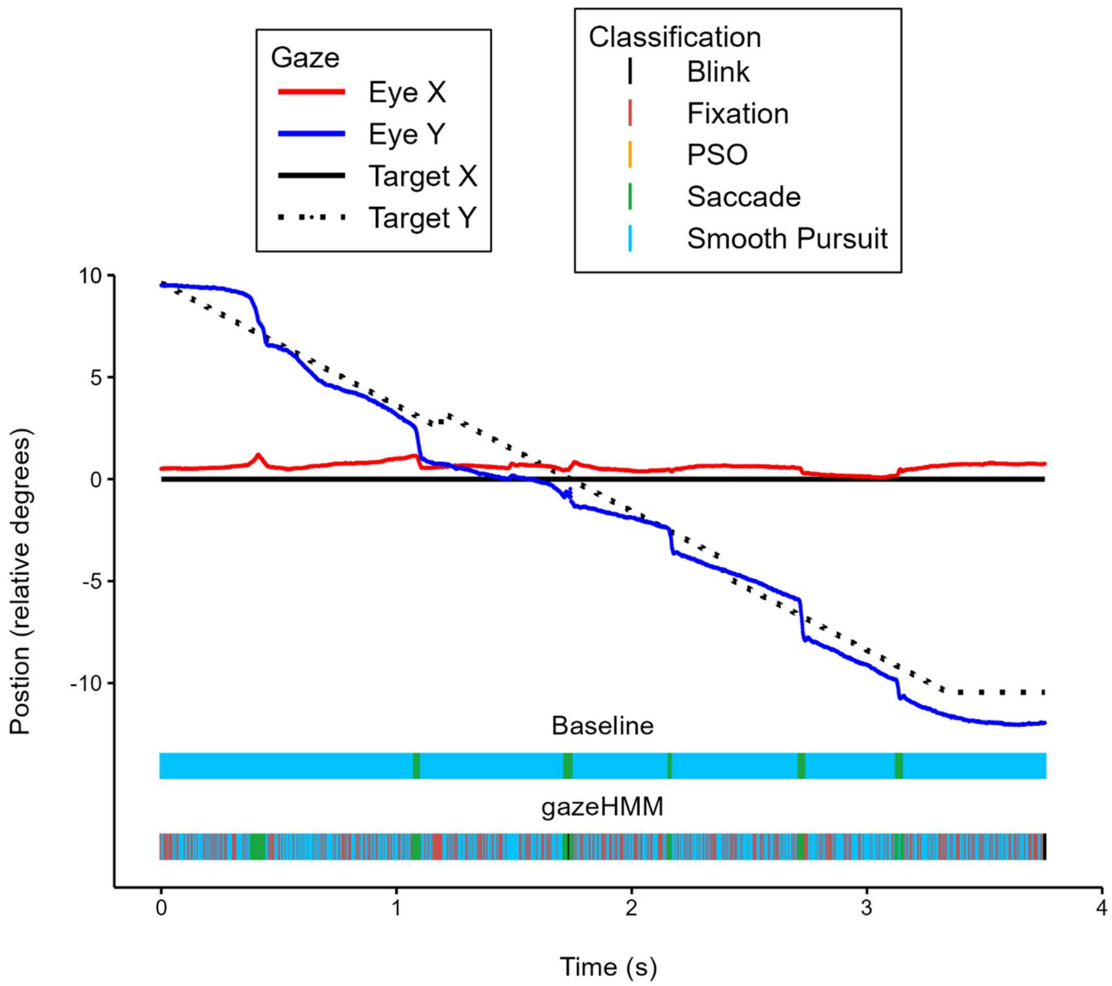
Gaze Path and Classification for Trial 126 of Participant 6cde27b5 (Jumping Circle, 6°/s, ↑)



Note. This plot shows trial 126 for participant 6cde27b5. In this *jumping circle* trial, the target jumped 6° at 1000ms intervals along a vertical trajectory from the bottom to the top. According to the baseline this was a clear series of fixations and saccades. gazeHMM fitted to this single trial suffered from label switching between smooth pursuits and saccades but agreed almost perfectly with the baseline. gazeHMM fitted to all 6°/s trials of this participant classified saccades well without label switching, but rapidly switched between fixations and smooth pursuits.

Figure 6

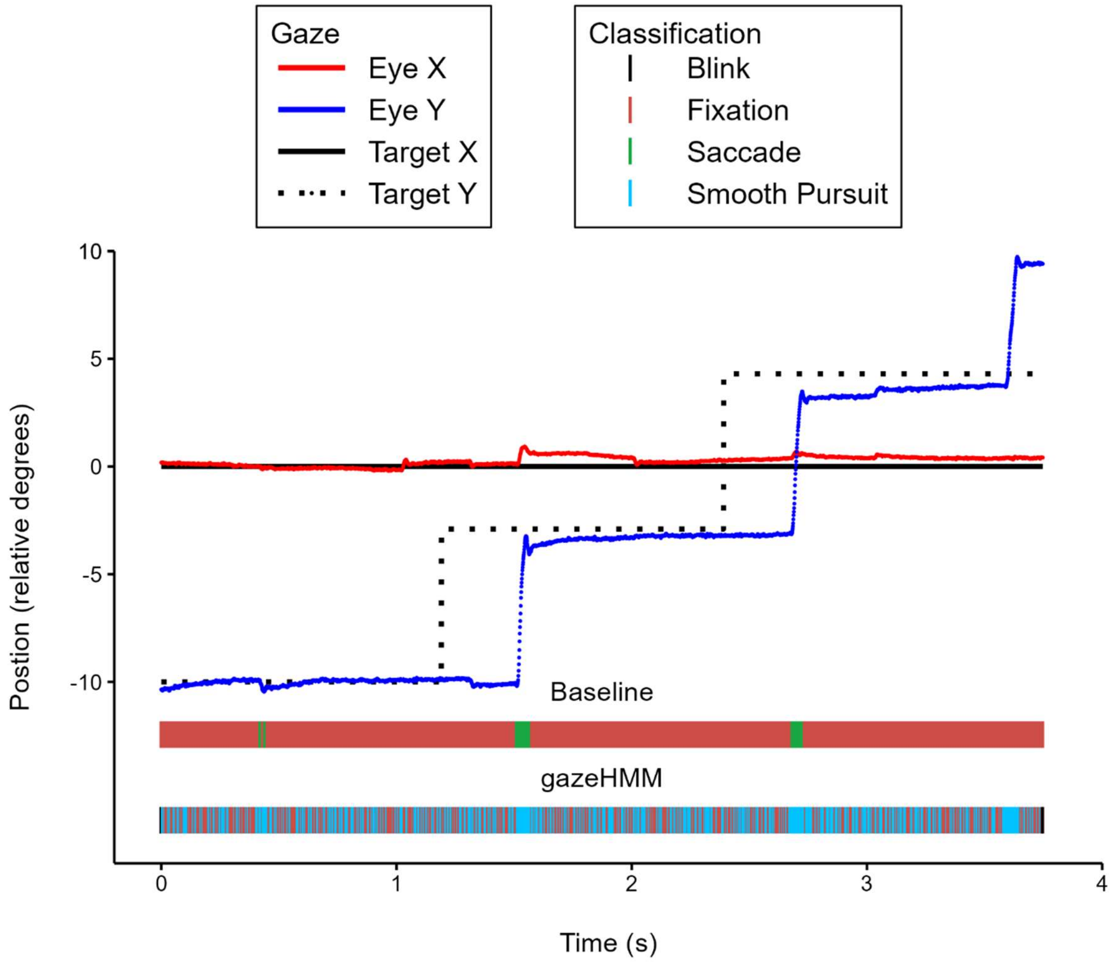
Gaze Path and Classification for Trial 126 of Participant 68471e16 (BF Circle, 6°/s, ↓)



Note. This plot shows trial 126 for participant 68471e16. In this *back-and-forth circle* trial, the target moved consistently at 6°/s on a vertical trajectory from the bottom to the top. Additionally, it jumped back and forth at 1000ms intervals. According to the baseline this was a clear smooth pursuit with saccades when jumps occurred. gazeHMM classified largely the same saccades as classified for the baseline but switched rapidly between fixations and smooth pursuits.

Figure 7

Gaze Path and Classification for Trial 126 of Participant f7bb2338 (Jumping Circle, 6°/s, ↑)



Note. This plot shows trial 126 for participant f7bb2338. In this *jumping circle* trial, the target jumped 6° at 1000ms intervals along a vertical trajectory from the bottom to the top. According to the baseline this was a clear series of fixations and saccades. gazeHMM classified no saccades and switched rapidly between fixations and smooth pursuits.

Discussion

As expected gazeHMM failed to distinguish fixations and smooth pursuits on the current data set. This is consistent with my expectation that the features used by gazeHMM are distributed similarly for fixations and smooth pursuits and rejects the alternative explanation that naturalistic data is too noisy.

Overall, gazeHMM classification performance was very inconsistent. It classified some trials nearly perfectly while switching rapidly between fixations and smooth pursuits on other similar trials. Interestingly, gazeHMM classified more smooth pursuits than fixations overall. One possible explanation for this would be that the 4-state model did not fit *jumping circle* trials well because they include fewer actual states. However, I addressed this by fitting the forked gazeHMM (Lüken & Kucharský, 2023). While this approach prevented label switching, it made the classification of fixations worse for some trials. This suggests that gazeHMM has a bias towards smooth pursuits for slow eye movements and indicates that its underlying assumptions about fixations and smooth pursuits may be wrong. Another way to address the problem that a 4-state model doesn't align with the number of expected states would be to fit 2-state models to all trials. However, in its current state gazeHMM only detects smooth pursuits if four states are selected by the user.

While gazeHMM struggled to distinguish fixations and smooth pursuits, its classification of saccades generally coincided with the baseline. This is evidence that saccades were accurately detected during preprocessing and that the data set can be considered a benchmark without human annotation.

Therefore, I used it to investigate gazeHMM features of ground truth saccades, fixations, and smooth pursuits in the next section.

gazeHMM Features

Method

To investigate my expectation that features are distributed very similarly for fixations and smooth pursuits, I calculated the velocity, acceleration, and sample-to-sample angle according to (Lüken et al., 2022). These calculations were built on the original R code (Lüken, 2020) but were done in Python (Van Rossum & Drake Jr, 1995).

The velocity and acceleration were defined as the first and second degree of a Savitzky-Golay filter with a polynomial order of three and a window size of five. Importantly, I identified a potential mistake in the gazeHMM calculation of velocity and acceleration. Specifically, they treated samples separated by missing values as consecutive samples. I avoided this and therefore my calculations differ from the original gazeHMM calculations for trials with missing values.

The sample-to-sample angle was defined as the relative angle of two consecutive samples. My calculation followed the original code. However, I converted radians to degrees and did not mirror negative values. Therefore, sample-to-sample angles calculated here vary between -360° and 360° which results in distributions that are easy to read.

For each feature, I looked at plots across all speeds and for each target speed for the entire training set and each participant. Plots across targets speeds were based on 1008 trials for the complete training set and 144 trials for individual participants. Plots for each target speed were based on 336 trials for the complete training set and 48 trials for individual participants. Generally, the distributions of saccades included larger values which made it more difficult to compare fixations and smooth pursuits. As the focus of this study lied on these two eye movements, I only included saccades in the plots across all trials.

To investigate my expectation that feature distributions are similar for fixations and smooth pursuits especially at low target speeds, I inspected these plots visually.

The analysis was conducted in R (R Core Team, 2022) and Python (Van Rossum & Drake Jr, 1995). See Appendix A and B for a list of the most central packages used.

Results

Planned analysis. As expected, the plots show that distributions of velocity, acceleration, and sample-to-sample angle are distinct for saccades but very similar for fixations and smooth pursuits (see Figures 8-10). Also as expected, feature distributions for fixations and smooth pursuit become more distinct with higher target speeds (see Figures 11-13). However, this pattern is least visible for sample-to-sample angles (see Figure 13). Also, Figures 14-19 illustrate that this pattern is more pronounced for some participants (e.g., 06b8d2d3) than others (e.g., 7d248f8f).

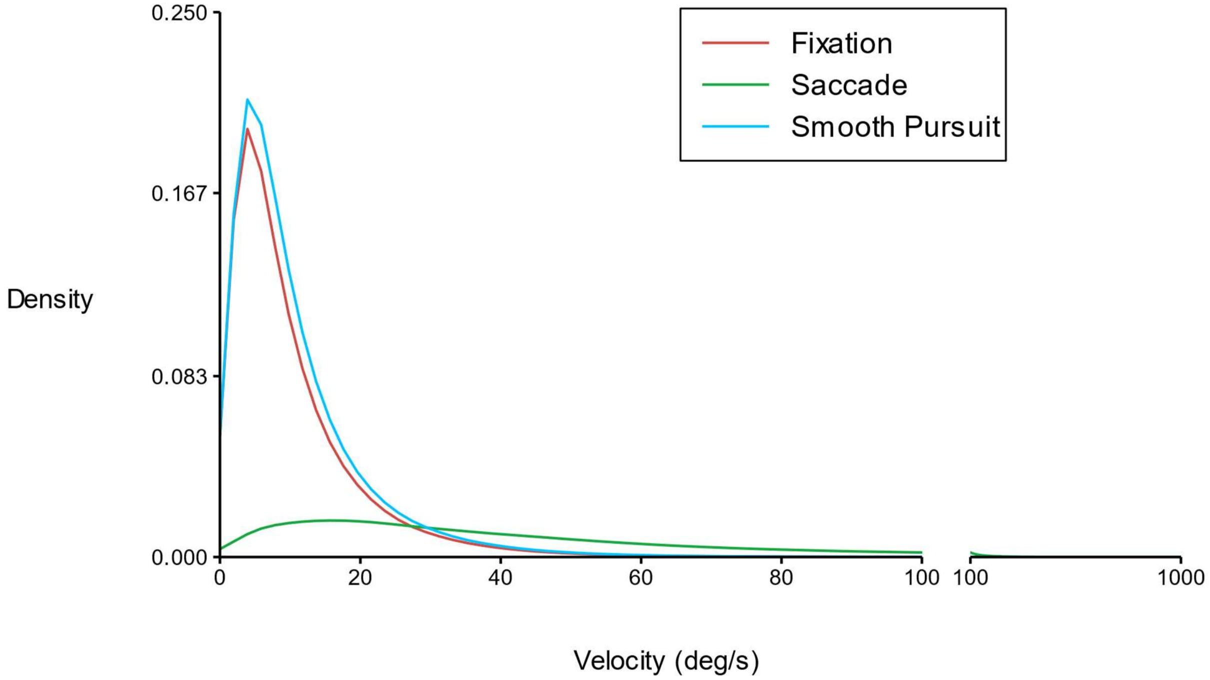
Exploratory analysis. In addition to my planned analysis, I want to point out some exploratory findings about sample-to-sample angles.

Are fixations random walks? Lüken et al. (2020) expected that fixations are random walks characterized by uniform sample-to-sample angle distributions. Instead, fixations are clearly distributed around zero degrees. This suggests, that fixational movements like saccades and smooth pursuits are directed movements on the scale of milliseconds.

Multiple peaks. Sample-to-sample angle distributions for fixations and smooth pursuits clearly peak at zero. However, interestingly they also have several smaller peaks at 45-degree intervals. The highest of these peaks are at 45- and 90-degrees and they become lower for larger angles. Additionally, peaks at 90°, 180°, and 270° are higher than the 45° steps between them (i.e., 135°, 225°). To exclude the possibility that this was an artifact of specific movement patterns or trajectories, I plotted the distributions separately for only *moving circles* (Figure 20), *jumping circles* (Figure 21), and *back-and-forth circles* (Figure 22), and only horizontal trajectories from left to right (Figure 23). The figures show that the pattern was stable across all conditions, thereby suggesting that this is not an artifact of stimulus design. This phenomenon was also consistent across all participants in the training set and very clear for all but two participants (e.g., 7d248f8f; see Figure 19).

Figure 8

Velocity Distributions Across All Trials and All Participants

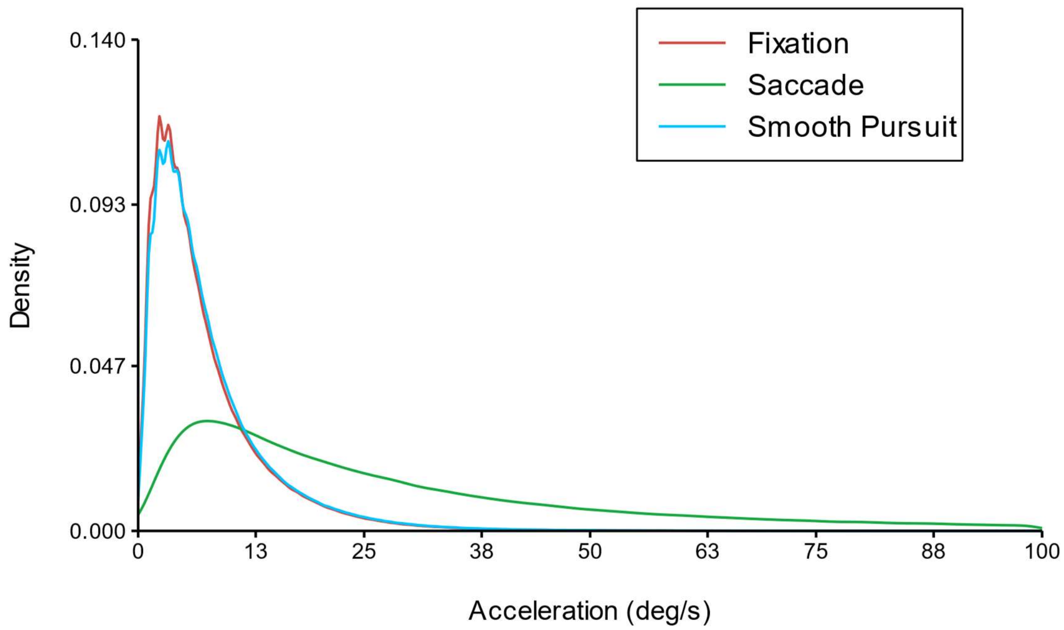


Note. This plot shows velocity densities for saccades (green), fixations (red), and smooth pursuits (blue) for all 1008 trials included in the training set. The distributions of fixations and smooth pursuits are very similar.

The x axis is broken to focus the plot on the area where most densities lie.

Figure 9

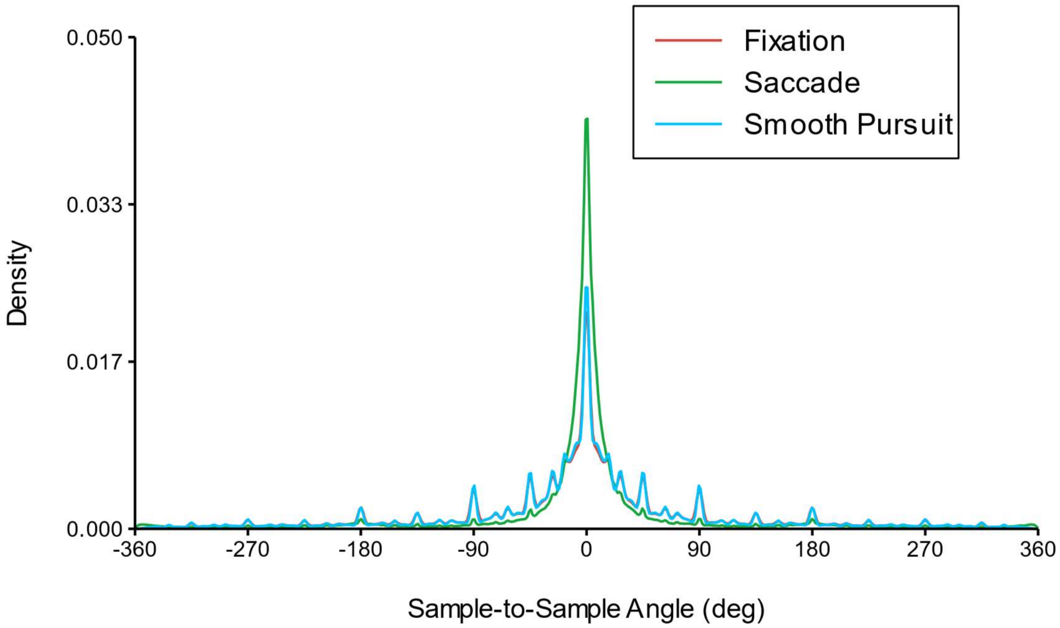
Acceleration Distributions Across All Trials and All Participants



Note. This plot shows acceleration densities for saccades (green), fixations (red), and smooth pursuits (blue) for all 1008 trials included in the training set. The distributions of fixations and smooth pursuits are very similar.

Figure 10

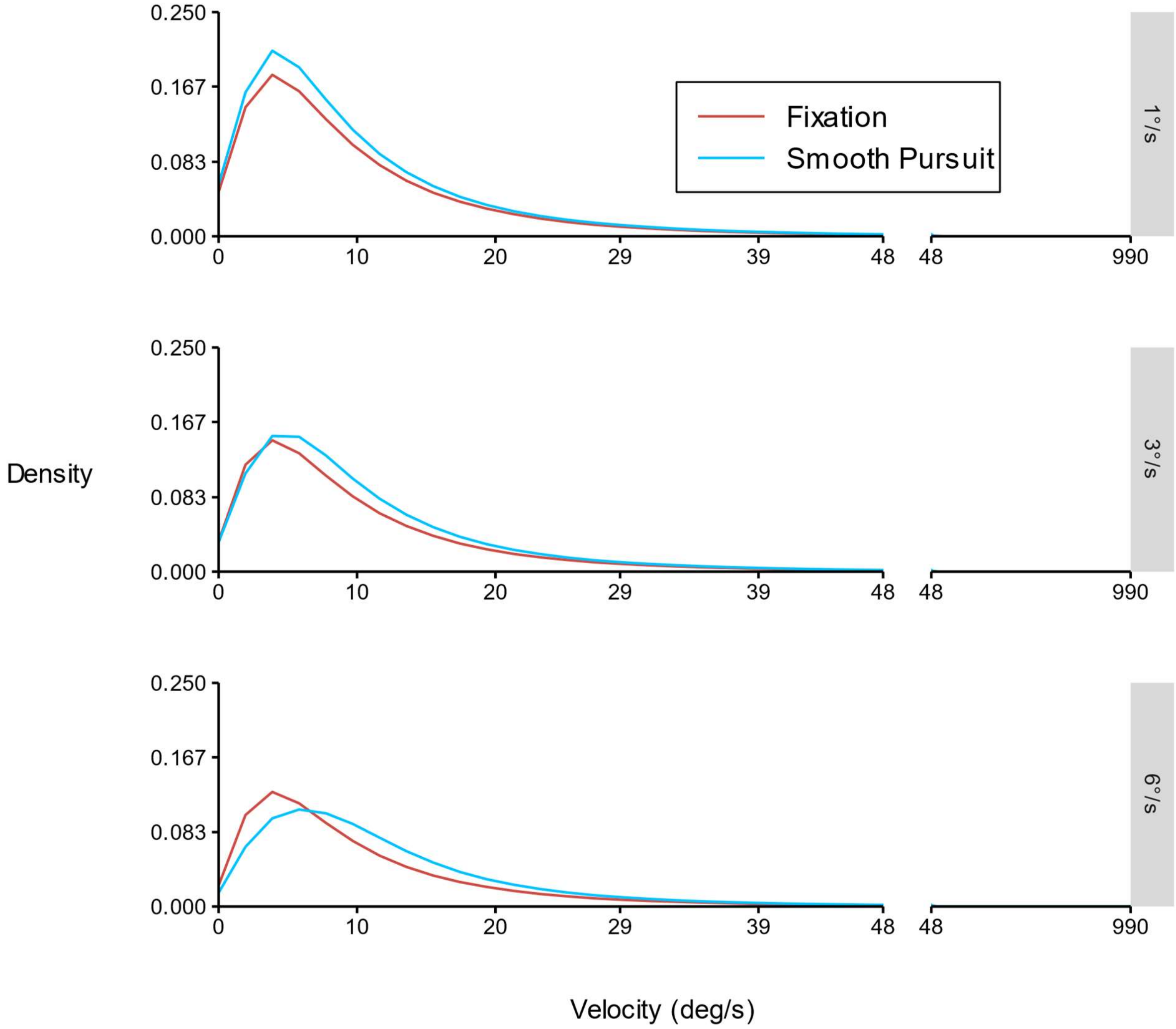
Sample-to-Sample Angle Distributions Across All Trials and All Participants



Note. This plot shows sample-to-sample angle densities for saccades (green), fixations (red), and smooth pursuits (blue). It is based on all 1008 trials included in the training set. The distributions of fixations and smooth pursuits are very similar. The distributions show several peaks at 45-degree intervals that get smaller with larger angles.

Figure 11

Velocity Distributions at Different Target Speeds for All Participants

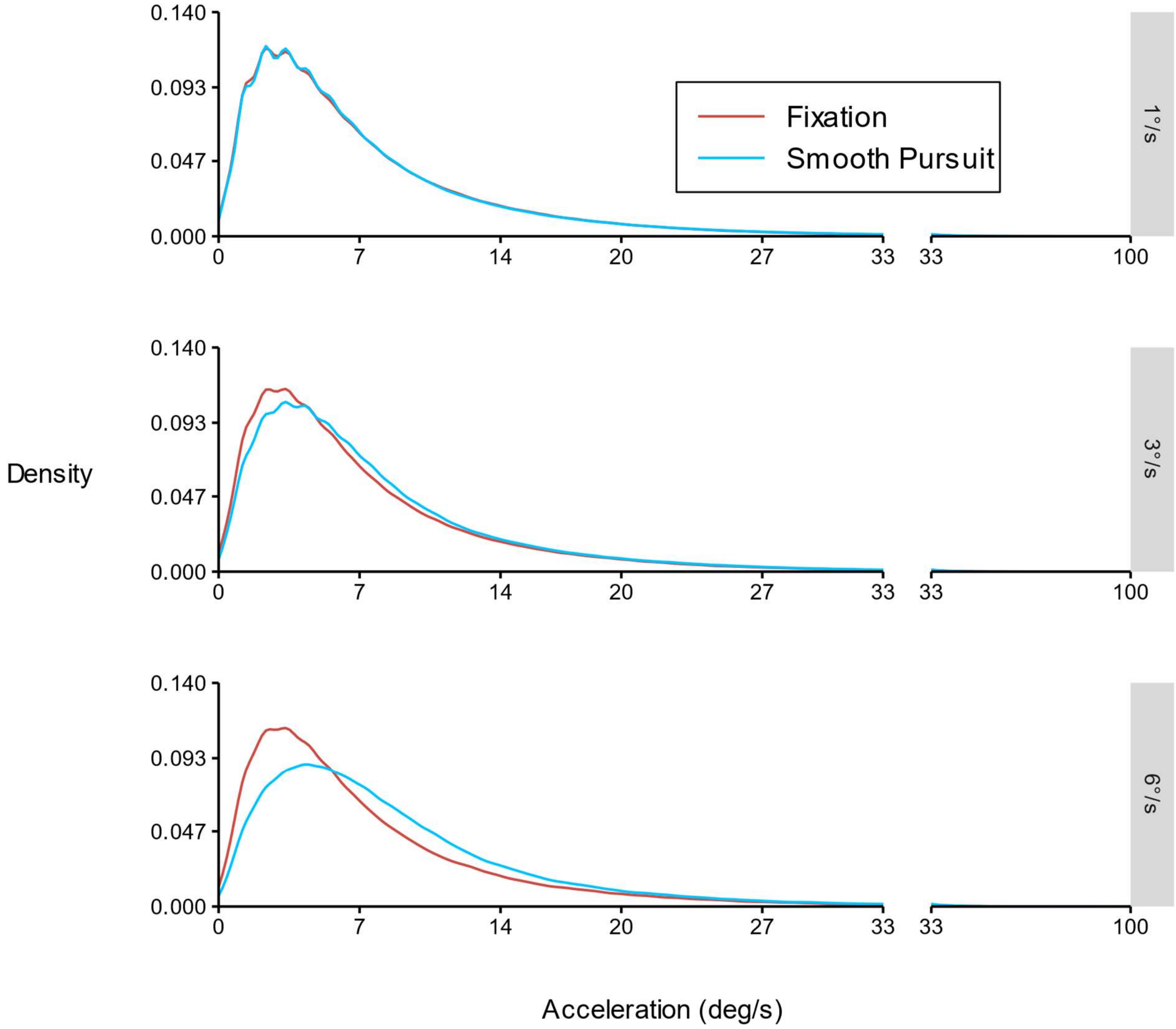


Note. These plots show velocity densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 336 trials of a given speed included in the training set. Distributions for fixations and smooth pursuits are very similar, but less so for faster targets.

The x axes are broken to focus the plots on the area where most densities are.

Figure 12

Acceleration Distributions at Different Target Speeds for All Participants

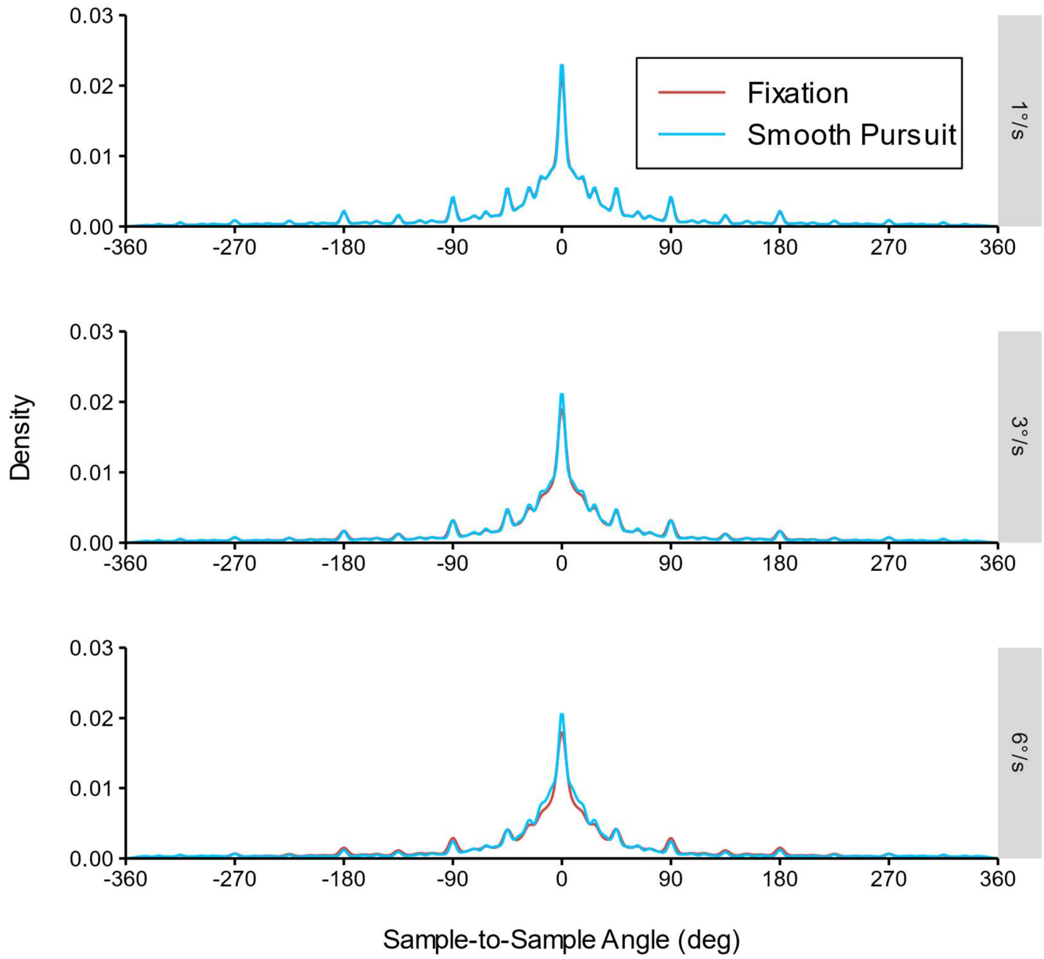


Note. These plots show acceleration densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 336 trials of a given speed included in the training set. Distributions for fixations and smooth pursuits are very similar, but less so for faster targets.

The x axis is broken to focus the plot on the area where most densities are.

Figure 13

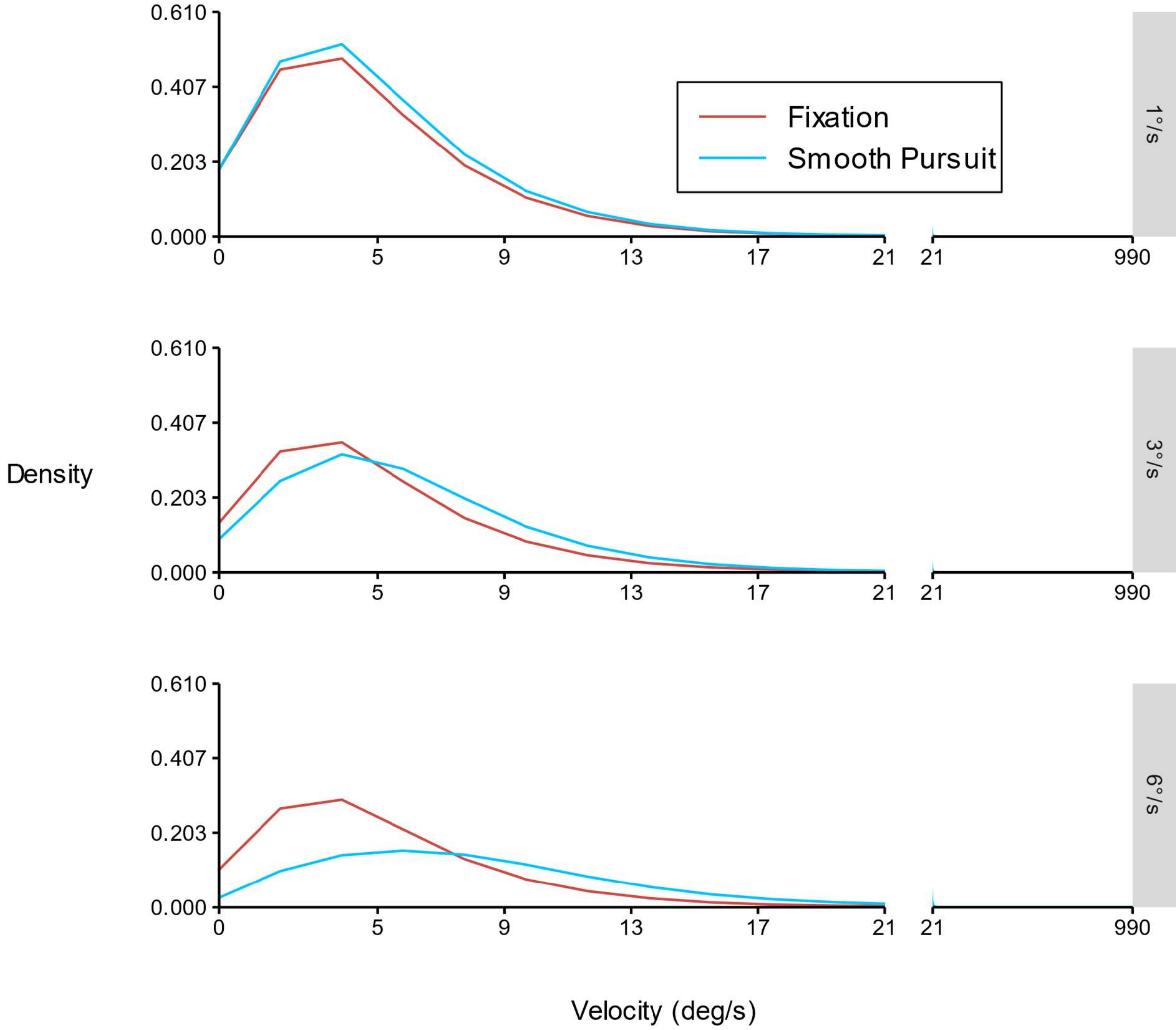
Sample-to-Sample Angle Distributions at Different Target Speeds for All Participants



Note. These plots show sample-to-sample angle densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 336 trials of a given speed included in the training set. Distributions for fixations and smooth pursuits are very similar with only slight differences for faster target speeds.

Figure 14

Velocity Distributions at Different Target Speeds for Participant 06b8d2d3

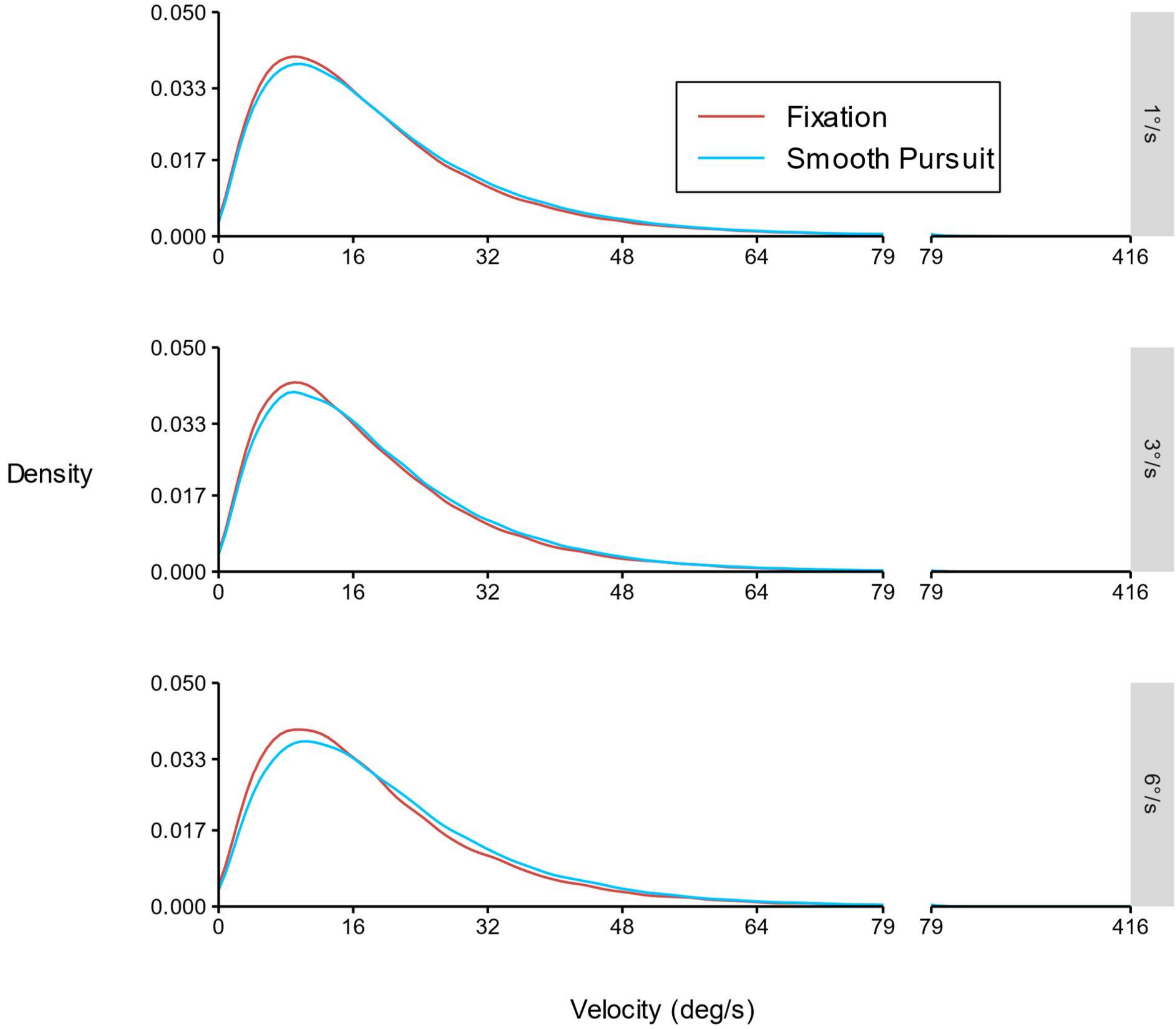


Note. These plots show velocity densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 48 trials of a given speed of participant 06b8d2d3. Distributions for fixations and smooth pursuits become clearly more distinct for faster targets for this participant.

The x axes are broken to focus the plots on the area where most densities are.

Figure 15

Velocity Distributions at Different Target Speeds for Participant 7d248f8f

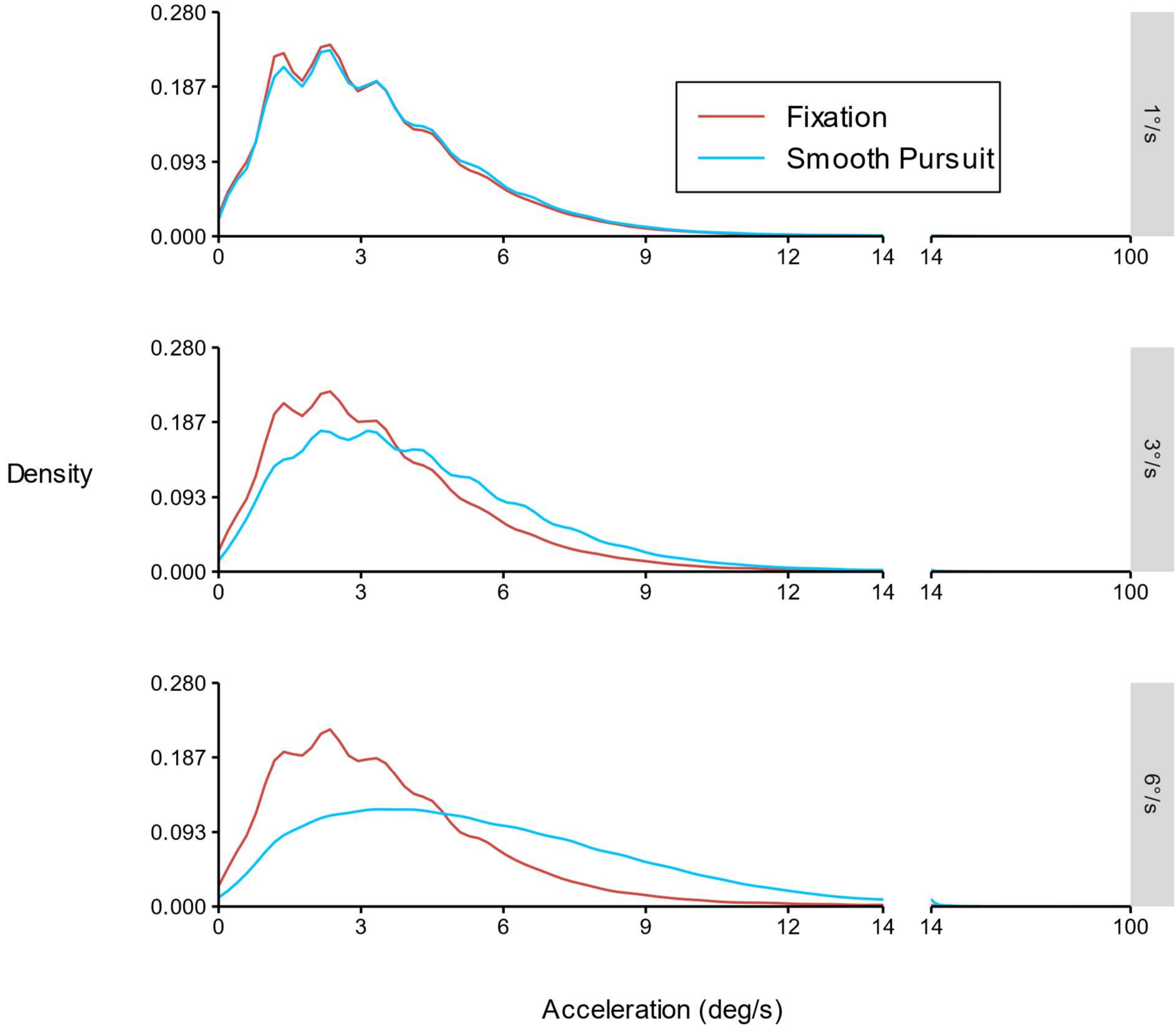


Note. These plots show velocity densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 48 trials of a given speed of participant 7d248f8f. Distributions for fixations and smooth pursuits stay very similar irrespective of target speed for this participant.

The x axes are broken to focus the plots on the area where most densities are.

Figure 16

Acceleration Distributions at Different Target Speeds for Participant 06b8d2d3

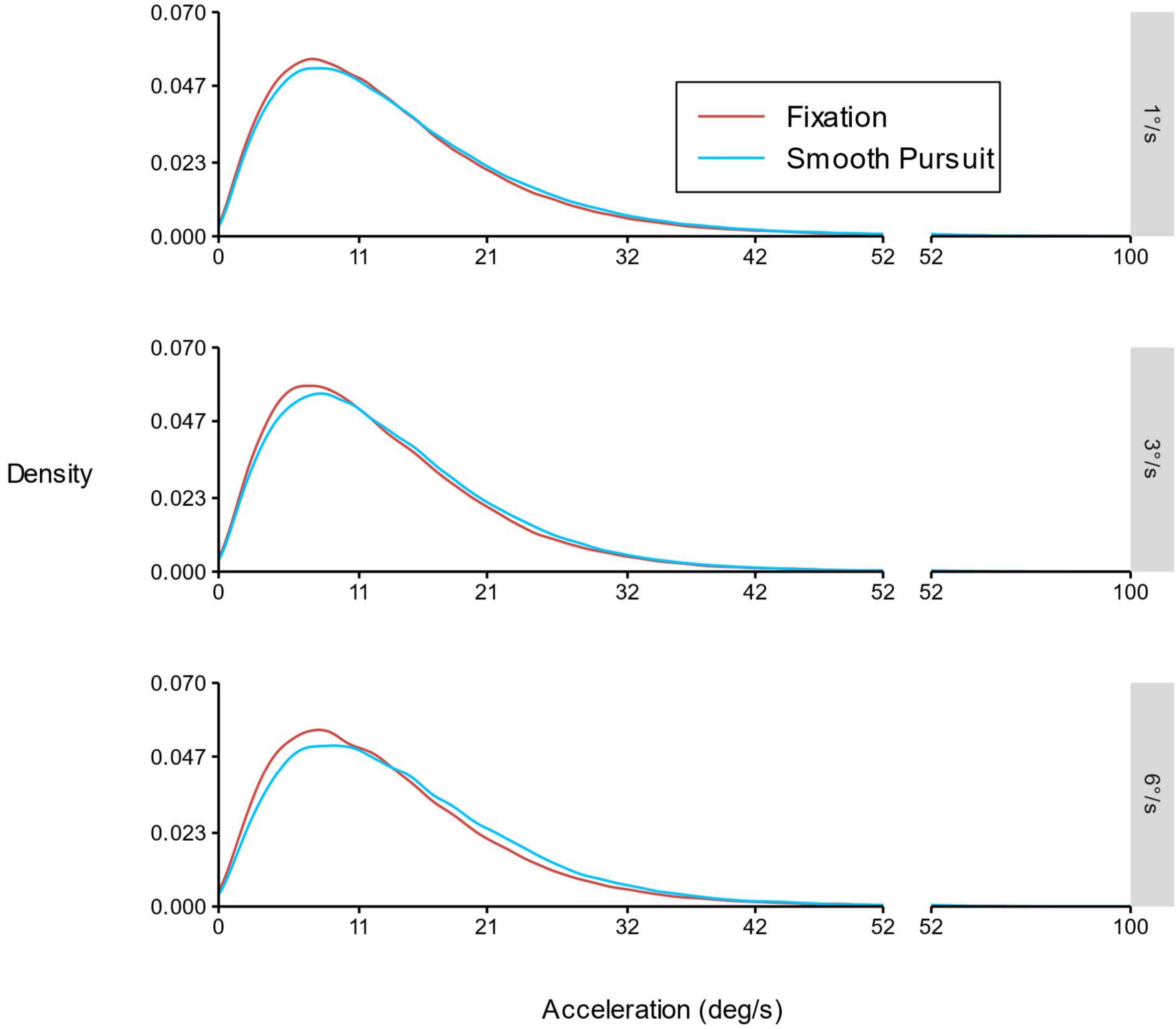


Note. These plots show acceleration densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 48 trials of a given speed of participant 06b8d2d3. Distributions for fixations and smooth pursuits become clearly more distinct for faster targets for this participant.

The x axes are broken to focus the plots on the area where most densities are.

Figure 17

Acceleration Distributions at Different Target Speeds for Participant 7d248f8f

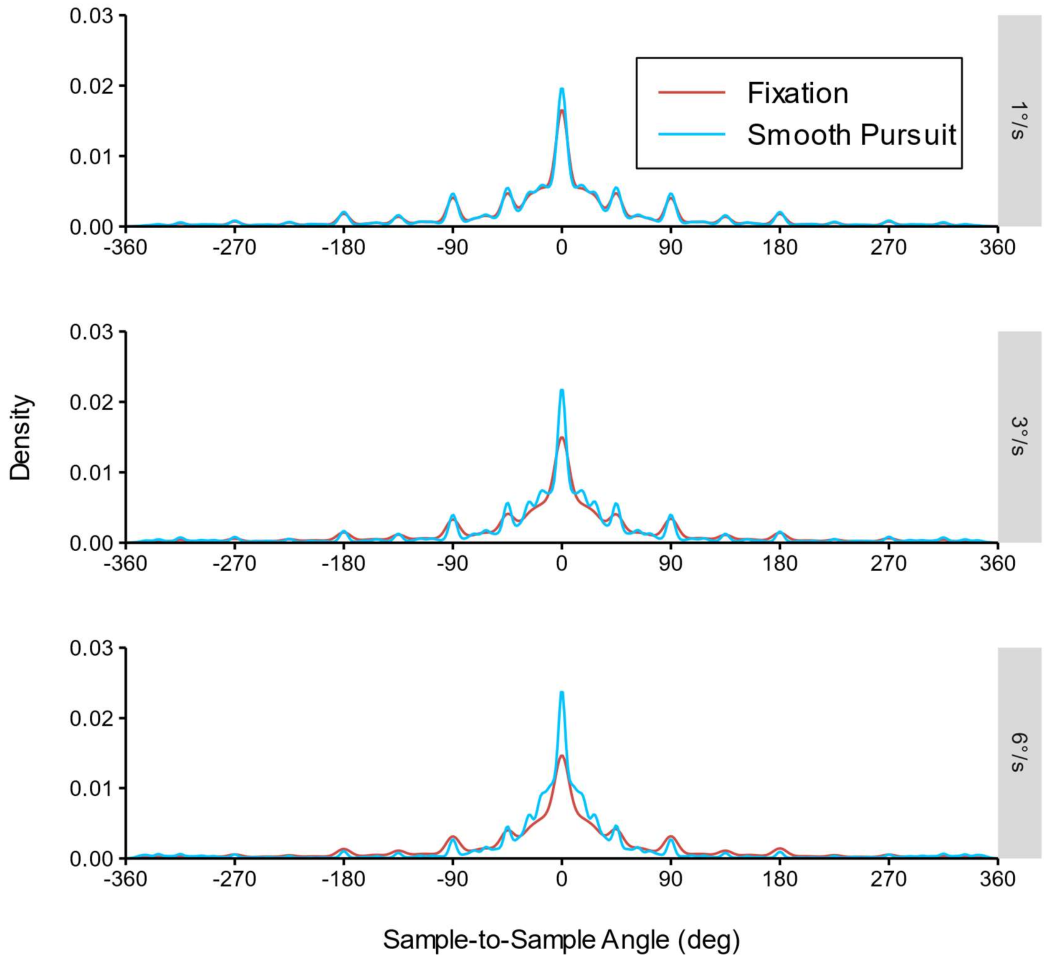


Note. These plots show acceleration densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 48 trials of a given speed of participant 7d248f8f. Distributions for fixations and smooth pursuits stay very similar irrespective of target speed for this participant.

The x axes are broken to focus the plots on the area where most densities are.

Figure 18

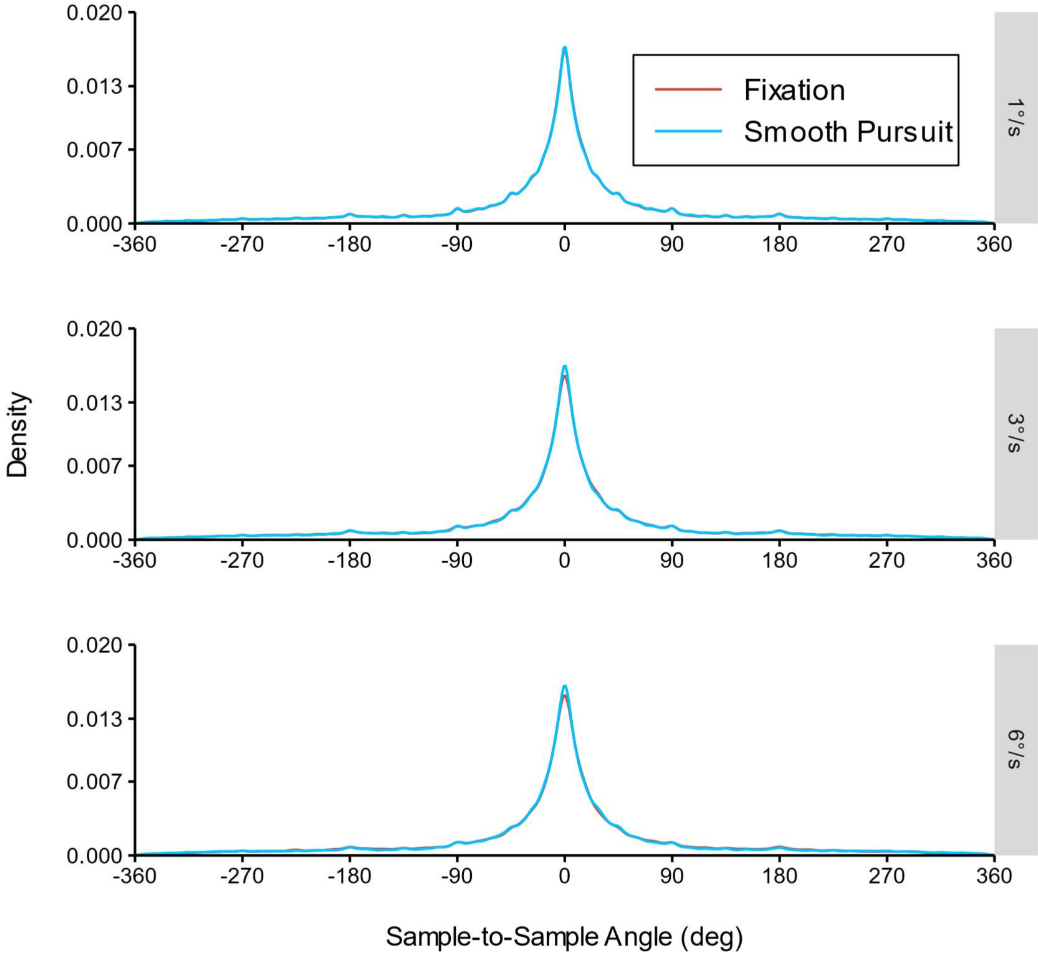
Sample-to-Sample Angle Distributions at Different Target Speeds for Participant 06b8d2d3



Note. These plots show sample-to-sample angle densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 48 trials of a given speed of participant 06b8d2d3. Distributions for fixations and smooth pursuits become clearly more distinct for faster targets for this participant. The distributions show very clear peaks at 45-degree intervals that get smaller with larger angles for this participant.

Figure 19

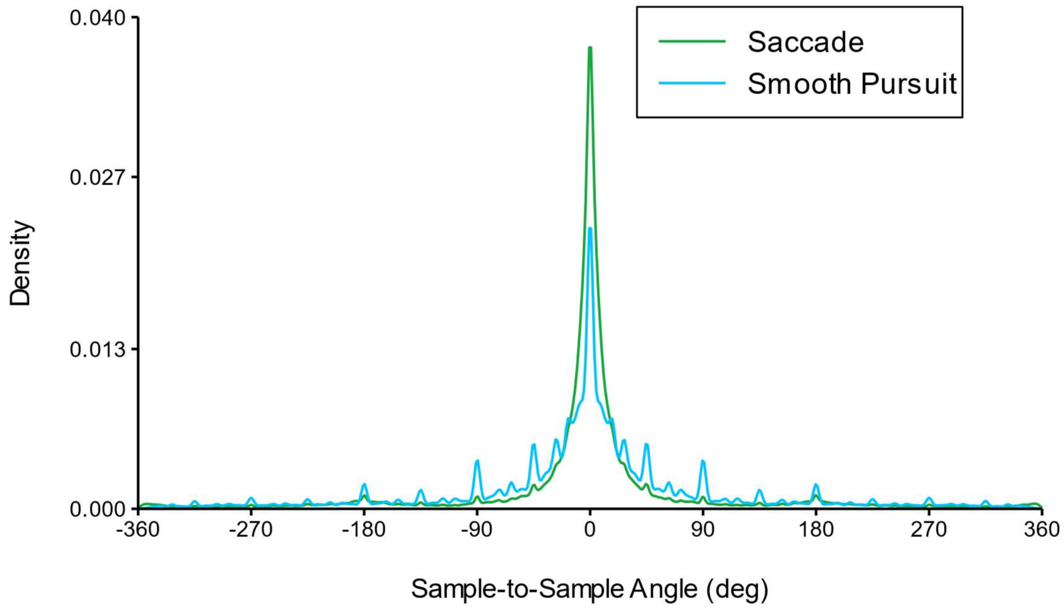
Sample-to-Sample Angle Distributions at Different Target Speeds for Participant 7d248f8f



Note. These plots show sample-to-sample angle densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 48 trials of a given speed of participant 7d248f8f. Distributions for fixations and smooth pursuits stay very similar irrespective of target speed for this participant. The distributions show only very small peaks at 45-degree intervals for this participant.

Figure 20

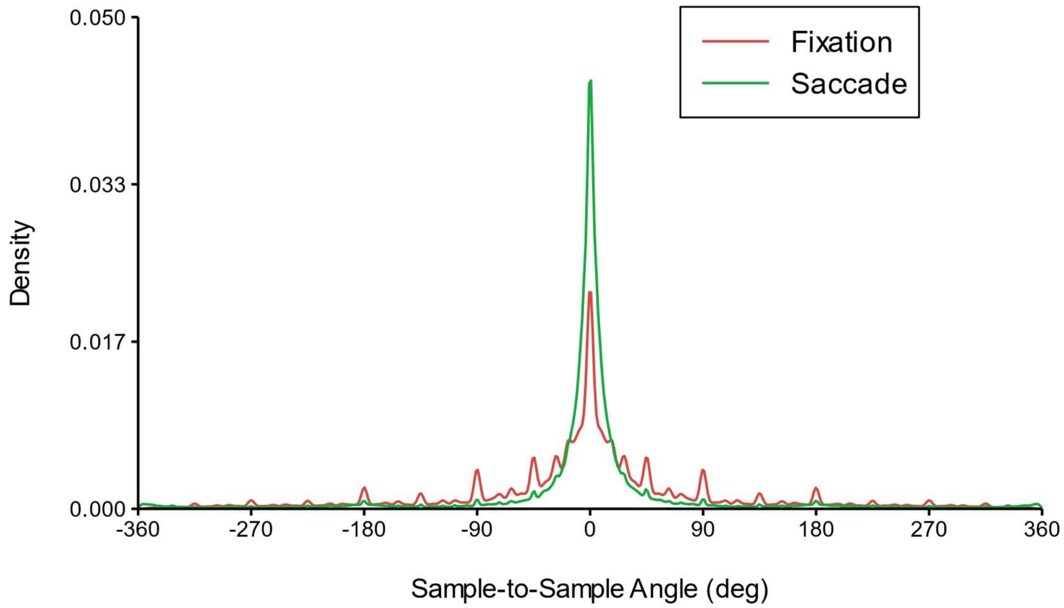
Sample-to-Sample Angle Distributions for Moving Circle Trials and All Participants



Note. This plot shows sample-to-sample angle densities for saccades (green), and smooth pursuits (blue). It is based on all 336 *moving circle* trials included in the training set. The distribution for smooth pursuits shows clear peaks at 45-degree intervals.

Figure 21

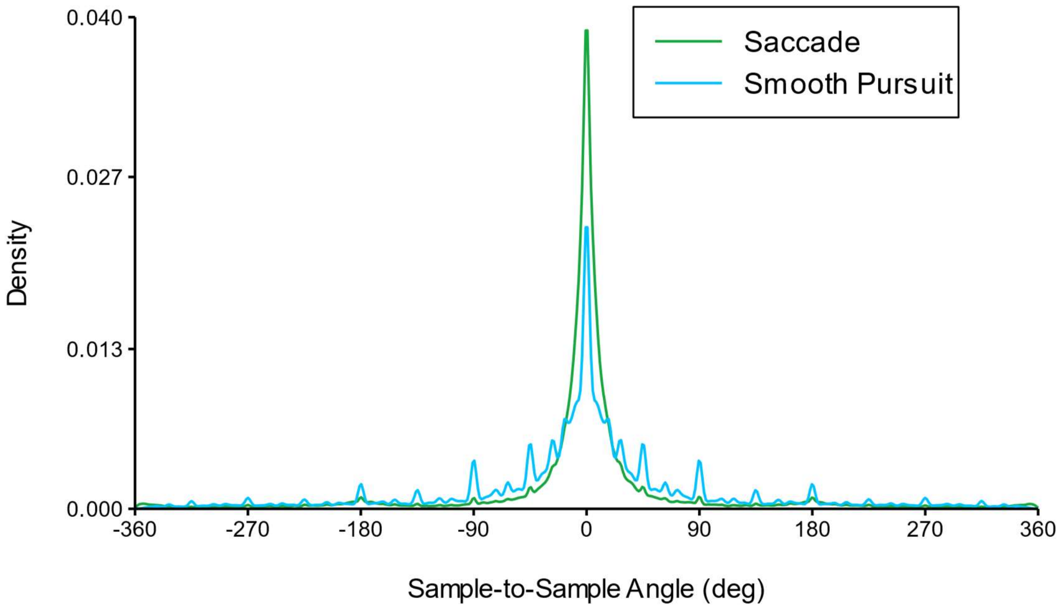
Sample-to-Sample Angle Distributions for Jumping Circle Trials and All Participants



Note. This plot shows sample-to-sample angle densities for saccades (green), and fixations (red). It is based on all 336 *jumping circle* trials included in the training set. The distribution for fixations shows clear peaks at 45-degree intervals.

Figure 22

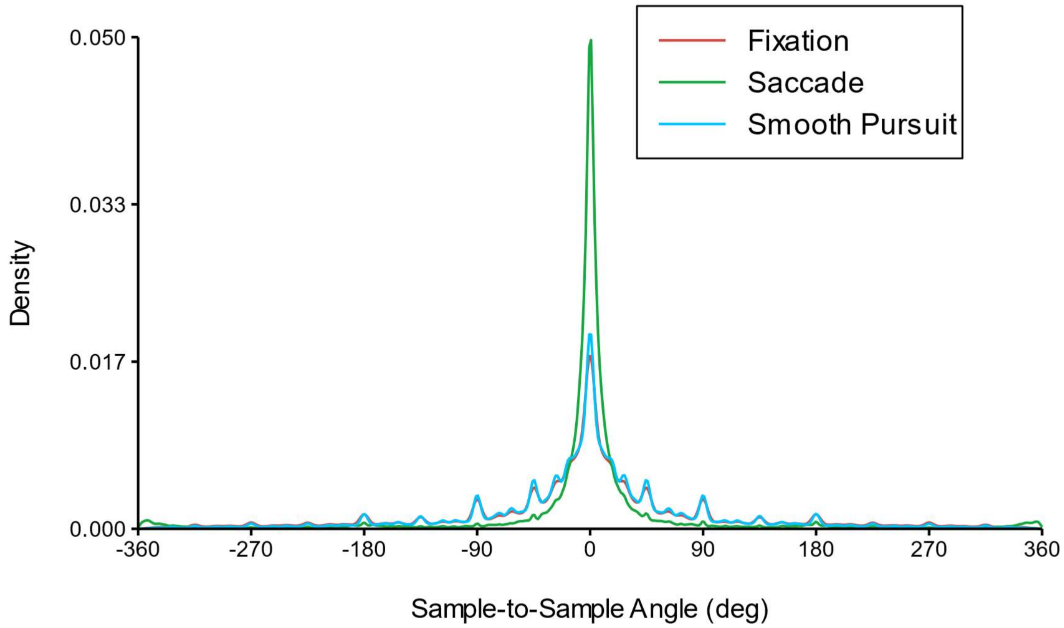
Sample-to-Sample Angle Distributions for Back-and-Forth Circle Trials and All Participants



Note. This plot shows sample-to-sample angle densities for saccades (green), and fixations (red). It is based on all 336 *back-and-forth circle* trials included in the training set. The distribution for smooth pursuits shows clear peaks at 45-degree intervals.

Figure 23

Sample-to-Sample Angle Distributions for Trials with Horizontal-Right Trajectories and All Participants



Note. This plot shows sample-to-sample angle densities for saccades (green), fixations (red), and smooth pursuits (blue). It is based on all 126 trials with a horizontal target trajectory from left to right included in the training set. The distributions for fixations and smooth pursuits shows clear peaks at 45-degree intervals.

Discussion

As expected gazeHMM features are distributed very similarly for fixations and smooth pursuits especially for slow targets. This is particularly the case for sample-to-sample angle distributions. These findings are in line with my expectation that gazeHMM fails to distinguish fixations and smooth pursuits because they are too similar in terms of their features.

Importantly, the plots for faster target speeds also supported the assumptions of (Lüken et al., 2022) to some degree. Specifically, smooth pursuits had slightly higher velocities and acceleration than fixations for targets moving 3°/s and 6°/s. Nevertheless, the sample-to-sample angle, which was supposed to distinguish fixations and smooth pursuits, was distributed very similarly at all target speeds. Importantly, both distributions were clearly distributed around zero which matches their expectations for smooth pursuits but not fixations (Lüken et al., 2022). This serves as a good explanation for the bias towards smooth pursuits that was revealed in the last section. Specifically, as most samples of both fixations and smooth pursuits have sample-to-sample angles around zero, they tend to fall into the smooth pursuit category rather than the fixation category which assumes uniformly distributed sample-to-sample angles.

Additionally, this indicates that fixations are directed movements at least on the scale of milliseconds. This is potentially the result of fixational eye movements like drift and micro saccades. Previous research found that these movements are not random but underlie some cognitive control (Krauzlis et al., 2017). A recent study found directional drift when participants were searching for different letters on a noisy background (Lin et al., 2023). Additionally, Chen and Hafed (2013) found in monkeys that drift corrected against micro saccadic displacement by moving in the opposite direction. This suggests that fixations are not random walks but potentially series of micro saccades and corrective drifts which would be consistent with the distribution of sample-to-sample angles found in the current study.

Interestingly sample-to-sample angles showed peaks at 45-degree intervals that decreased in height for larger angles and appeared larger for 90-degree steps than 45-degree steps in between. I excluded the possibility that this was an artifact of the stimulus design and expect that this is due to the muscles that control eye movements. Specifically, the eyes are moved by three pairs of antagonistic muscles which pull at the eyes to move them in a desired direction. Two of these muscle pairs attach to the eyeball horizontally and vertically (Lee et al., 2019). I expect that this is the reason why it is easier for these muscles to perform 45-and 90-degree angle changes than more precise movements but did not find any studies investigating this. This could also be an artifact based of the eye tracking algorithm of the EyeLink. However, as there were two participants for whom the pattern was much less pronounced, I deem this unlikely. Nevertheless, future studies should investigate sample-to-sample angles in data collected with different eye trackers or other methods that measure eye movements to investigate this possibility.

The current analysis is limited with respect to the feature distributions of saccades. Specifically, because sample velocity was used to classify saccades for the baseline, it is guaranteed that their velocity distribution is much more skewed towards extreme values than those of fixations and smooth pursuits. This also extends to the acceleration as it is highly related to the velocity. However, this is irrelevant for the current study because I focus on fixations and smooth pursuits. Nevertheless, this needs to be kept in mind if any future studies use the benchmark data set to investigate saccades.

Another limitation is that some fixations probably occurred during *moving circle* and *back-and-forth circle* trials and consequently were misclassified as smooth pursuits. If fixations during these trials were frequent, they would make the smooth pursuit distributions more like the distribution for fixations in *jumping circle* trials. This would seriously affect the interpretability of my findings. However, I have addressed this limitation before (see discussion of benchmark data set section). Additionally, if I would find other features that clearly distinguish between ground truth fixations and smooth pursuits in the benchmark data set, this would indicate that fixations and smooth pursuits are correctly classified in the baseline and that my findings are valid. Therefore, I investigated the direction deviation of different eye movements in the benchmark data set and developed new eye movement features in the next section.

Feature Engineering

I wanted to come up with a suitable candidate to replace the sample-to-sample angle feature in gazeHMM with one that would distinguish fixations and smooth pursuits more reliably. I did this exploratorily after investigating the classification performance and feature distributions of gazeHMM.

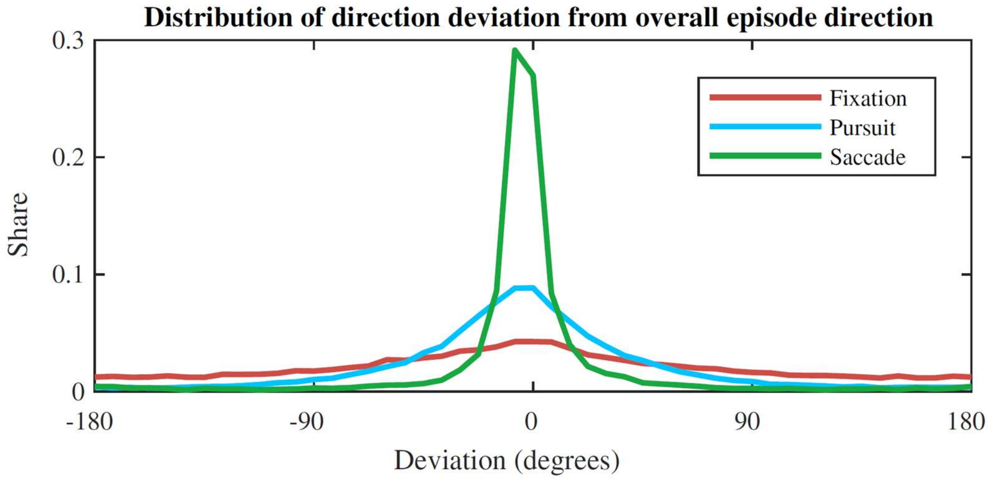
Many authors tried to develop features based on the idea that smooth pursuits are regular movements that can be described by some overall trajectory or shape. This idea is best illustrated by Vidal et al. (2012) who created seven "shape features" to characterize smooth pursuits. Similarly, sample-to-sample angle are used by gazeHMM because smooth pursuits are expected to be directed eye movements while fixations are not (Lüken et al. 2022).

Startsev et al., (2019) found some evidence supporting this general idea. They calculated windowed (e.g., 16ms) angles as features to classify eye movements. For their discussion they also calculated the angles between the first and last sample of each classified event. They showed that 16ms sample directions (angles) of smooth pursuits deviated less from the overall event directions than fixations (Startsev et al., 2019). Figure 24 shows a screenshot of their findings.

(Startsev et al., 2019) calculated the direction deviation after classifying eye movements because it relies on the direction of classified events. However, if the event direction and direction deviation could be accurately estimated before classification, it could potentially distinguish fixations and smooth pursuits better than the features used by gazeHMM.

Figure 24

Direction Deviation from Startsev et al. (2018, p. 565)



Note. This is a screenshot taken from Startsev et al. (2018, p. 565). It illustrates that direction deviation distributions differ for fixations (red) and smooth pursuits (blue).

Direction Deviation

Method

I calculated direction deviations to see if I could reproduce the distributions reported by Startsev et al. (2019) in the benchmark data set. Therefore, I calculated the 16ms direction (angle) of each sample in the training set and the direction of each event according to the ground truth. Then I subtracted the 16ms direction of each sample from their respective event direction to calculate the direction deviation. Finally, I plotted distributions for saccades, fixations, and smooth pursuits and compared them with the distribution in Figure 24. Like for the gazeHMM features I created plots for across all speeds and separately for each speed, for the entire training set and each individual participant.

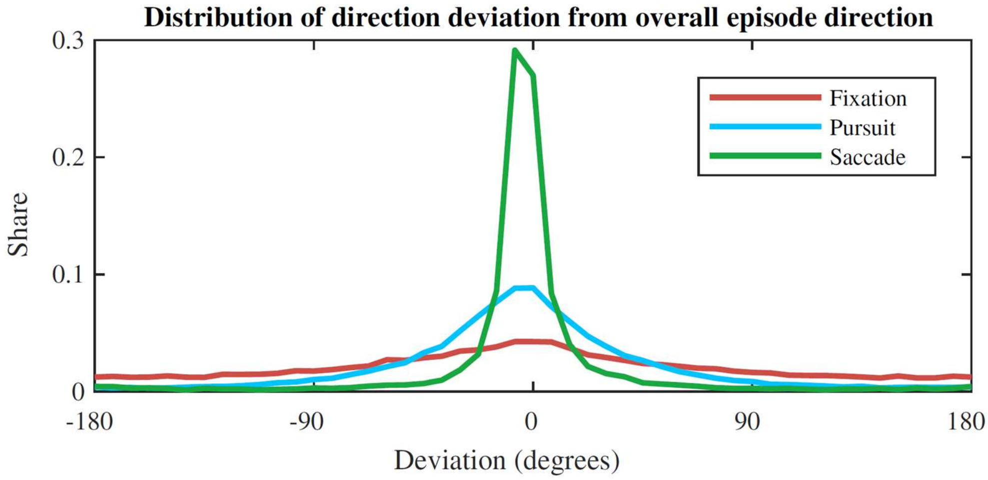
The analysis was conducted in R (R Core Team, 2022) and Python (Van Rossum & Drake Jr, 1995). See Appendix A and B for a list of the most central packages used.

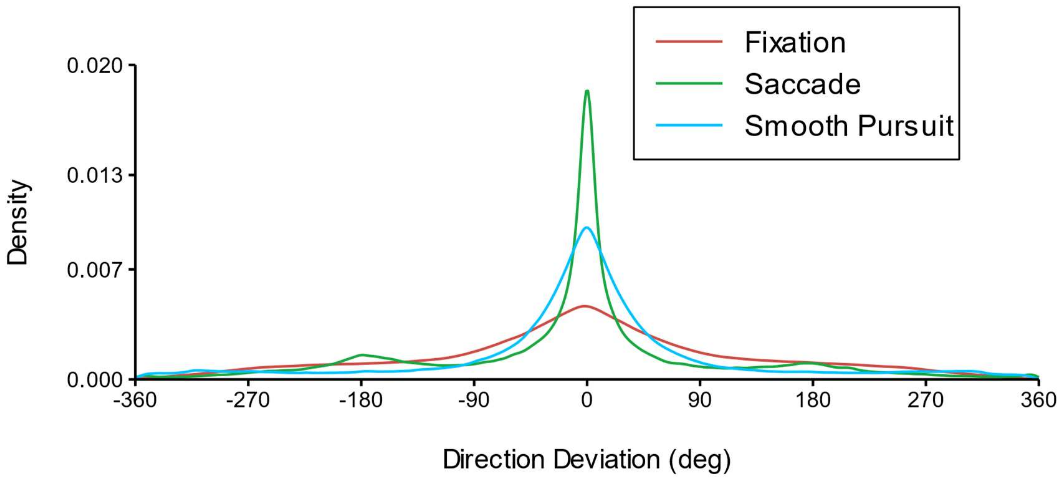
Results

Exploratory analysis. I found direction deviation distributions that are remarkably similar to those reported by (Startsev et al., 2019) for all participants in the training set (see Figure 18). Specifically, saccades display the lowest, and fixations the largest deviations. Consistent with my findings on the gazeHMM features, the distributions of fixations and smooth pursuits become more distinct at higher target speeds (see Figure 19). Importantly, in contrast to gazeHMM features, the distributions are also distinct for slow targets (i.e., 1°/s) for all participants in the training set.

Figure 25

Startsev et al (2018) and Current Direction Deviation Distributions Across All Trials and All Participants

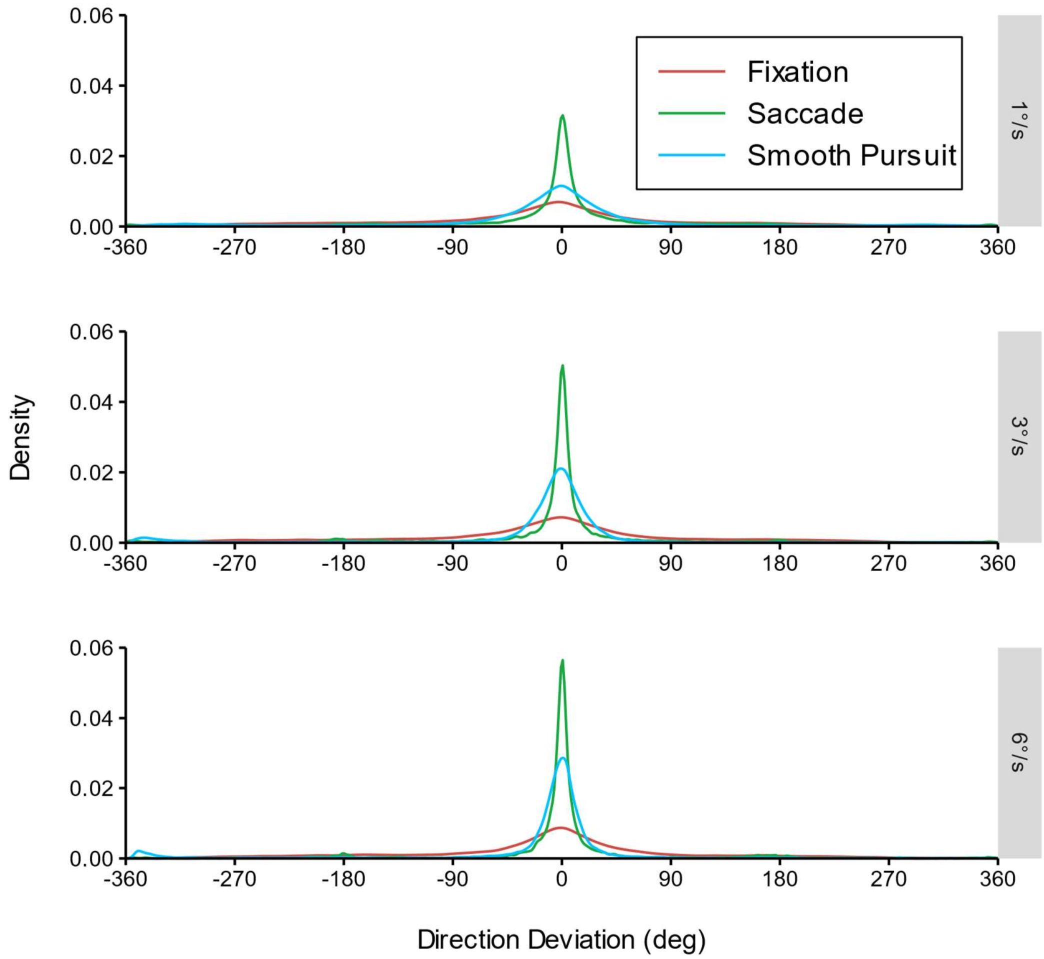




Note. These plots show direction deviations reported by Startsev et al. (2018; top) and direction deviation densities for saccades (green), fixations (red), and smooth pursuits (blue) for all 1008 trials included in the training set (bottom). The distributions of the three eye movements in the training set are clearly distinct and resemble those that were found by Startsev et al. (2018).

Figure 26

Direction Deviation Distributions at Different Target Speeds for All Participants



Note. These plots show direction deviation densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 336 trials of a given speed included in the training set. Distributions for fixations and smooth pursuits are distinct at all target speeds and become more distinct for faster targets.

Discussion

Direction deviation distributions in the benchmark data set are remarkably similar to those reported by Startsev et al. (2019) for the GazeCom data set. GazeCom consists of naturalistic eye movements in response to naturalistic videos and is based on a human annotated ground truth (Dorr et al., 2010; Startsev et al., 2016). In contrast, the current benchmark data set consists of somewhat unnatural eye movements (e.g., long fixations, clear smooth pursuits along straight trajectories) in response to synthetic stimuli and is based on a ground truth that did not rely on human annotation. Thus, finding distributions this similar suggests that direction deviation differences between fixations and smooth pursuits generalize to different contexts and are a promising candidate to distinguish fixations and smooth pursuits. Additionally, finding distinct distributions is especially promising because slow target speeds, which were least distinct for the gazeHMM features, are heavily overrepresented in the current data set. Finding clearly distinct distributions despite this, suggests that the direction deviation can also distinguish between fixations and slow smooth pursuits.

Finding clear differences between the fixations and smooth pursuits classified for the baseline also provides further evidence that ground truth labeling during preprocessing was accurate. This is evidence that current data set provides a good benchmark without human annotation and that my findings about gazeHMM features are valid.

Finally, these findings support my conclusion that gazeHMM features are inadequate to distinguish fixations and smooth pursuits. Therefore, I tried to estimate the direction deviation before classification in the next section to develop a new feature for event classification.

Estimated Direction Deviation (EED)

Calculating the direction deviation requires event directions which in turn are dependent on events that have already been classified. Therefore, I tried to estimate the event direction before classification to be able to calculate an *estimated direction deviation* (EDD).

Method

To estimate the event direction, I calculated the direction in 30 windows (10ms - 300ms; 10ms step size) by calculating the angle from the first to the last sample in each window for each sample. I then took the mean (*mean direction approach*) and median (*median direction approach*) as estimates for the actual event direction. Finally, I subtracted the 16ms direction of each sample from the two estimates to calculate two versions of the EDD.

To evaluate which estimation procedure to base my analysis on, I calculated the mean squared error (MSE) between the estimated event direction and actual event direction for both procedures. Additionally, I calculated the MSE for the EDD and actual direction deviation for both procedures with the intention to use the approach with the lower MSE.

Finally, I plotted EED distributions for the selected approach across all trials and separately for each target speed, for the entire training set and for each participant individually.

The analysis was conducted in R (R Core Team, 2022) and Python (Van Rossum & Drake Jr, 1995). See Appendix A and B for a list of the most central packages used.

Results

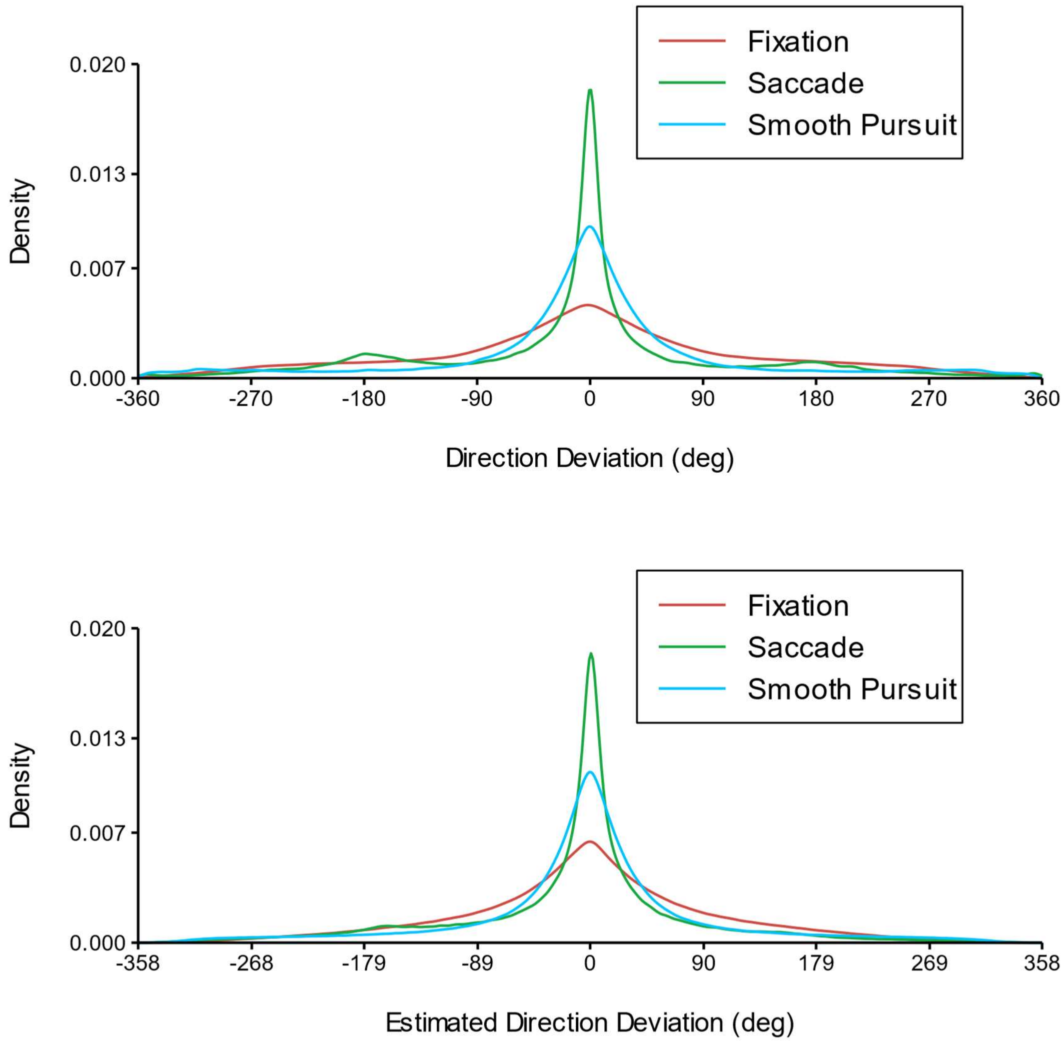
Exploratory analysis. The MSE for the estimated event direction from the actual direction deviation was lower for the *mean direction approach* (MSE = 2.33) than the *median direction approach* (MSE = 2.869). Similarly, the MSE of the EED from the actual direction deviation was lower for the *mean direction approach* (MSE = 2.324) than for the *median direction approach* (MSE = 2.863). Therefore, I analyzed the feature distributions of the EED based on the *mean direction approach*.

Figure 20 shows that the EED distributions closely resemble those of the actual direction deviation in the benchmark data set. Its distribution is clearly distinct for fixations and smooth pursuits for targets moving 3°/s and 6°/s. However, for targets moving 1°/s, distributions for fixations and smooth pursuits were only distinct for four out of seven of the participants (e.g., 06b8d2d3; see Figure

18) in the training set. For the remaining three participants (e.g., 7d248f8f; see Figure 19) they were not distinct for these slow smooth pursuits.

Figure 27

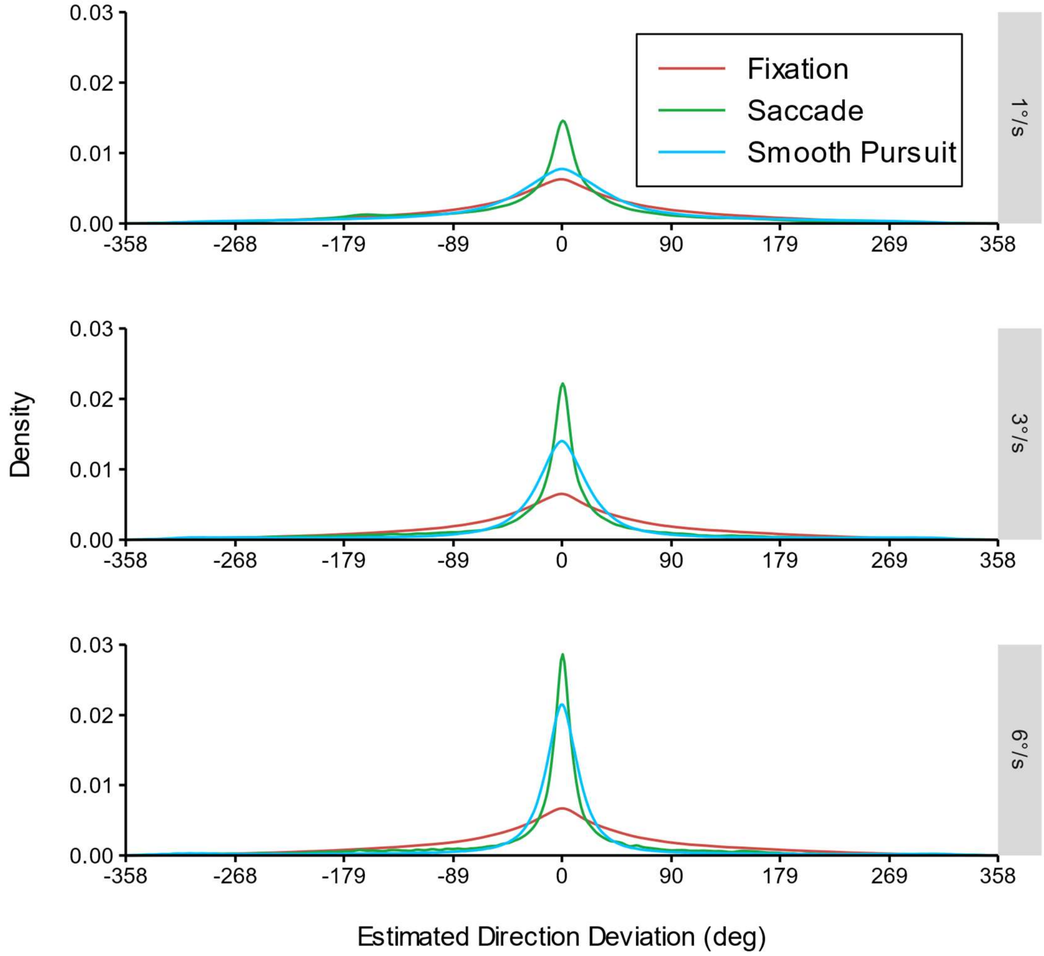
Actual and Estimated Direction Deviation Distributions Across All Trials and All Participants



Note. These plots show the direction deviation (top) and estimated direction deviation (bottom) densities for saccades (green), fixations (red), and smooth pursuits (blue) for all 1008 trials included in the training set. The estimated direction deviation clearly resembles the actual direction deviation and distribution for fixations and smooth pursuits are clearly distinct.

Figure 28

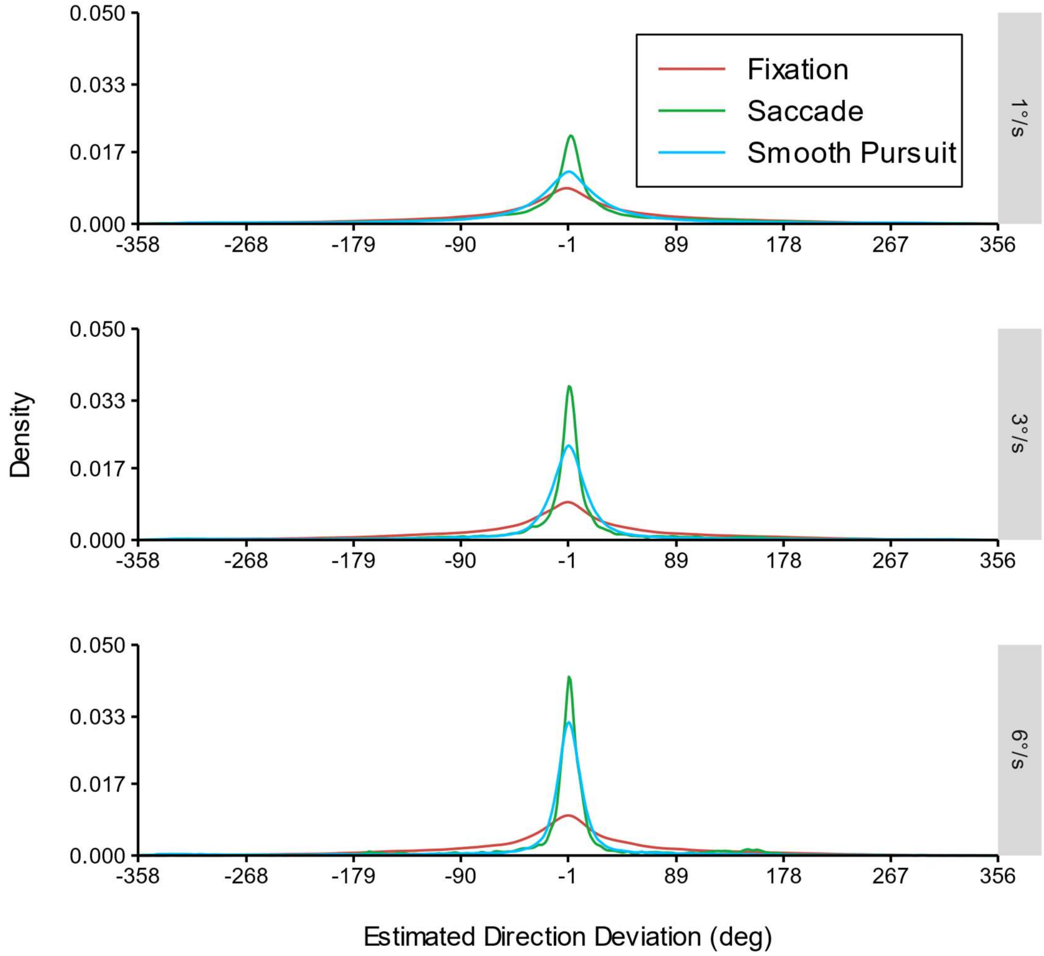
EDD Distributions at Different Target Speeds for All Participants



Note. These plots show estimated direction deviation densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 336 trials of a given speed included in the training set. Distributions resemble the actual direction deviations in Figure 26. They are distinct for fixations and smooth pursuits at all target speeds and become more distinct for faster targets.

Figure 29

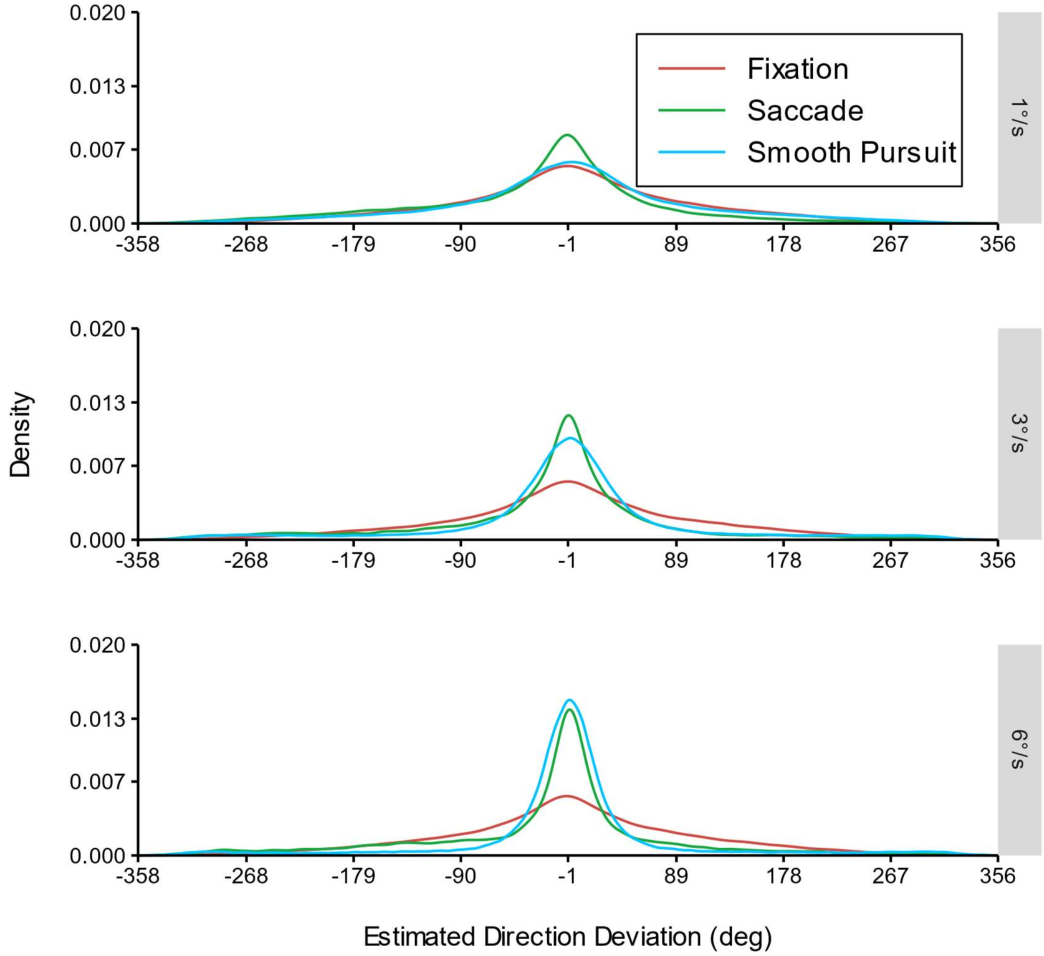
EDD Distributions at Different Target Speeds for Participant 06b8d2d3



Note. These plots show estimated direction deviation densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 48 trials of a given speed of participant 06b8d2d3. Distributions for fixations and smooth pursuits are distinct at all target speeds and become clearly more distinct for faster targets for this participant.

Figure 30

EDD Distributions at Different Target Speeds for Participant 7d248f8f



Note. These plots show estimated direction deviation densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 48 trials of a given speed of participant 7d248f8f. Distributions for fixations and smooth pursuits at low target speeds are very similar and become more distinct for faster targets for this participant.

Discussion

The *estimated direction deviation* closely resembled the actual direction deviation for all features. Importantly, fixations and smooth pursuits were distinct at all target speeds for most participants. Additionally, the EDD was more distinct for fixations and smooth pursuits than any gazeHMM features for all these cases. This suggests that the EDD is a promising feature and could replace the sample-to-sample angle feature in gazeHMM. To test its actual capability to distinguish fixations and smooth pursuits, I want to implement it in gazeHMM in a future study.

One limitation is that distributions for fixations and slow smooth pursuits (i.e., 1°/s) were not distinct for two participants in the training set. Interestingly, these participants were the same that had the most similar distributions for the gazeHMM features. This indicates that fixations and smooth pursuits are more difficult to distinguish for some people regardless of which feature is being used. Another explanation could be that these participants were unable to perform consistent smooth pursuits at 1°/s without intermittent fixations. My own experience during stimulus design and evidence by Drewes et al. (2018) suggest that smooth pursuits turn into fixations for targets moving below 1°/s. However, Meyer et al. (1985) found that the upper limit for fast smooth pursuits differs between people and it seems likely that the same would be true for lower limits. Future studies are therefore needed, to investigate if the direction deviation was not distinct for some participants because they made more intermittent fixations or if there are other explanations for this inconsistency.

Another limitation concerns the generalizability of my approach to more naturalistic eye movements. While I was able to accurately estimate event directions using 30 windows from 10ms to 300ms, it is questionable whether the same window sizes generalize well to data with shorter fixations and smooth pursuits. Nevertheless, findings from past studies indicate that larger window sizes are often preferred to detect smooth pursuits (Komogortsev & Karpov, 2013; Startsev et al., 2019; Vidal et al., 2012). Additionally, it would be easy to estimate the event direction of shorter events by using smaller window- and step sizes. Finally, the EDD distributions resembled the direction deviation distributions found by Startsev et al. (2019) on naturalistic data which suggests the estimation procedure could generalize. Nevertheless, future studies should investigate how the number and size of windows affects the accuracy of the EDD in different data sets.

In conclusion, the EDD is a promising feature to distinguish fixations and smooth pursuits and is distributed much more distinct for these eye movements than any of the gazeHMM features. Nevertheless, its distributions are not distinct for slow target speeds of some participants and the distributions all have the same location and are only distinct in terms of their spread. To see, if I could create an even more promising candidate, I developed the estimated direction deviation spread (EDD-S) feature in the next section.

Estimated Direction Deviation Spread (EED-S)

Even though the EED distributions for fixations and smooth pursuits are distinct in most cases, their distributions have similar shapes with peaks at zero degrees. Consequently, they are mainly distinct in terms of their spread. Following this logic, I came up with the idea to quantify the spread of the estimated direction deviation instead, hoping this would result in distributions that are even more distinct for fixations and smooth pursuits.

Method

To calculate the estimated direction deviation spread (EDD-S), I first calculated the same 30 sample directions like for the EDD (i.e., 10ms – 300ms; 10ms step size). However, instead of taking the mean to estimate the event direction, I subtracted the 16ms direction from each of the windowed directions. Therefore, I had multiple windowed direction deviations for each sample. Finally, I calculated the EED-S as the standard deviation of these windowed direction deviations for each sample.

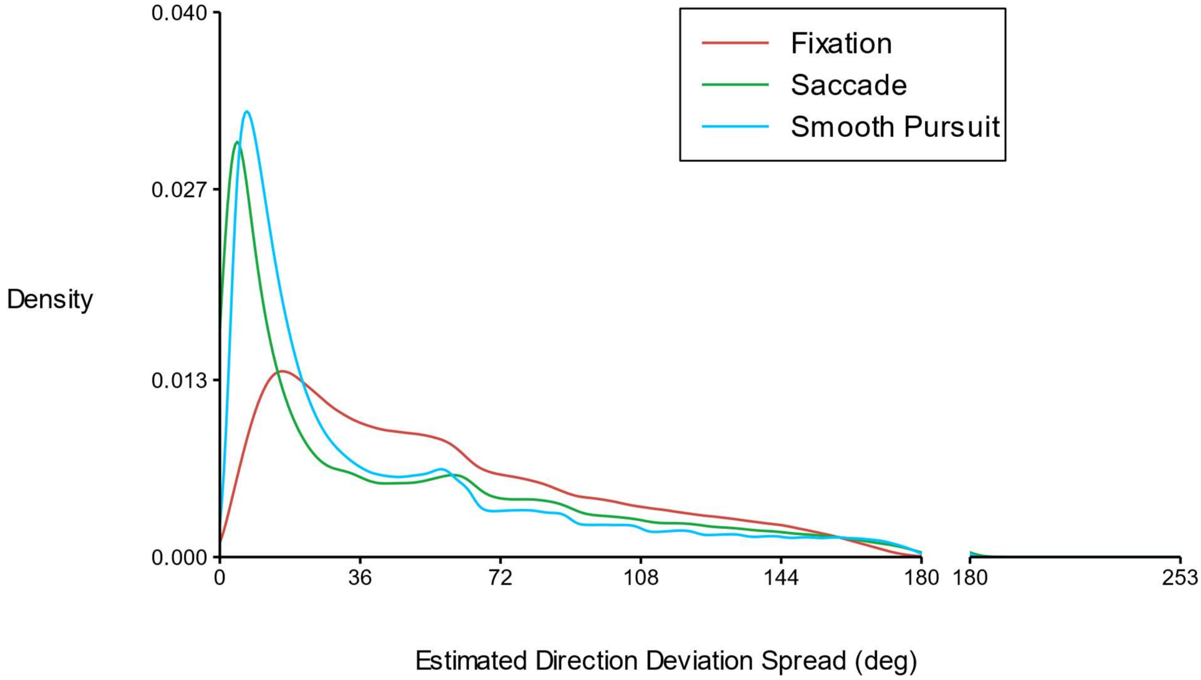
Like in previous sections, I plotted the EDD-S across all speeds, and separately for each speed, for the entire training set and each individual participant.

Results

Exploratory analysis. Figure 31 shows that the EED-S distributions for fixations and smooth pursuits are clearly distinct. Like the gazeHMM features and the EDD, they become more distinct for faster targets (see Figure 32). For the individual participants, the overall pattern is like the EED. Specifically, for most participants in the training set (e.g., 06b8d2d3; see Figure 33*)* the EDD-S is distinct at all target speeds, but for others (e.g., 7d248f8f; see Figure 34) the distributions at 1°/s are very similar. Nevertheless, while the EDD distributions for slow targets looked almost identical for the latter (see Figure 30), the EDD-S is markedly flatter for fixations and at least somewhat different.

Figure 31

EDD-S Distributions Across All Trials and All Participants

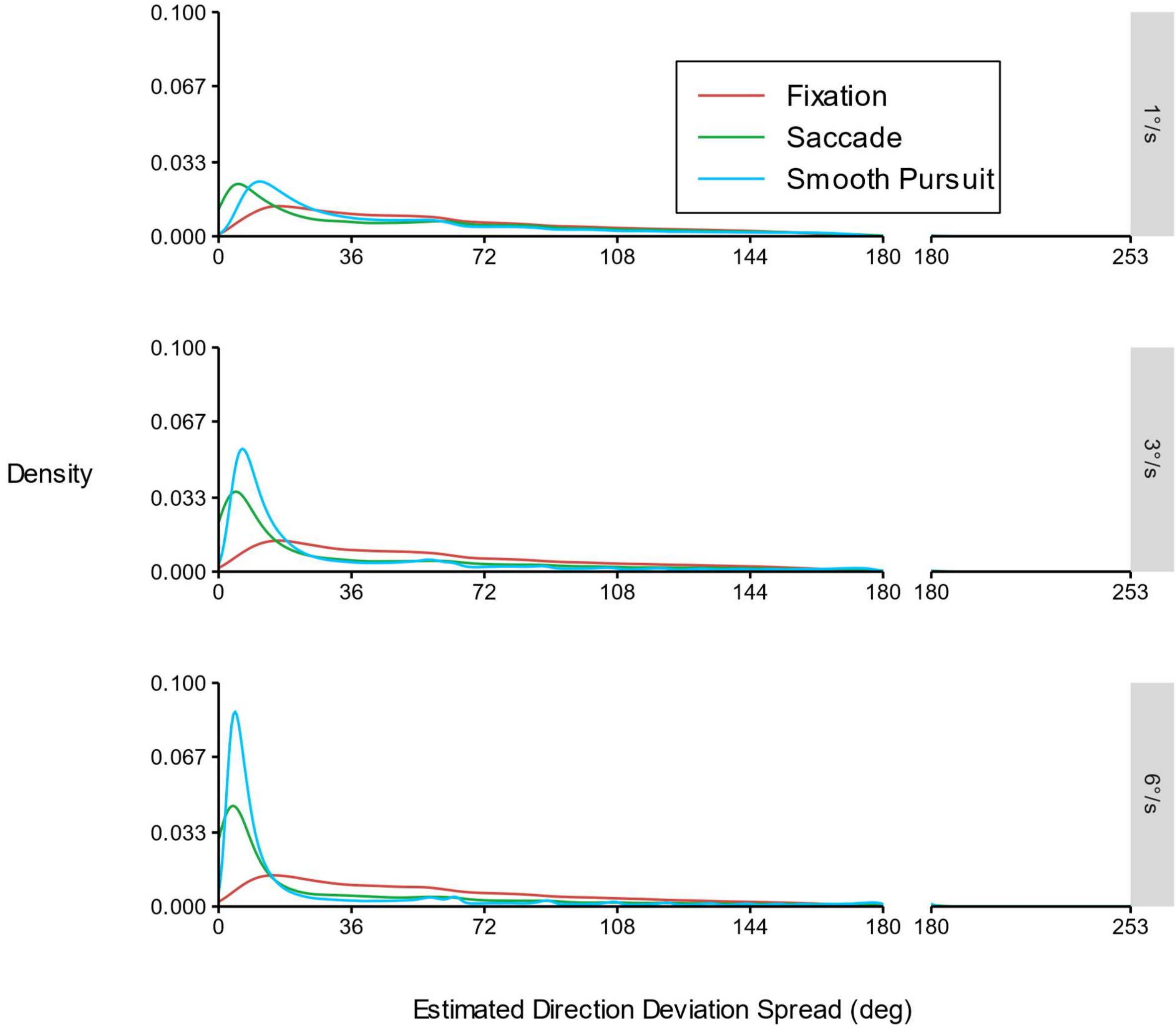


Note. This plot shows estimated direction deviation spread densities for saccades (green), fixations (red), and smooth pursuits (blue) for all 1008 trials included in the training set. The distribution of fixations is clearly distinct from those of saccades and fixations.

The x axis is broken to focus the plot on the area with the highest densities.

Figure 32

EDD-S Distributions at Different Target Speeds for All Participants

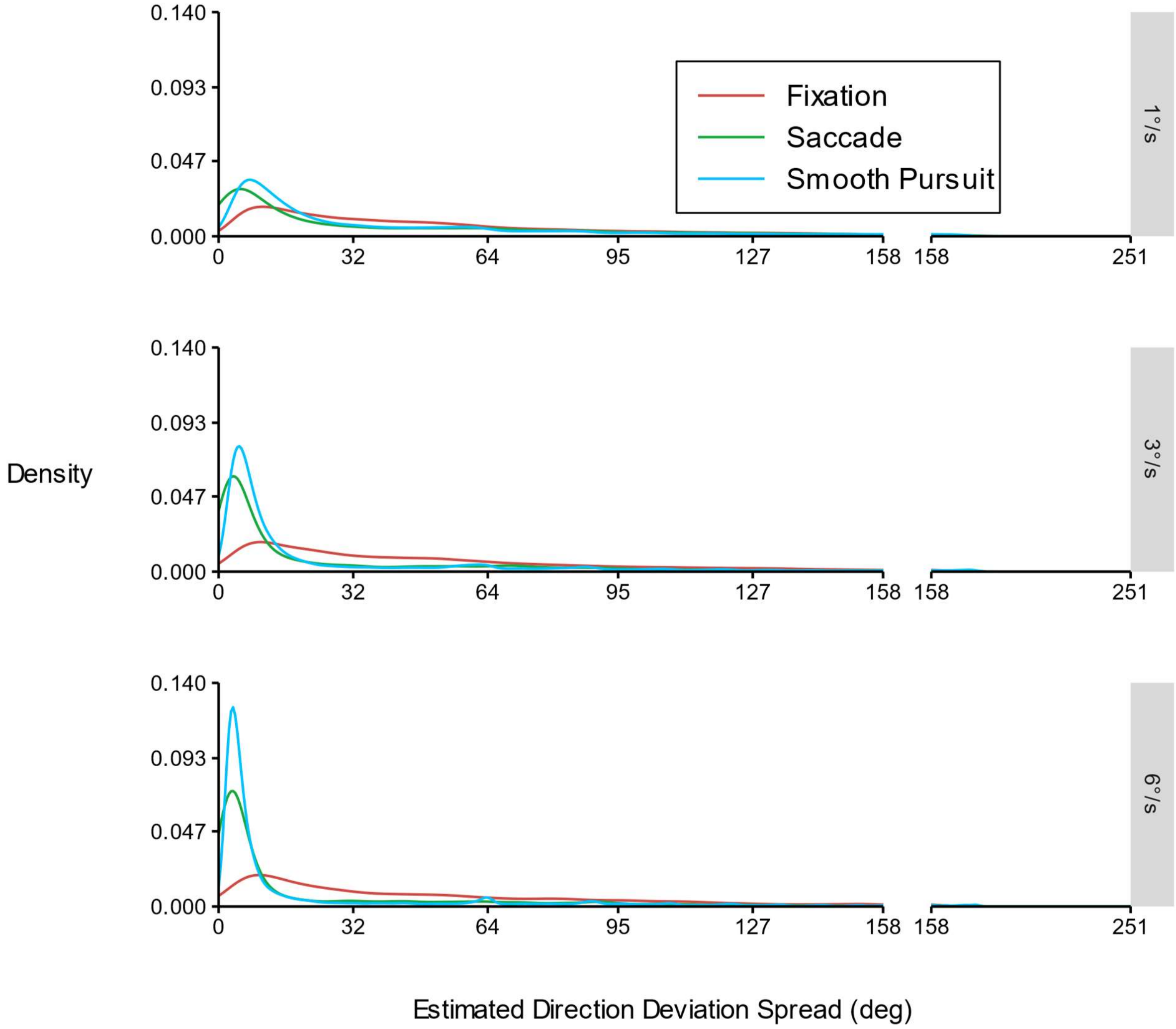


Note. These plots show estimated direction deviation spread densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 336 trials of a given speed included in the training set. Distributions for fixations are clearly distinct from those of saccades and smooth pursuits for all target speeds.

The x axes are broken to focus the plots on the area with the highest densities.

Figure 33

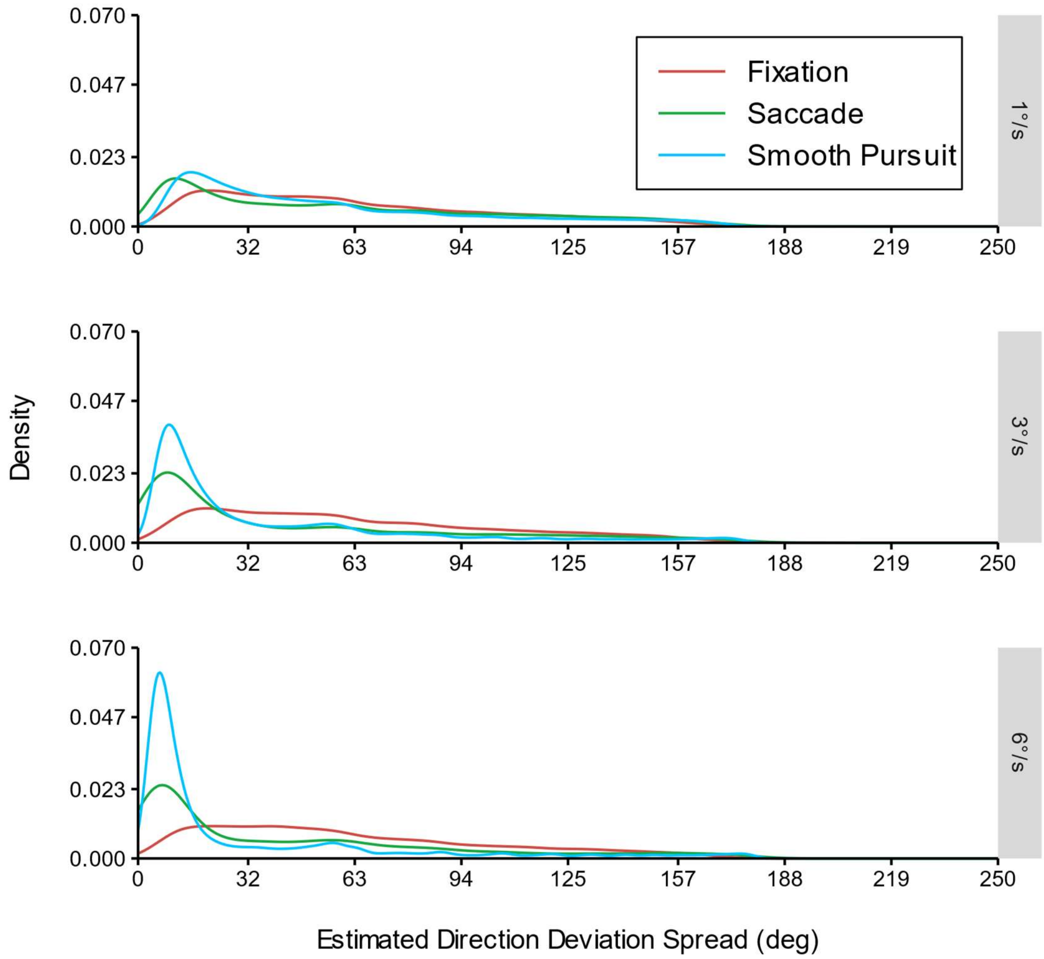
EDD-S Distributions at Different Target Speeds for Participant 06b8d2d3



Note. These plots show estimated direction deviation spread densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 48 trials of a given speed of participant 06b8d2d3. Distributions for fixations and smooth pursuits are distinct at all target speeds and become clearly more distinct for faster targets for this participant. The x axes are broken to focus the plots on the area with the highest densities.

Figure 34

EDD-S Distributions at Different Target Speeds for Participant 7d248f8f



Note. These plots show estimated direction deviation spread densities for fixations (red), and smooth pursuits (blue) for targets moving 1°/s (top), 3°/s (middle), and 6°/s (bottom). They are based on all 48 trials of a given speed of participant 7d248f8f. Distributions for fixations and smooth pursuits at low target speeds are distinct but relatively similar, and they become more distinct for faster targets for this participant.

Discussion

Like the EDD, the EDD-S is clearly distinct for fixations and smooth pursuits. Importantly, the EDD-S is at least slightly distinct for all participants including those who had very similar EDD distributions for targets moving 1°/s. Additionally, the EDD-S is the only feature investigated during this study that distinguishes fixations and smooth pursuits in terms of distribution location and not only distribution spread.

Together, these findings make the EDD-S an even more promising candidate to distinguish fixations and smooth pursuits. Therefore, I want I want to implement them in gazeHMM in a future study and evaluate if this would increase gazeHMM performance.

The same limitations with regards to generalizability and window sizes discussed for the EDD apply to the EDD-S and should be investigated in future studies.

General Discussion

Fixations and smooth pursuits are frequently confused by human annotators and event classification algorithms (Andersson et al., 2017; Komogortsev & Karpov, 2013). Additionally, human annotations do not necessarily represent a gold standard of eye movement classification (Andersson et al., 2017; Hooge et al., 2018). To address both shortcomings, I created a benchmark data set without relying on human annotation and investigated the feature characteristics of fixations and smooth pursuits.

The data set consists of almost four hours of eye movements by ten participants and results indicate that ground truth labels are relatively accurate. Furthermore, results suggest that algorithms like gazeHMM (Lüken et al., 2022) fail to distinguish between fixations and smooth pursuits because their feature characteristics overlap. Specifically, the velocity, acceleration, and sample-to-sample angles of these eye movements were distributed very similarly. To solve this problem, I developed two new eye movement features based on findings of Startsev et al. (2019). The *estimated direction deviation* (EDD) and *estimated direction deviation spread* (EDD-S) show that sample directions within smooth pursuits deviate less from the overall direction than those within fixations. In contrast to the gazeHMM features, their distributions are very distinct for fixations and smooth pursuits in the current data set. Therefore, they represent promising candidates to improve automatic eye movement classification in the future.

Despite avoiding manual annotation, ground truth labels in the current data set are not entirely free from human bias because the velocity threshold to separate fast and slow eye movements was manually finetuned. Additionally, some fixations were likely misclassified as smooth pursuits because it was impossible to fully prevent them from cooccurring. Nevertheless, the different results of this study consistently indicate that the current data set represents a successful first attempt at creating a benchmark data set without human annotation.

Importantly, because of researcher error, eye positions and target positions were shifted by 100ms, and some target positions were incorrectly interpolated. Deviations were small and likely did not influence results. Nevertheless, this needs to be corrected and the current analysis needs to be repeated before conducting further investigations.

Afterwards, I want to implement the EDD and EDD-S in gazeHMM and test if this would improve its performance on the current data set, and the Andersson et al. (2017) dataset.

Additionally, future studies should further investigate the benchmark data set and evaluate different preprocessing procedures to improve its ground truth labels and reduce human influence further. As an initial step the evaluation of different algorithms done by (Andersson et al., 2017) could be repeated with the current benchmark data set.

Finally, more eye movement features should be investigated using the benchmark data set. Specifically, future studies could investigate the distributions of features used by other authors to classify smooth pursuits or develop entirely new features like the EDD and EDD-S. Until someday a published paper titles “automatic event classification algorithm accurately distinguishes fixations and smooth pursuits”.

References

Agtzidis, I., Startsev, M., & Dorr, M. (2016). Smooth pursuit detection based on multiple observers. *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, 303–306. https://doi.org/10.1145/2857491.2857521

Agtzidis, I., Startsev, M., & Dorr, M. (2019). 360-degree Video Gaze Behaviour: A Ground-Truth Data Set and a Classification Algorithm for Eye Movements. *Proceedings of the 27th ACM International Conference on Multimedia*, 1007–1015. https://doi.org/10.1145/3343031.3350947

Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (2017). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods*, *49*(2), 616–637. https://doi.org/10.3758/s13428-016-0738-9

Armstrong, T., & Olatunji, B. O. (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clinical Psychology Review*, *32*(8), 704–723. https://doi.org/10.1016/j.cpr.2012.09.004

Berg, D. J., Boehnke, S. E., Marino, R. A., Munoz, D. P., & Itti, L. (2009). Free viewing of dynamic stimuli by humans and monkeys. *Journal of Vision*, *9*(5), 19. https://doi.org/10.1167/9.5.19

Braunagel, C., Kasneci, E., Stolzmann, W., & Rosenstiel, W. (2015). Driver-Activity Recognition in the Context of Conditionally Autonomous Driving. *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 1652–1657. https://doi.org/10.1109/ITSC.2015.268

Chen, C.-Y., & Hafed, Z. M. (2013). Postmicrosaccadic Enhancement of Slow Eye Movements. *Journal of Neuroscience*, *33*(12), 5375–5386. https://doi.org/10.1523/JNEUROSCI.3703-12.2013

Dar, A. H., Wagner, A. S., & Hanke, M. (2020). *REMoDNaV: Robust Eye-Movement Classification for Dynamic Stimulation* (p. 619254). bioRxiv. https://doi.org/10.1101/619254

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 28. https://doi.org/10.1167/10.10.28

Drewes, H., Khamis, M., & Alt, F. (2018). Smooth Pursuit Target Speeds and Trajectories. *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, 139–146. https://doi.org/10.1145/3282894.3282913

Duchowski, A. T. (2017). Taxonomy and Models of Eye Movements. In A. T. Duchowski (Ed.), *Eye Tracking Methodology: Theory and Practice* (pp. 39–46). Springer International Publishing. https://doi.org/10.1007/978-3-319-57883-5_4

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hein, O., & Zangemeister, W. (2017). Topology for gaze analyses—Raw data segmentation. *Journal of Eye Movement Research*, *10*(1), 10.16910/jemr.11.1.2. https://doi.org/10.16910/jemr.10.1.1

Hershman, R., Henik, A., & Cohen, N. (2018). A novel blink detection method based on pupillometry noise. *Behavior Research Methods*, *50*(1), 107–114. https://doi.org/10.3758/s13428-017-1008-1

Hooge, I. T. C., Niehorster, D. C., Nyström, M., Andersson, R., & Hessels, R. S. (2018). Is human classification by experienced untrained observers a gold standard in fixation detection? *Behavior Research Methods*, *50*(5), 1864–1881. https://doi.org/10.3758/s13428-017-0955-x

Hunfalvay, M., Roberts, C.-M., Murray, N. P., Tyagi, A., Barclay, K. W., Bolte, T., Kelly, H., & Carrick, F. R. (2020). Vertical smooth pursuit as a diagnostic marker of traumatic brain injury. *Concussion*, *5*(1), CNC69. https://doi.org/10.2217/cnc-2019-0013

Khamis, M., Oechsner, C., Alt, F., & Bulling, A. (2018). VRpursuits: Interaction in virtual reality using smooth pursuit eye movements. *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, 1–8. https://doi.org/10.1145/3206505.3206522

Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D. H., & Gowda, S. M. (2010). Standardization of Automated Analyses of Oculomotor Fixation and Saccadic Behaviors. *IEEE Transactions on Biomedical Engineering*, *57*(11), 2635–2645. https://doi.org/10.1109/TBME.2010.2057429

Komogortsev, O. V., & Karpov, A. (2013). Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. *Behavior Research Methods*, *45*(1), 203–215. https://doi.org/10.3758/s13428-012-0234-9

Krauzlis, R. J., Goffart, L., & Hafed, Z. M. (2017). Neuronal control of fixation and fixational eye movements. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1718), 20160205. https://doi.org/10.1098/rstb.2016.0205

Kucharský, Š., Visser, I., Truțescu, G.-O., Laurence, P. G., Zaharieva, M., & Raijmakers, M. E. J. (2020). Cognitive strategies revealed by clustering eye movement transitions. *Journal of Eye Movement Research*, *13*(1), 10.16910/jemr.13.1.1. https://doi.org/10.16910/jemr.13.1.1

Lappi, O. (2016). Eye movements in the wild: Oculomotor control, gaze behavior & frames of reference. *Neuroscience and Biobehavioral Reviews*, *69*, 49–68. https://doi.org/10.1016/j.neubiorev.2016.06.006

Larsson, L., Nyström, M., Andersson, R., & Stridh, M. (2015). Detection of fixations and smooth pursuit movements in high-speed eye-tracking data. *Biomedical Signal Processing and Control*, *18*, 145–152. https://doi.org/10.1016/j.bspc.2014.12.008

Larsson, L., Nyström, M., & Stridh, M. (2013). Detection of Saccades and Postsaccadic Oscillations in the Presence of Smooth Pursuit. *IEEE Transactions on Biomedical Engineering*, *60*(9), 2484–2493. https://doi.org/10.1109/TBME.2013.2258918

Lee, W. J., Kim, J. H., Shin, Y. U., Hwang, S., & Lim, H. W. (2019). Differences in eye movement range based on age and gaze direction. *Eye*, *33*(7), Article 7. https://doi.org/10.1038/s41433-019-0376-4

Lin, Y.-C., Intoy, J., Clark, A. M., Rucci, M., & Victor, J. D. (2023). Cognitive influences on fixational eye movements. *Current Biology*, *33*(8), 1606-1612.e4. https://doi.org/10.1016/j.cub.2023.03.026

Lisberger, S. G., Evinger, C., Johanson, G. W., & Fuchs, A. F. (1981). Relationship between eye acceleration and retinal image velocity during foveal smooth pursuit in man and monkey. *Journal of Neurophysiology*, *46*(2), 229–249. https://doi.org/10.1152/jn.1981.46.2.229

Lüken, M. (2020). *GazeHMM: Classify Gaze Data With a Hidden Markov Model* (0.1.0) [Computer software]. https://github.com/maltelueken/gazeHMM/tree/master

Lüken, M., & Kucharský, Š. (2023). *gazeHMM@ntimes: Classify Gaze Data With a Hidden Markov Model* (0.1.1) [Computer software]. https://github.com/Kucharssim/gazeHMM

Lüken, M., Kucharský, Š., & Visser, I. (2022). Characterising eye movement events with an unsupervised hidden markov model. *Journal of Eye Movement Research*, *15*(1), Article 1. https://doi.org/10.16910/jemr.15.1.4

MacAskill, M. R., & Anderson, T. J. (2016). Eye movements in neurodegenerative diseases. *Current Opinion in Neurology*, *29*(1), 61. https://doi.org/10.1097/WCO.0000000000000274

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, *5*(3), Article 3. https://doi.org/10.1038/nrn1348

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56–61). https://doi.org/10.25080/Majora-92bf1922-00a

Meyer, C. H., Lasker, A. G., & Robinson, D. A. (1985). The upper limit of human smooth pursuit velocity. *Vision Research*, *25*(4), 561–563. https://doi.org/10.1016/0042-6989(85)90160-9

O'Driscoll, G. A., & Callahan, B. L. (2008). Smooth pursuit in schizophrenia: A meta-analytic review of research since 1993. *Brain and Cognition*, *68*(3), 359–370. https://doi.org/10.1016/j.bandc.2008.08.023

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, 141–144. https://doi.org/10.1145/1743666.1743701

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pekkanen, J., & Lappi, O. (2017). A new and general approach to signal denoising and eye movement classification based on segmented linear regression. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-17983-x

Pinkhardt, E. H., & Kassubek, J. (2011). Ocular motor abnormalities in Parkinsonian syndromes. *Parkinsonism & Related Disorders*, *17*(4), 223–230. https://doi.org/10.1016/j.parkreldis.2010.08.004

Plotly TechnologiesInc. (2015). *Collaborative data science*. Plotly Technologies Inc. https://plot.ly

Poole, A., & Ball, L. J. (2006). *Eye Tracking in HCI and Usability Research* [Chapter]. Encyclopedia of Human Computer Interaction; IGI Global. https://doi.org/10.4018/978-1-59140-562-7.ch034

R Core Team. (2022). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rigas, I., Friedman, L., & Komogortsev, O. (2018). Study of an Extensive Set of Eye Movement Features: Extraction Methods and Statistical Analysis. *Journal of Eye Movement Research*, *11*(1), 10.16910/jemr.11.1.3. https://doi.org/10.16910/jemr.11.1.3

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, 71–78. https://doi.org/10.1145/355017.355028

Santini, T., Fuhl, W., Kübler, T., & Kasneci, E. (2016). Bayesian identification of fixations, saccades, and smooth pursuits. *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, 163–170. https://doi.org/10.1145/2857491.2857512

Schafer, R. W. (2011). What Is a Savitzky-Golay Filter? [Lecture Notes]. *IEEE Signal Processing Magazine*, *28*(4), 111–117. https://doi.org/10.1109/MSP.2011.941097

Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., & Willemsen, M. C. (2017). Process-Tracing Methods in Decision Making: On Growing Up in the 70s. *Current Directions in Psychological Science*, *26*(5), 442–450. https://doi.org/10.1177/0963721417708229

SR Research. (n.d.). *EyeLink 1000 Plus*. Fast, Accurate, Reliable Eye Tracking. Retrieved 27 July 2023, from https://www.sr-research.com/eyelink-1000-plus/

SR Research. (2020, August 27). *Getting Started with Python and PyLink*. SR Research Support. https://www.sr-research.com/support/thread-48.html

Startsev, M., Agtzidis, I., & Dorr, M. (2019). 1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits. *Behavior Research Methods*, *51*(2), 556–572. https://doi.org/10.3758/s13428-018-1144-2

Startsev, M., Agtzidis, I., & Dorr, M. (2016). *Smooth Pursuit*. https://michaeldorr.de/smoothpursuit/

Startsev, M., & Zemblys, R. (2022). Evaluating Eye Movement Event Detection: A Review of the State of the Art. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01763-7

Tao, L., Wang, Q., Liu, D., Wang, J., Zhu, Z., & Feng, L. (2020). Eye tracking metrics to screen and assess cognitive impairment in patients with neurological disorders. *Neurological Sciences*, *41*(7), 1697–1704. https://doi.org/10.1007/s10072-020-04310-y

Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Verdonck, T., Baesens, B., Óskarsdóttir, M., & vanden Broucke, S. (2021). Special issue on feature engineering editorial. *Machine Learning*. https://doi.org/10.1007/s10994-021-06042-2

Vidal, M., Bulling, A., & Gellersen, H. (2011). Analysing EOG signal features for the discrimination of eye movements with wearable devices. *PETMEI'11 - Proceedings of the 1st International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction*. https://doi.org/10.1145/2029956.2029962

Vidal, M., Bulling, A., & Gellersen, H. (2012). Detection of smooth pursuits using eye movement shape features. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 177–180. https://doi.org/10.1145/2168556.2168586

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Visser, I., & Speekenbrink, M. (2022). *Mixture and Hidden Markov Models with R*. Springer International Publishing. https://doi.org/10.1007/978-3-031-01440-6

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation* (1.1.2) [Computer software]. https://CRAN.R-project.org/package=dplyr

Xu, S., Chen, M., Feng, T., Zhan, L., Zhou, L., & Yu, G. (2021). Use ggbreak to Effectively Utilize Plotting Space to Deal With Large Datasets and Outliers. *Frontiers in Genetics*, *12*. https://www.frontiersin.org/articles/10.3389/fgene.2021.774846

Appendix

Appendix A

List of Central R Packages and Their Respective Functions

- gazeHMM (Lüken, 2020): original gazeHMM fitted to single trials
- gazeHMM@ntimes (Lüken & Kucharský, 2023): gazeHMM fork to fit multiple trials
- dplyr (Wickham et al., 2023): manipulation and filtering of data frames
- ggplot2 (Wickham, 2016): plots
- ggbreak (Xu et al., 2021): break in x-axis for plots

Appendix B

List of Central Python Packages and Their Respective Functions

- PsychoPy (Peirce et al., 2019): programming of the study for data collection
- PyLink (SR Research, 2020): using EyeLink with PsychoPy
- Numpy (Harris et al., 2020): data structures for statistical analysis in python
- Pandas (McKinney, 2010): data structures for statistical analysis in python
- Plotly (Plotly TechnologiesInc, 2015): plots
- SciPy (Virtanen et al., 2020): savitzky-golay-filter to calculate velocity, and acceleration