Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades

Oleg V. Komogortsev ·
Alex Karpov

Abstract Ternary eye movement classification, which separates fixations, saccades, and smooth pursuit from the raw eye positional data, is extremely challenging. This article develops new and modifies existing eye-tracking algorithms for the purpose of conducting meaningful ternary classification. To this end, a set of qualitative and quantitative behavior scores is introduced to facilitate the assessment of classification performance and to provide means for automated threshold selection. Experimental evaluation of the proposed methods is conducted using eye movement records obtained from 11 subjects at 1000 Hz in response to a step-ramp stimulus eliciting fixations, saccades, and smooth pursuits. Results indicate that a simple hybrid method that incorporates velocity and dispersion thresholding allows producing robust classification performance. It is concluded that behavior scores are able to aid automated threshold selection for the algorithms capable of successful classification.

Keywords Eye movements · Classification · Algorithm · Analysis · Scoring · Metrics · Smooth-pursuit

Introduction

The identification of the basic eye movement types from a noisy and frequently inaccurate raw eye positional signal is of the utmost importance to researchers and practitioners who employ eyetrackers in their studies. The human oculomotor system (HOS) primarily exhibits six eye movement types: fixations, saccades, smooth pursuits (SPs),

O. V. Komogortsev (✉) · A. Karpov
Department of Computer Science, Texas State University,
601 University Drive,
San Marcos, TX 78666, USA
e-mail: ok11@txstate.edu

optokinetic reflex, vestibulo-ocular reflex, and vergence (Leigh & Zee, 2006). Among those eye movement types, fixations, saccades, and SPs are most frequently studied. The following brief definitions can be provided for these eye movement types: An eye fixation is an eye movement that keeps an eye gaze stable on a selected stationary target, a saccade is a very rapid eye rotation moving the eye from one fixation point to the next, and SP is an eye movement that follows a moving object with the purpose of keeping the object on a high acuity vision zone called the fovea (Duchowski, 2007; Poole & Ball, 2004). Eye fixations are frequently employed for human–computer interactions as an input modality (Istance, Hyrskykari, Immonen, Mansikkamaa, & Vickers, 2010); saccades and SPs are frequently employed to diagnose pathologies of the HOS or to assess HOS performance in clinical populations (Elina, Aalto, & Pyykkö, 2009). Therefore, accurate automated classification of eye movements is an important topic of research.

Accurate automated eye movement classification is exceedingly difficult due to the noise and inaccuracies inherited from the eye-tracking equipment, the dynamics of the HOS behavior, and variability between and within eye movement classification algorithms. Variation of the single threshold value, in cases in which only fixations and saccades are classified, is reported to substantially affect metrics such as number of detected saccades and fixations, average fixation duration, and saccade amplitude (Ceballos, Komogortsev, & Turner, 2009; Garbutt et al., 2003; Komogortsev, Gobert, Jayarathna, Koh, & Gowda, 2010; Poole & Ball, 2004). Frequently, researchers perform manual classification to avoid misidentification issues associated with automated algorithms. However, in such cases, classification becomes a very long and tedious process. Selection of the thresholds that provide meaningful classification is frequently done empirically, with default values suggested by either eye-tracking vendors or related literature. Given rapid developments of the eye-tracking

technologies that vary in hardware, sampling frequencies, and calibration algorithms (Hansen & Qiang, 2010), it is easy to "copy and paste" suggested thresholds; however, it is hard to validate classification accuracy. During empirical threshold selection by "eye balling" small part of the classified data, it is easy to misclassify some of the recordings or to misidentify corrective behavior such as corrected undershoots, overshoots, dynamic saccades, and so forth (Leigh & Zee, 2006).

It is hard to define meaningfulness of the automated classification given a threshold value. For example, it is possible to assume that the quality of saccade detection can be ultimately judged by such properties as the amplitude–duration relationship, main-sequence relationship, and saccades' waveform[1] (Leigh & Zee, 2006). However, there is a substantial amount of variability for some of those metrics between people (Bollen et al., 1993) and even directional differences for the same person (Smit, Opstal, & Gisbergen, 1990). In this type of circumstances, it is very difficult to judge when the selected threshold produces accurate performance as measured by the above-mentioned metrics, because this performance might depend on multiple factors. Recently, Komogortsev and colleagues proposed a set of behavior scores with the purpose of selecting a meaningful classification threshold using a fixed stimulus (Komogortsev et al., 2010). Behavior scores assume that the amount of saccadic and fixational behavior encoded in a simple step-stimulus is matched by the HOS of a normal person, therefore providing an opportunity to find a threshold value that ensures such performance. Researchers have reported that thresholds selected according to these criteria have provided meaningful classification results (Komogortsev et al., 2010).

It should be noted that the purpose of the scores is not to substitute already established metrics, such as the amplitude–duration relationship and so forth, but to provide an opportunity for the automated selection of the classification parameters immediately after the calibration procedure. In cases in which the experimental stimulus contains step or step-ramp stimulus classification, performance for the whole experiment can be benchmarked with behavior scores, in addition to any other metric employed by the experimenters. The goal of automated threshold selection for a step-ramp stimulus with subsequent employment of the same threshold for dynamic stimuli is particularly attractive, because the step-ramp stimulus is already presented as a part of the calibration procedure. Recording equipment's performance for a given setup and subject is unlikely to change from calibration to the actual recording. Therefore, it is possible to assume that selected thresholds will continue to provide

meaningful classification performance even during a presentation of the stimuli that is different from the calibration.

Automated classification of SPs in the presence of fixations and saccades is an even more difficult task and continues to be a topic of active research (Agustin, 2009; Larsson, 2010). The most difficult part of ternary eye movement classification is separation between fixations and SP. Two main factors contribute to the challenge. (1) A fixation consists of the three submovement types, such as tremor, drift, and microsaccades. As a result, a velocity range during a fixation (velocities up to 30º/s are possible, as computed by the main-sequence relationship; Leigh & Zee, 2006) and SP (velocity up to 100º/s is reported in Carpenter, 1977) overlap. (2) Eye-tracking noise further blurs quantitative boundaries between fixation and pursuit.

Given the importance of ternary eye movement classification and its challenges, it is necessary to find out what degree of meaningful classification can be obtained and whether the accuracy of classification performance can be verified by a set of simple behavior scores.

To start answering these questions, this work (1) introduces behavior scores related to SP, (2) proposes an algorithm for ternary eye movement classification, (3) evaluates automated and manual ternary classification on the basis of the proposed scores, and (4) establishes automated selection of the classification thresholds on the basis of the ideal values of behavior scores.

Overview

Classification of fixations and saccades

In general eye movement, classification algorithms consider different properties of the signal that is captured by an eye-tracker. In cases in which fixations have to be separated from saccades, classification algorithms can be broken into the following groups: (1) position based (dispersion threshold identification [I-DT], minimum spanning tree identification [I-MST]), (2) velocity based (velocity threshold identification [I-VT], hidden Markov model identification [I-HMM], Kalman filter identification [I-KF]), and (3) acceleration based (finite input response filter identification [I-FIR]) (Komogortsev et al., 2010; Salvucci & Goldberg, 2000; Tole & Young, 1981). To the best of our knowledge, these algorithms have not been "successfully" applied to the problem of ternary classification.

Human visual system performance during pursuit stimuli

The SP movement consists of three phases: initiation, steady-state, and termination (Bahill & McDonald, 1983; Leigh & Zee, 2006; Mohrmann & Thier, 1995; Robinson,

---

[1] An example where saccades are detected on the basis of "the peak-velocity-magnitude-duration parameters" "manually" can be found in Bahill, Iandolo, and Troost (1980). However, this manual procedure is obviously very time consuming.

1965). The initiation phase can be broken into three steps: (1) the SP latency when the brain programs the movement, (2) the initial SP represented by an exponential rise in eye movement velocity, and (3) a corrective saccade that brings the target closer to the fovea. The steady-state consists of continuous SP movement that might be interspersed by the corrective saccades. The termination phase consists of three steps: (1) the response latency, (2) an exponential decay in velocity, and (3) an optional corrective saccade(s) that brings the eye to a new target.

It is possible to assume that separation of the steady-state SP from the fixation signal is simple via a velocity threshold; however, the following factors challenge accurate classification.

The first factor is jitter during fixations. Jitter is frequently caused by inaccuracies in the eyetracker's gaze position estimation. Good eyetracker's positional accuracy performance is varied in the range of 0.25°–1°.

The second factor is the presence of miniature eye movements such as drift, microsaccades, and tremors (Leigh & Zee, 2006), which result in high spread of the amplitudes for the positional (e.g., up to 1.5°) and velocity (e.g., up to 40°/s) signals. This spread does not greatly impact classification accuracy if only fixations and saccades are present; however, in cases of low-velocity SPs (e.g., 20–40°/s), the results of classification might be poor.

The third factor is variability of eye movement behavior among people and its dependence on the task. For example, response times can be different for "express saccade" makers and naïve subjects in gap-step-ramp experiments (Kimmig, Biscaldi, Mutter, Doerr, & Fischer, 2002). Also, humans are capable of matching velocities of up to 90°/s during the SP exhibited to a ramp stimulus with constant velocity (Meyer, Lasker, & Robinson, 1985). For unpredictable motion, it has been suggested that humans cannot pursue small targets at speeds faster than 40°/s (Young & Stark, 1963).

Existing algorithms for automated classification of smooth pursuit

In the previous research, the separation of SP was done in cases in which only SP and saccades were present. For example, a single threshold-based algorithm was employed by Bahill et al. (1980). The researchers used a velocity threshold of 50°/s to separate saccades from SPs. All sequences of samples with a velocity greater than threshold were checked upon matching main-sequence relationships. If a sequence of points met these criteria, all its samples were marked as a saccade. Otherwise, the samples were discarded. Bahill et al. were able to use main sequence relationship as a criterion for meaningful classification because only horizontal saccades were considered.

For ternary classification, an interesting approach was proposed by Agustin (2009) and further enhanced by Larsson (2010). The approach monitors the direction of movement and the rate of movement to separate fixations from SP. This approach, together with a new proposed approach, is discussed in the section describing the classification algorithms.

Behavior scores for smooth pursuit classification

Considering the multitude of factors affecting SP performance and especially between-subjects variability, it is important to develop simple metrics that can assess automated eye movement classification performance against a ramp stimulus with constant velocity, signaling the cases of classification success or failures, with an ultimate goal of suggesting parameters/thresholds for meaningful classification even in cases of unpredictable SP-exhibiting content.

Previously, Komogortsev et al. (2010) created a set of behavior scores that allows assessment of classification quality or even determining the optimal threshold values when only fixations and saccades are present. This work continues in the same direction, fine-tuning already existing scores and creating additional scores to assess meaningfulness of ternary classification. For the purposes of the initial assessment, behavior scores assume that the amount of fixational, saccadic and SP behavior encoded in step and ramp stimuli are matched by the HOS in a normal subject.

Scores when only fixations and saccades are present

Komogortsev et al. (2010) originally proposed three behavior scores—namely, the fixation quantitative score (FQnS), fixation qualitative score (FQlS), and saccade quantitative score (SQnS). The scores were originally designed to measure classification quality if only fixations and saccades are present in the raw eye positional trace. We perform the following additions and modifications that allow extending the utility of behavior scores for ternary classification.

Modified saccade quantitative score

The SQnS measures the amount of saccadic behavior in response to a stimulus. The SQnS is defined as the ratio of all detected saccade amplitudes to all saccade amplitudes encoded in the stimulus (Komogortsev et al., 2010). To avoid counting corrective saccades during SP, the SQnS is modified to consider saccades that directly correspond only to the stimulus saccades represented by the instantaneous jump of the target's location. To attain this goal, a temporal window is introduced that considers saccades in response to the step part of the stimulus only. This is achieved by use of

a temporal window that monitors the eye positional signal in a fixed time interval prior to and after stimulus change. This logic allows correctly considering anticipatory saccades and corrective saccades for the SQnS computation.

The ideal SQnS score, which is achieved only if the HOS perfectly executes a saccade within the temporal window and the classifier accurately detects it, is 100 %. In practice, the SQnS value might be lower, because of some amount of the anticipatory and the corrective saccadic behavior that might fall outside of the temporal window.

Modified ideal fixation quantitative score

The FQnS measures the amount of fixational behavior in response to a stimulus. The FQnS is defined as the amount of eye position points, which are part of the fixation related to stimulus fixation,[2] divided by the total number of stimulus fixation points (Komogortsev et al., 2010). The ideal FQnS score presented in Komogortsev et al. (2010) did not consider the effect of SP on score computation; therefore, in the present work, we provide a modified formula that accounts for SP effects:

$$\text{Ideal\_FQnS} = 100 \left(1 - \frac{mS_l + kP_l + \sum_{j=1}^{m} D_{sac\_dur_j}}{\sum_{i=1}^{n} D_{stim\_fix\_dur_i}} \right) \tag{1}$$

where n is the number of stimulus fixations, $D_{stim\_fix\_dur_i}$ is duration of the ith stimulus fixation, S_l is saccadic latency, m is the number of stimulus transitions between fixations and saccades, $D_{sac\_dur_j}$ is the expected duration of a saccade in response to the stimulus saccade j, k is the number of stimulus transitions from SP to fixations, and P_l is the duration of the SP termination phase during fixation stimulus.

Smooth pursuit qualitative scores

The intuitive idea behind the smooth pursuit qualitative scores (PQlS) is to compare the proximity of the detected SP signal with the signal presented in the stimuli. Two scores are indicative of positional (PQlS_P) and velocity (PQlS_V) accuracy.

The PQlS_P and PQlS_V calculations are similar to the FQnS (Komogortsev et al., 2010); that is, for every SP point (x_s, y_s) of the presented stimuli, the check is made for the point in the eye position trace (x_e, y_e). If such a point is classified as part of SP, the Euclidean distance between

these two points and the difference between their speeds are computed. Then the sum of such distances and speed differences are normalized by the amount of points compared:

$$\text{PQlS\_P} = \frac{1}{N} \cdot \sum_{i=1}^{N} pursuit\_distance_i \tag{2}$$

$$\text{PQlS\_V} = \frac{1}{N} \cdot \sum_{i=1}^{N} pursuit\_speed\_difference_i \tag{3}$$

N is the amount of stimuli position points where the stimulus SP is matched with corresponding eye position sample detected as SP; $pursuit\_distance_i = \sqrt{(x_s^i - x_e^i)^2 + (y_s^i - y_e^i)^2}$ and represents the distance between stimuli SP position and the corresponding SP point; $pursuit\_speed\_difference_i = |v_s^i - v_e^i|$ and represents the difference between speeds in the ith stimuli point and the corresponding point in the raw eye positional sequence.

Ideal PQlS scores, which can be achieved only if HOS perfectly matches positional/velocity characteristics of the moving target and no calibration errors are present, are PQlS_P = 0º and PQlS_V = 0º/s. In practice, ideal scores might not be achieved, due to calibration errors, corrective behavior, and classification inaccuracies.

It should be noted that qualitative scores are indicative of two things: (1) how well the HOS follows the target and (2) how accurately the tracking equipment works for a given subject. Considering this, we did not use the existing SP gain metric, defined as peak eye velocity/peak target velocity (Leigh & Zee, 2006), due to the fact that SP gain is designed to measure HOS performance only.

Smooth pursuit quantitative score

The smooth pursuit quantitative score (PQnS) measures the amount of detected SP behavior given the SP behavior encoded in the stimuli. To calculate PQnS, two separate quantities are computed. One measures the total length of the SP trajectories presented by the stimuli. The second one measures the overall length of the properly detected SP by the classifier. The ratio of these two values defines the score:

$$\text{PQnS} = 100 \cdot \frac{total\_detected\_SP\_length}{total\_stmuli\_SP\_length} \tag{4}$$

The computation of the ideal PQnS can be performed as

$$\text{Ideal\_PQnS} = 100 \cdot \left(1 - \frac{n \cdot P_l + \sum_{j=1}^{n} D\,cor\_sac\_durj}{\sum_{i=1}^{n} D\,stim\_pur\_dur_i} \right) \tag{5}$$

---

[2] For simplicity of writing, we use the following definitions that describe stimulus signal behavior: stimulus saccade, step part of the stimulus signal that elicits eye saccades; stimulus SP, ramp part of the signal that elicits eye SP; and stimulus fixation, flat part of the stimulus signal that elicits eye fixations. Eye fixation, saccade, and SP are called simply *fixation*, *saccade*, and *SP*, without any prefix.

where n is the number of stimulus pursuits, $D_{stim\_pur\_dur_i}$ is duration of the ith stimulus pursuit, P_l is pursuit's latency prior to the onset of the corrective saccade that brings the fovea to the target, and $D_{cor\_sac\_dur_j}$ is the expected duration of the corrective saccade. The Ideal_PQnS assumes that the HOS exhibits the SP for the duration of the target's movement immediately after the initial corrective saccade. Subsequently, accurate SP classification has to be performed for the duration of the movement. In practice, ideal score might not be achieved due to classification errors or additional corrective saccades occurring during the SP stimulus.

Misclassified fixation score (MisFix)

Misclassification error of the SP can be determined during a fixation stimulus, when correct classification is most challenging. SP_fixation_points is the number of points in the eye position trace that were classified as SP but the corresponding stimuli point for them is a fixation; total_stimuli_fixation_points is the total number of fixation points in the stimuli. To calculate MisFix, two separate quantities are calculated: SP_fixation_points is the number of points in the eye position trace that were classified as SP but the corresponding stimuli point for them is fixation; total_stimuli_fixation_points is the total number of fixation points in stimuli:

$$\text{MixFix} = 100 \cdot \frac{SP\_fixation\_points}{total\_stimuli\_fixation\_points} \qquad (6)$$

Computation of the ideal MisFix should take into consideration the fact that termination phase of SP continues during fixational stimulus after the SP stimulus is over. Therefore, the following formulation is employed for the computation of the ideal MisFix score:

$$\text{Ideal\_MixFix} = 100 \cdot \left(\frac{n \cdot P_{lt} + \sum_{j=1}^{n} D_{cor\_sac\_dur_j}}{\sum_{i=1}^{n} D_{stim\_fix\_dur_i}} \right) \qquad (7)$$

where n is the number of SP present in the stimuli, average duration of the latency of the termination phase P_{lt} prior to the last corrective saccade leading to fixational stimulus position. $D_{cor\_sac\_dur_j}$ is duration of the corrective saccade, if present. In calculation of the Ideal MisFix, we assumed that each stimulus SP is followed by a stimulus fixation.

Algorithms for smooth pursuits detection

Velocity and velocity threshold identification (I-VVT)

We modify the I-VT algorithm to perform ternary classification. For the purposes of separating SPs from fixations, a second velocity threshold is introduced. To highlight such

modification, the algorithm's name is changed to *velocity and velocity threshold identification* (I-VVT). Figure 1 presents the pseudocode. The pseudocode contains a filter function that accepts a list of the preclassified saccades for the purpose of filtering noisy saccade-like events according to minimum amplitude and duration. In case of this work, such events with amplitudes of less than 3.5° and 4 ms in duration were discarded. The Filter Function" subsection in the Discussion section provides an additional description on the topic of filtering.

The I-VVT algorithm is capable of real-time performance; however, it is not able to provide accurate classification, as is discussed in the Results section.

Note that our implementation of each algorithm presented here and behavior scores, together with eye movement recordings, can be downloaded here (Komogortsev, 2011).

Velocity and movement pattern identification (I-VMP)

We call the approach proposed by Agustin (2009) and enhanced by Larsson (2010) *velocity and movement pattern identification* (I-VMP) because it employs velocity threshold to first identify saccades, similarly to I-VVT. Subsequently, it analyzes the movement patterns to separate SPs from fixations. The movement pattern is analyzed in a temporal window with a size of T_w. In that window, the magnitude of movement is computed by analyzing angles created by every pair of the adjacent positional points and the horizontal coordinate axis. Once the value representing the magnitude of movement is computed, it is compared against a

---

Algorithm: *Smooth Pursuit Classification (I-VVT)*
Input: array of eye position points, saccade velocity threshold - T_{Vs}, smooth pursuit velocity threshold - T_{Vp}
Output: list of fixations, saccades, and smooth pursuits

Calculate point-to-point velocities for each point

Mark all points above T_{Vs} as saccades

Mark all unclassified points below T_{Vp} as fixations

Mark all unclassified points as smooth pursuit

Filter Function (array of pre classified saccades)

Merge Function(array of pre classified smooth pursuits, fixations, and saccades)

> *Merge every group of consecutive saccade points into a saccade with identified onset, offset, and amplitude.*
>
> *Merge every group of fixation points into a fixation with identified centroid coordinates, onset, and duration.*
>
> *Merge every group of smooth pursuit points into a smooth pursuit with identified onset, offset, and trajectory.*

Return saccades, fixations, and smooth pursuits

Fig. 1 Pseudocode for Velocity and Velocity Threshold Identification algorithm

threshold T_m. Values above the threshold are marked as SP, and those below the threshold are marked as fixations. Figure 2 presents the pseudocode. A more detailed description of the algorithm is provided elsewhere (Larsson, 2010).

Velocity and dispersion threshold identification (I-VDT)

In this work, we propose a ternary classification algorithm called *velocity and dispersion threshold identification* (I-VDT). It performs the initial separation of saccades similarly to the I-VVT and the I-VMP. Subsequently, it separates SPs from fixations by employing a modified dispersion threshold identification method, which, within a temporal window of the size T_w, monitors dispersion of the points (corresponding threshold is T_d). Figure 3 presents the pseudocode. Dispersion of the points is computed in the same way as that presented in Salvucci and Goldberg (2000).

Experimental setup

Apparatus

The data were recorded using the EyeLink 1000 eyetracker (EyeLink, 2010) at 1000 Hz on a 21-in. CRT monitor with a screen resolution of 1,024 × 768 pixels and a refresh rate of 80 Hz. The vendor-reported spatial resolution for EyeLink

---

Algorithm: *Smooth Pursuit Classification (I-VMP)*
Input: array of eye position points, saccade velocity threshold - T_V, temporal window size – T_W, movement threshold - T_M
Output: array of fixations, saccades, and smooth pursuits

Calculate point-to-point velocities for each point

Mark all points above T_V as saccades

Filter Function (array of pre classified saccades)

 While temporal window T_W does not reach the end of array

 Mark all unclassified points inside T_W as fixations

 For all pairs of adjacent points inside of T_W calculate an angle created by the pair and horizontal axis

 Represent computed angles as points on circumference of a unit circle

 Calculate mean (M_X, M_Y) of x and y coordinates of those points)

 If (distance between (M_X, M_Y) and (0,0) > T_M)

 Mark fixation points inside the T_W as smooth pursuit

 End if

 End while

Merge Function(array of pre classified smooth pursuits, fixations, and saccades)

Return saccades, fixations, and smooth pursuits

Fig. 2 Pseudocode for Velocity and Movement Pattern Identification algorithm

---

Algorithm: *Smooth Pursuit Classification (I-VDT)*
Input: array of eye position points, velocity threshold - T_V, dispersion threshold - T_D, temporal window size - T_W
Output: array of fixations, saccades, and smooth pursuits

Calculate point-to-point velocities for each point

Mark all points above T_V as saccades

Filter Function (array of pre classified saccades)

Initialize temporal window over first points in the remaining eye movement trace

While temporal window does not reach the end of array

 Calculate dispersion of points in window

 If (dispersion < T_D)

 While dispersion < T_D

 Add one more unclassified point to window

 Calculate dispersion of points in window

 End while

 Mark the points inside of the window as fixations

 Clear window

 Else

 Remove first point from window

 Mark first point as a smooth pursuit

 End if

End while

Merge Function(array of pre classified smooth pursuits, fixations, and saccades)

Return saccades, fixations, and smooth pursuits

Fig. 3 Pseudocode for Velocity and Dispersion Threshold Identification algorithm

---

1000 is 0.01° (EyeLink, 2010). To ensure high accuracy of the eye movement recording, a chinrest was employed. The chinrest was positioned at 70 cm in front of the monitor. The recordings were performed in the monocular mode for the eye that provided the best calibration accuracy. The height of the chinrest was adjusted to ensure that primary position of the recorded eye corresponded to the center of the screen. The stimulus screen was rectangular to the line of view. Recorded raw eye positional signal was first processed by the heuristic one-sample filter described in Stampe (1993) and as implemented by the EyeLink 1000 vendor. The raw eye positional signal was subsequently translated to the coordinates presented in the degrees of visual angle, with the center of coordinate system corresponding to the center of the screen. The procedure of converting the signal from the eye-tracking units to the degrees of the visual angle is described elsewhere (Duchowski, 2007).

Stimulus signal

A 2-D step-ramp stimulus was presented by a moving target. The presented range of stimulus saccades' amplitudes was 14.2°–28.5° (M = 20.2, SD = 6.7). The presented range of stimulus SPs' velocities was 20.1–53.7 °/s (M = 38.0, SD = 11.3). Stimulus SP velocity was constant at each interval. Only a single target was continuously presented throughout

the experiment. Total stimulus duration was approximately 35 s. Detailed target's behavior is described in Table 1 and is supplied as an additional video file attached to this article. The target was presented as a white dot with a size of approximately 1º in diameter and the center marked with a small black dot to facilitate higher targeting accuracy for HOS. The remaining screen's background was black.

The data recorded for the above-described task was part of a larger study that had its purpose in establishing a normal baseline among healthy subjects for subsequent comparison of data with people who had mTBI injuries. Specifically, the task described here was presented as the last stimulus in a battery of eight other step and ramp stimuli tasks. Each task in this sequence was preceded by calibration and calibration

Table 1 Presented step- ramp stimulus characteristics. Ramp characteristics are highlighted with grey. Step characteristics are described by the remaining rows. Value A presents the amplitude of the target's jump (step stimulus) for saccades or distance traveled for the SP eliciting target (ramp stimulus). Value V represents the velocity of target's movement during ramp stimulus. Within each single time interval velocity value was a constant. Target was stationary between step and ramp signal, therefore invoking eye fixations

Onset time, ms	Length, ms	Stimulus onset coordinates, deg		Stimulus Signal	
		X	Y	A	V
1000	709	0	0.00	14.2°	20.1º/s
2709	606	-10	10.15	14.2°	23.5º/s
4314	-	0	0.00	14.2°	-
5314	-	-10	10.15	14.2°	-
6315	531	0	0.00	14.2°	26.8º/s
7846	471	-10	-10.16	14.2°	30.2º/s
9316	-	0	0.00	14.2°	-
10316	-	-10	-10.16	14.2°	-
11317	425	0	0.00	14.2°	33.5º/s
12742	386	10	-10.16	14.2°	36.9º/s
14127	-	0	0.00	14.2°	-
15127	-	10	-10.16	14.2°	-
16128	353	0	0.00	14.2°	40.3º/s
17481	653	10	10.15	28.5°	43.6º/s
19134	607	-10	-10.16	28.5°	46.9º/s
20740	-	10	10.15	28.5°	-
21740	-	-10	-10.16	28.5°	-
22741	400	10	10.15	20.0°	50.0º/s
24141	566	-10	10.15	28.5°	50.3º/s
25707	531	10	-10.16	28.5°	53.6º/s
27237	-	-10	10.15	28.5°	-
28237	-	10	-10.16	28.5°	-
29237	-	-10	10.15	28.5°	-
31237	-	10	-10.16	20.3°	-
32237	-	10	10.15	20.0°	-
33237	-	-10	10.15	14.2°	-

verification procedures. The battery of tasks was designed under the guidance of a physical therapist, with one of the goals to prevent excessive fatigue as part of the task completion. Specifically, in that sequence, a 1-min break (or longer by request) was given to subjects between each individual tasks. The duration of the whole experiment for each subject, on average, was approximately 25 min.

Subjects and recordings

The test data consisted of a heterogeneous subject pool, 18–25 years of age, with normal or corrected-to-normal vision. A total of 11 subjects volunteered for the evaluation test. None of the subjects had prior experience with eye tracking. The mean percentage of invalid data was 1.24 %, with a maximum of 7.61 %. All recordings were employed during automated classification assessment. Only three recordings selected by the criteria described next were employed during manual assessment.

Manual classification

Manual classification was performed by a postdoctoral researcher to establish performance baseline and was done by a visual inspection of the recorded data where raw positional coordinates were converted to the coordinates in degrees of the visual angle, as described in the Apparatus subsection. The process of visual inspection consisted of examining the horizontal and vertical components of movement and, in the most difficult cases, a 3-D view of the signal (Komogortsev, 2011). Saccades were separated when the signal's positional change was large. Fixations were separated when the signal stayed within a certain positional proximity, with jitter, tremor, and microsaccades present in the signal. SP was characterized as a signal with very low jitter and continuous directional change of the eye gaze position. Initial corrective saccades in response to the onset of stimulus SP were classified as saccades.

Due to the considerable time necessary to classify the signal manually (approximately 2.5 h per recording), only three records were classified manually and were labeled as "good," "medium," and "bad." Please note that "good," "medium," and "bad" categorization contains a description of the quality of signal as recorded by the eye-tracking equipment and the quality of the HOS for matching stimulus behavior. Next, we have provided a qualitative description of the recorded signal for each category; however, the process of manual classification and categorization can be considered as subjective. "Good" (subject 7) was selected due to low jitter (approximate average amplitude of 0.2° during fixations), lack of large saccade overshoots/undershoots due to initial accurate saccade in response to the step signal change (the amplitude of the correctional behavior

did not exceed 1.5°), and close match of SP to the stimulus SP position (corrective saccade during SP are infrequent and small in amplitude—i.e., 1°–3°). The "medium" (subject 1) recording had higher jitter (amplitude range, 0.3°–1.5°); corrective saccades to compensate for the initial over-shoots/undershoots in response to the step signal change were large (amplitude range 3-4°), and ramp signal was not well matched by the HOS (more frequent corrective saccades with larger amplitudes—e.g., 1.5°–4°). "Bad" (subject 10) was selected due to high jitter (amplitude range, 1.5°–2°), prolonged corrective behavior during the fixa-tional stimulus that consisted of a sequence of corrective saccades and drifts, and pure matching of ramp signal by the HOS (all corrective saccades had amplitudes higher than 2°).

Ideal scores To compute the Ideal_FQnS by Eq. 1 for the stimulus described by Table 1, the following assumptions are made: Average saccade latency is 200 ms, saccade duration is computed by Eq. 3 in Komogortsev et al. (2010), and average duration of the SP termination phase is 130 ms. As a result, computed Ideal_FQnS is 83.9 %. To compute the Ideal_PQnS by Eq. 5, the following assumption is made after manual inspection of the recorded data: SP latency for stimulus pursuits with velocity <20 °/s is 0 ms, <30 °/s is 230 ms, <40 °/s is 210 ms, <50 °/s is 180 ms, and >50 °/s is 210 ms. Latency numbers estimated here already contain the duration of the initial corrective saccade. As a result, computed Ideal_PQnS is 52 %. Average latency duration in termination phase is 153 ms for our data. Therefore, Ideal_MisFix is computed to be approximately 7.1 %. The latency number estimated here already contains the dura-tion of the final corrective saccade.

Results

Manual classification

Table 2 presents the behavior scores computed for manually classified data, and Fig. 6 presents an example of manually classified data. The FQnS had the closest value to the ideal score of 71 %. The SQnS was lower than the ideal score of 100 % However, the difference was not substantial; that is, average SQnS computed as a result of manual classification was 90 %. The PQnS value was lower than the ideal value of 52 %. However. the difference was not large; that is, average PQnS computed as a result of manual classification was 42 %. This result can be attributed to the fact that frequently, HOS exhibits corrective saccades interspersed with fixations to follow ramp stimuli. Such corrective saccades lower the PQnS value. The average MisFix was higher than the ideal number of 7.1 %, due to variations in the SP termination phase and misclassification errors; however, for the record marked as

Table 2 Manual classification results and ideal behavior scores

Evaluated Quality	Good	Medium	Bad		
Score/Subject	S7	S1	S10	Average	Ideal Scores
SQnS	96 %	84 %	89 %	90 %	100 %
FQnS	71 %	63 %	42 %	56 %	84 %
PQnS	39 %	47 %	40 %	42 %	52 %
MisFix	6 %	13 %	33 %	17 %	7.1 %
FQlS	0.44°	0.46°	0.58°	0.49°	0°
PQlS_P	3.15°	3.07°	2.58°	2.93°	0°
PQlS_V	23°/s	39°/s	30°/s	31°/s	0°/s

"good," the MisFix was almost the same as the ideal number. All behavior qualitative scores present reasonable values indi-cating relatively small positional and velocity errors between the presented stimulus and recorded eye movements.

Automated classification

The velocity threshold that separates saccades from fixa-tions and SP was set to 70º/s for all classification algorithms considered in this work. Such a threshold was selected following the recommendations presented in Komogortsev et al. (2010) allowing fixing the saccade classification per-formance and investigating the performance of the most challenging part of the classification—that is, separation of SPs from fixations. The resulting SQnS was 92 % for all classification algorithms, which is quite close to the ideal score of 100 % discussed in Komogortsev et al. (2010) and to the results of the manual classification.

I-VVT Figure 4 presents behavior scores. The FQnS starts extremely low and increases together with the value of the SP threshold. The PQnS score starts at 41 % and decreases. The MisFix starts high and decreases to 0 % when the SP threshold reaches saccade threshold. Increase of the FQnS and the parallel decrease of the PQnS depict classification failure of the I-VVT,
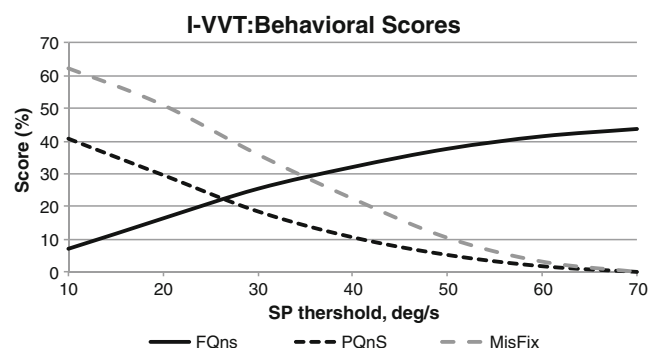

Fig. 4 Behavior scores for the I-VVT. x axis – represents the value of SP threshold (TVp). y – axis represents score(s) value

which is represented by the impossibility of accurately classi-fying both fixations and SPs at the same time. The intersection point at SP threshold of 26°/s yielding FQnS = PQnS = 22 % is far from the values provided by manual classification. At the same time, mismatch scores are too high.

I-VMP Figure 5 presents the behavior scores. The values of the FQnS and the PQnS immediately indicate that magni-tude of movement threshold T_m with values of 0.1 and 0.4 does not yield acceptable classification performance; that is,

in the case of $T_m = 0.1$, the FQnS is too low and the PQnS is too high, and in the case of $T_m = 0.4$, the FQnS is too high and the PQnS is too low. The threshold value of $T_m = 0.2$ provides the most usable case, where the FQnS slightly grows, when temporal window size increases, and essential-ly reaches the value of 63 %. The PQnS slightly decreases, eventually reaching the value of 49 % and stabilizing at that value, starting temporal window threshold of $T_w = 120$ ms. Obtained quantitative scores are not far from the average values depicted in Table 2. Mismatch score (MisFix) starts
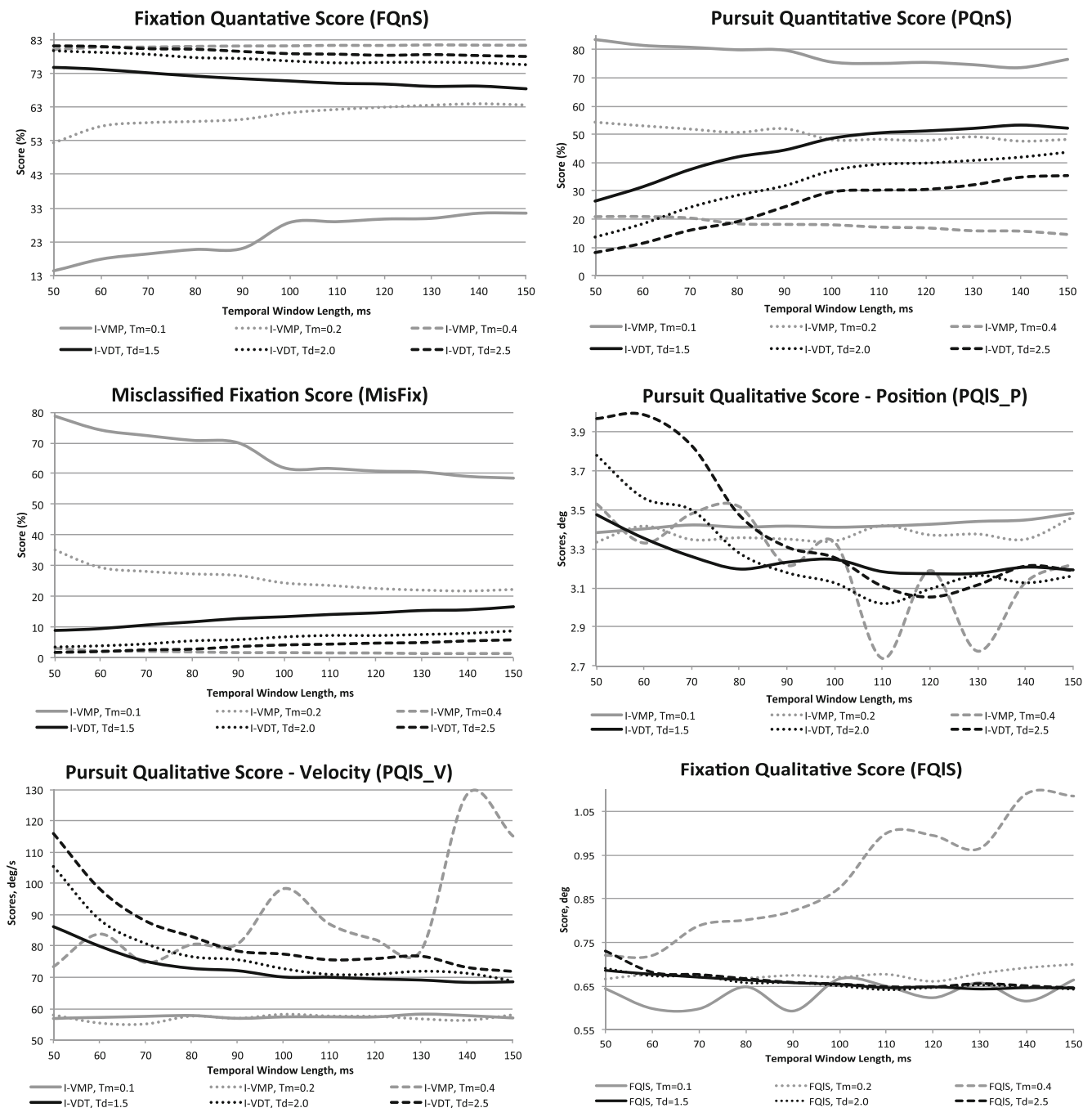

Fig. 5 Behavior scores for the I-VMP & the I-VDT. x axis – represents the size of the temporal sampling window. y – axis represents score(s) value

with relatively high value but decreases with the increase of the temporal window. Score value stabilizes and becomes close to the average depicted in Table 2 after the temporal window reaches 120 ms. The FQlS does not exceed 1.1°. The PQlS_V remains relatively stable at 58 %. The PQlS_P fluctuates at approximately 3.4°.

I-VDT Figure 5 presents classification performance of the I-VDT algorithm. The impact of two factors on the I-VDT performance is investigated: dispersion threshold and the size of the temporal window. The increase in dispersion threshold increases the FQnS, however, slightly yielding a maximum at 82 %. The increase in dispersion threshold T_d significantly decreases the PQnS. At $T_d = 1.5º$, the PQnS almost reaches 53 %, while at $T_d = 2.5º$, the PQnS reaches only 37 %. The size of the temporal window does not impact the FQnS; however, the growth in the temporal window size produces substantial growth in the PQnS. The PQnS starts saturating at window sizes exceeding 110 ms. Eventually, obtained quantitative scores are not far from the average values depicted by Table 2. MisFix is higher for smaller dispersions. The growth of the temporal window makes MisFix grow slowly, essentially, reaching the value obtained by the manual classification (Table 2). Qualitative scores, with the exception of the PQnS_V, are not affected by either the dispersion threshold or the temporal window size. The velocity error represented by the PQnS_V goes down when temporal window size is increased. The PQnS_V value is saturated after the temporal window size reaches 110 ms. A smaller dispersion value yields a smaller PQnS_V value. The FQlS stays below 0.75° for all threshold and temporal window sizes.

I-VDT versus I-VMP From our experimentation, we conclude that the performance of the I-VDT is less impacted by the thresholds than is the performance of the I-VMP. If the optimum thresholds are selected for the I-VMP, classification performance becomes very similar to that of the I-VDT; however, qualitative scores (FQlS, PQnS_V, PQnS_P) and

MisFix are slightly better for the I-VDT when most usable thresholds are considered.

Discussion

Manual selection of meaningful thresholds

On the basis of the results of classifications presented by Figs. 4 and 5, we can manually select classification thresholds that provide the best classification performance for each algorithm. Please note that the saccade-related threshold is fixed to 70°/s for all algorithms during manual classification .

For the I-VVT, the optimal value of the fixation threshold is 26°/s, for which more or less balanced performance is achieved. However, this optimal point produces low qualitative scores and high mismatch scores when compared with the average values presented by Table 2.

For the I-VMP, the optimal value of the magnitude of movement threshold is $T_M = 0.2$, with a temporal window range between 120 and 140 ms. Such thresholds produce the scores that are close to the average values depicted in Table 2.

For the I-VDT, the optimal dispersion threshold is $T_D = 2º$, with the temporal window of 110–150 ms. These thresholds allow one to obtain scores that are close to the average scores presented by Table 2. An example of the raw eye positional signal classified by the above-mentioned thresholds is depicted by Fig. 7.

Automated selection of meaningful thresholds on the basis of the ideal behavior scores

In this section, we investigate the feasibility of automated selection of classification thresholds on the basis of the values of the ideal behavior scores. The idea is to select classification threshold values that allow minimizing the difference between actual and the ideal values of the behavior scores. For this purpose, the following objective function is selected for the minimization process:

$$F(T_1, T_2, \ldots, T_i) = \sqrt{(\text{Ideal\_SQnS} - \text{SQnS})^2 + (\text{Ideal\_FQnS} - \text{FQnS})^2 + (\text{Ideal\_PQnS} - \text{PQnS})^2} \qquad (8)$$

where T_1, T_2, …,T_i are thresholds employed for the identification and SQnS, FQnS, PQnS actual behavior scores that are achieved for given thresholds.

We have employed the Nelder–Mead simplex algorithm (Lagarias, Reeds, Wright, & Wright, 1998) (fminsearch implementation in MATLAB) with an objective function presented by Eq. 7 and 8 to select optimal threshold values.

For the I-VVT, the optimal velocity threshold for saccades is $T_{Vs} = 90 °/s$; the optimal velocity threshold for SP is $T_{Vp} = 50 °/s$. These thresholds allow one to obtain the

following behavior scores: SQnS = 93.1 %, FQnS = 40.3 %, PQnS = 11.4 %, and MisFix = 22.4 %.

For the I-VMP, the optimal velocity threshold is $T_V = 90 °/s$; the dispersion threshold is $T_D = 2.5º$, with the temporal window of $T_W = 80$ ms. These thresholds allow one to obtain the following behavior scores: SQnS = 90.4 %, FQnS = 68.9 %, PQnS = 44.3 %, and MisFix = 15.9 %.

For the I-VDT, the optimal velocity threshold is $T_V = 75 °/s$; the dispersion threshold is $T_D = 1.9º$, with the temporal window of $T_W = 150$ ms. These thresholds allow one to obtain

the following behavior scores: SQnS = 91.6 %, FQnS = 74.95 %, PQnS = 46.07 %, and MisFix = 9.4 %. An example of the raw eye positional signal classified by the above-mentioned thresholds is depicted by Fig. 8.

Manual versus automated selection of classification thresholds

There is little difference (e.g., Fig. 7 vs. Fig. 8) when classification thresholds are selected manually on the basis of the overall picture of the classified data (e.g., Fig. 5) and with the fully automated approach on the basis of the proposed objective function (Eq. 7 and 8). Practically, automated threshold selection might be preferable, due to the reduced burden on the facilitator and the reasonable outcome of the classification performance.

Variability of HOS performance

During manual inspection of the recorded data, we have noticed substantial variability of the HOS performance between and within subjects. SP latency, the size of the corrective saccades, and the quality of target's tracking vary substantially. Often, during ramp stimulus, the HOS exhibits a sequence of corrective saccades interspersed by fixations, rather than tracking the target smoothly. Corrective saccades were more frequent for faster moving targets. We hypothesize that this behavior is exhibited due to dot jumps (step part of the stimulus) that occur in between smooth dot movements (ramp part of the stimulus). Participating subjects do not know when the step or ramp part is going to occur, and therefore, there is a tendency to compensate more with saccadic movement even in case of the ramp stimulus. This hypothesis is supported, in part, by evidence that the properties of the previous stimulus affect HOS performance during the current task (Collins & Barnes, 2009). Another
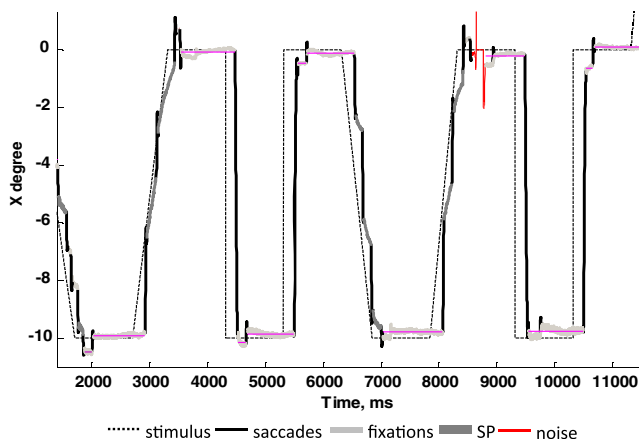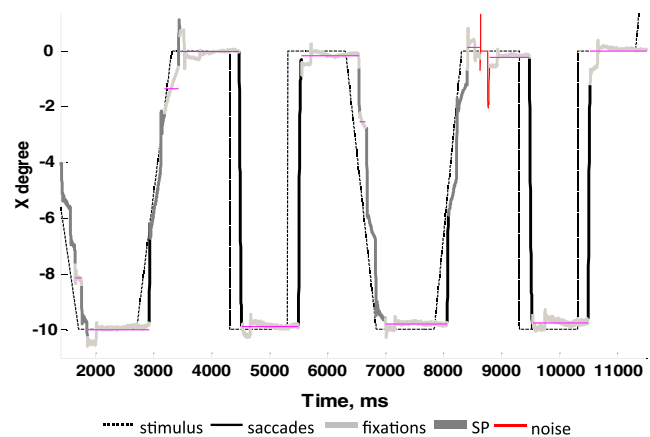

Fig. 7 Results of classification performed by I-VDT for Subject 7 (S7). Classification thresholds are selected manually based on classification performance depicted by Fig. 5

explanation is possible fatigue due to the fact that the stimulus explored here was presented as the last task in a sequence of other tasks, even though the whole sequence of tasks was designed not to cause excessive fatigue. There is evidence that fatigue might result in excessive presence of corrective saccades during SP stimuli (Bahill et al., 1980). Sometimes, closer to the end of the recording, when a subject has experienced a variety of SP stimuli, the HOS started exhibiting, during fixational stimulus, occasional movements with characteristics resembling the SP even after the termination phase was over. A portion of MisFix errors documented in Table 2 highlight this peculiarity. We have not found any literature that documented similar HOS performance. Such HOS performance further complicates ternary classification and necessitates very careful construction of the ideal behavior scores.
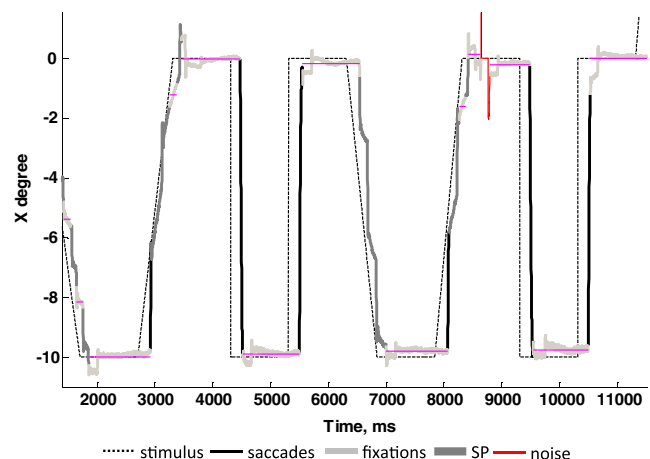

Fig. 6 Results of classification performed manually Subject 7 (S7). Only horizontal component of movement is displayed

Fig. 8 Results of classification performed by I-VDT for Subject (S7). Classification thresholds are selected automatically by proposed objective function (Eq. 7 and 8))

Filter function

A raw eye positional signal frequently contains jitter and also spikes of noise caused by blinks, equipment slippage, and so forth. An example of noise can be seen as red spikes in Fig. 6, which are observable during the 8 or 9 s of the recording. It is important to filter out such events in order to exclude their impact on signal classification and computation of the behavior scores. "Filter Function," presented in the pseudocodes of the algorithms described earlier, performs this role. Automated detection of proper noise events is difficult. Therefore, in our implementation of the "Filter Function," we filter out events initially classified as saccades but that are too short to be actual saccades. A duration threshold of 4 ms is employed for these purposes. In addition, all saccades with an amplitude of less than $3.5°$[3] are marked for the reclassification to became part of the fixation, SP, or noise. This is done to prevent actual fixation or pursuit signal to be broken into a sequence of disconnected pseudosaccades, which might occur for the signal recorded at such a high temporal sampling rate if this amplitude threshold is lowered. The stability of the fixation/SP detection and noise might come at a price of filtering out actual microsaccades and corrective saccades of small amplitudes. Additional research is necessary for filtering tools that would accurately remove noise while keeping actual miniature eye movements intact.

In our setup, those empirically selected thresholds for the amplitude and duration allowed obtaining reasonable automated classification performance; however, other experimental setups might require adjustment of the filtering mechanisms and associate thresholds.

It should be noted that during manual classification, noise-related events are easier to identify due to complete overview of the waveform of the signal. Further research has to be conducted to investigate the impact of the filtering thresholds on classification performance and behavior scores for various types of the experimental setup.

Limitations of the study

A very specific hardware and step-ramp stimulus was employed in this work to establish a baseline on a high-accuracy, high-sampling-frequency eyetracker. A chinrest was employed for additional stability of the recorded data. Additional research is necessary to provide a more comprehensive performance picture of ternary eye movement classification algorithms that employ different hardwares,

allow freedom of head movement, and contain different stimuli characteristics. We expect that proposed behavior scores would be helpful for assessment of automated classification performance; however, careful consideration should be given to the calibration stimulus, types of subject groups used for the recording, and recording environment variables.

Conclusions

This article considered and introduced methods for reliable automated ternary classification that consists of three eye movement types: fixations, saccades, and smooth pursuit. This task is extremely challenging due to the substantial variability of oculomotor system performance between and within subjects, difficulties in separation of fixations from smooth pursuit, and substantial noisiness of the eye-tracking data.

We have extended the set of behavior scores originally introduced by Komogortsev and colleagues (Komogortsev et al., 2010) with the purpose of assessing the meaningfulness of ternary classification. Ideal score values were estimated, and an additional baseline in the form of manually classified data of various quality was established.

Our findings indicate that a simple extension of the popular velocity threshold method (I-VT) algorithm with an idea of separating fixations from SPs with an auxiliary velocity threshold will not provide meaningful ternary classification. Two additional algorithms were considered: velocity movement pattern identification (I-VMP), as introduced by Agustin (2009) and Larsson (2010), and the algorithm that we have developed in this work, velocity dispersion threshold identification (I-VDT). Both algorithms, when driven by the optimal thresholds, were able to provide classification results that were close to the results obtained via manual classification. However, within considered threshold intervals, the I-VDT had smaller performance variability and dependence on the thresholds than did the I-VMP, possibly indicating higher practical usefulness. Misclassification errors were also slightly smaller for the proposed I-VDT algorithm. Classification speed was linear for both algorithms.

It was possible to automatically select classification thresholds with an objective function on the basis of the ideal behavior scores to ensure meaningful classification for all algorithms except I-VVT, for which accurate identification of fixations and SP is impossible. Such an automated threshold selection method should be particularly useful for eye-tracking practitioners, who would be able to use suggested thresholds for a variety of stimuli recorded immediately after the calibration procedure.

---

[3] Please note that 3.5° represents 2-D amplitude of movement, which makes size of the filtered horizontal and vertical components of movement much smaller.

References

Agustin, J., S. (2009). *Off-the-shelf gaze interaction* (Unpublished doctoral thesis). IT University of Copenhagen.

Bahill, A. T., Iandolo, M. J., & Troost, B. T. (1980). Smooth pursuit eye movements in response to unpredictable target waveforms. *Vision Research, 20,* 923–931.

Carpenter, R. H. S. (1977). *Movements of the eyes*. London: Pion.

Ceballos, N., Komogortsev, O., & Turner, G. M. (2009). Ocular imaging of attentional bias among college students: Automatic and controlled processing of alcohol- related scenes. *Journal of Studies on Alcohol and Drugs, 70,* 652–659.

Collins, C. J. S., & Barnes, G. R. (2009). Predicting the unpredictable: Weighted averaging of past stimulus timing facilitates ocular pursuit of randomly timed stimuli. *Journal of Neuroscience, 29,* 13302–13314. doi:10.1523/jneurosci.1636-09.2009

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice* (2nd ed.). London: Springer.

Bollen, E., Bax, J., van Dijk, J. G., Koning, M., Bos, J. E., Kramer., C. G., . . . van der Velde, E. (1993). Variability of the main sequence. *Investigative Opthalmology & Visual Science, 34,* 3700–3704.

Elina, E., Aalto, H., & Pyykkö, I. (2009). Oculomotor findings mimicking a cerebellar disorder and postural control in severe Meniere's disease. *Auris, Nasus, Larynx, 36,* 36–41.

EyeLink (2010). EyeLink II. Retrieved from http://www.sr-research.com/EL_1000.html

Garbutt, S., Han, Y., Kumar, A. N., Harwood, M., Harris, C. M., & Leigh, R. J. (2003). Vertical optokinetic nystagmus and saccades in normal human subjects. *Investigative Ophthalmology & Visusal Science, 44,* 3833–3841. doi:10.1167/iovs.03-0066

Hansen, D. W., & Qiang, J. (2010). In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32,* 478–500.

Istance, H., Hyrskykari, A., Immonen, L., Mansikkamaa, S., & Vickers, S. (2010). *Designing gaze gestures for gaming: An investigation of performance.* Paper presented at the Symposium on Eye-Tracking Research and Applications.

Kimmig, H., Biscaldi, M., Mutter, J., Doerr, J., & Fischer, B. (2002). The initiation of smooth pursuit eye movements and saccades in normal subjects and in "express-saccade makers. *Experimental Brain Research, 144,* 373–384. doi:10.1007/s00221-002-1059-z

Komogortsev, O., V. (2011). Eye movement classification [Software]. Retrieved from http://www.cs.txstate.edu/~ok11/emd_offline.html

Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D., & Gowda, S. (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on Biomedical Engineering, 57,* 2635–2645.

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal on Optimization, 9,* 112–147.

Larsson, L. (2010). *Event detection in eye-tracking data* (Unpublished master's thesis). Lund University, Lund, Sweden.

Leigh, R. J., & Zee, D. S. (2006). *The neurology of eye movements*. Oxforod: Oxford University Press.

Meyer, C. H., Lasker, A. G., & Robinson, D. A. (1985). The upper limit of human smooth pursuit velocity. *Vision Research, 25,* 561–563.

Mohrmann, H., & Thier, P. (1995). The influence of structured visual backgrounds on smooth-pursuit initiation, steady-state pursuit and smooth-pursuit termination. *Biological Cybernetics, 73,* 83–93. doi:10.1007/bf00199058

Poole, A., & Ball, L. J. (2004). Eye tracking in human-computer interaction and usability research: Current status and future prospects. In C. Ghaoui (Ed.), *Encyclopedia of human computer interaction* (pp. 211–219). Hershey, PA: Idea Group.

Robinson, D. A. (1965). The mechanics of human smooth pursuit eye movement. *The Journal of Physiology, 180,* 569–591.

Salvucci, D. D., & Goldberg, J. H. (2000). *Identifying fixations and saccades in eye tracking protocols.* Paper presented at the Eye Tracking Research and Applications Symposium, New York.

Smit, A. C., Opstal, A. J., & Gisbergen, J. A. M. (1990). Component stretching in fast and slow oblique saccades in the human. *Experimental Brain Research, 81,* 325–334. doi:10.1007/bf00228123

Stampe, D. (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, 25,* 137–142. doi:10.3758/bf03204486

Terry Bahill, A., & McDonald, J. D. (1983). Smooth pursuit eye movements in response to predictable target motions. *Vision Research, 23,* 1573–1583. doi:10.1016/0042-6989(83)90171-2

Tole, J. R., & Young, L. R. (1981). Digital filters for saccade and fixation detection. In D. Fisher, R. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 185–199). Hillsdale, NJ: Erlbaum.

Young, L. R., & Stark, L. (1963). Variable feedback experiments testing a sampled data model for eye tracking movements. *IEEE Transactions on Human Factors in Electronics, 4,* 38–51. doi:10.1109/thfe.1963.231285