Measuring Saccade Latency Using Smartphone Cameras

Hsin-Yu Lai, *Student Member, IEEE*, Gladynel Saavedra-Peña, *Student Member, IEEE*, Charles G. Sodini, *Fellow, IEEE*, Vivienne Sze, *Senior Member, IEEE*, and Thomas Heldt, *Senior Member, IEEE*

Abstract*—*Objective: **Accurate quantification of neurodegenerative disease progression is an ongoing challenge that complicates efforts to understand and treat these conditions. Clinical studies have shown that eye movement features may serve as objective biomarkers to support diagnosis and tracking of disease progression. Here, we demonstrate that saccade latency—an eye movement measure of reaction time—can be measured robustly outside of the clinical environment with a smartphone camera.** ***Methods:*** **To enable tracking of saccade latency in large cohorts of patients and control subjects, we combined a deep convolutional neural network for gaze estimation with a model-based approach for saccade onset determination that provides automated signal-quality quantification and artifact rejection.** ***Results:*** **Simultaneous recordings with a smartphone and a high-speed camera resulted in negligible differences in saccade latency distributions. Furthermore, we demonstrated that the constraint of chinrest support can be removed when recording healthy subjects. Repeat smartphone-based measurements of saccade latency in 11 self-reported healthy subjects resulted in an intraclass correlation coefficient of 0.76, showing our approach has good to excellent test–retest reliability. Additionally, we conducted more than 19 000 saccade latency measurements in 29 self-reported healthy subjects and observed significant intra- and inter-subject variability, which highlights the importance of individualized tracking. Lastly, we showed that with around 65 measurements we can estimate mean saccade latency to within less-than-10-ms precision, which takes within 4 min with our setup.** ***Conclusion*** **and** ***Significance:*** **By enabling repeat measurements of saccade latency and its distribution in individual subjects, our framework opens the possibility of quantifying patient state on a finer timescale in a broader population than previously possible.**

***Index Terms*—Eye tracking, convolutional neural networks, health monitoring, saccade latency, mobile imaging.**

I. Introduction

OBJECTIVE and accurate tracking of neurodegenerative disease progression remains an ongoing challenge. Clinical examinations are typically spaced out across intervals over which the functional decline might be subtle, especially early in the disease process, and consequently difficult to ascertain using standard clinical tools. Patient assessment also relies on testimony from patients, family members or care-providers, which is subjective and prone to recall bias. Blood or cerebrospinal fluid sampling for determination of biomarker levels is invasive, and repeat imaging studies are costly. Finally, standard neurocognitive and neuropsychological test batteries require a trained observer to administer and score the test, demand significant patient time and cooperation, and can suffer from re-test variability and methodological limitations that may mask signs of the underlying disease progression [1]–[3].

The lack of objective and accurate assessment tools to quantify disease state and precisely track disease progression not only limits routine clinical assessments but also hinders the development and validation of novel treatment strategies. Since the quest for disease-modifying therapies in neurodegenerative diseases is increasingly focusing on the early or even prodromal stages of the disease process, the need for accurate and precise measures of disease progression and response to treatment has become urgent [4], [5]. It has been suggested that laboratory-based functional assessments, especially of eye movement patterns, may prove to be useful and informative adjuncts to the standard neurocognitive assessment tools in routine clinical care and clinical trials and may therefore help address this critical need [6]–[9].

Registration and analysis of eye movement patterns have attracted significant attention in neurophysiology, clinical medicine and —more recently—human-computer interfacing and gaming [10]–[13]. In the context of aiding in the differential diagnosis and tracking of neurocognitive diseases, rapid

Manuscript received January 31, 2019; revised March 28, 2019; accepted April 18, 2019. Date of publication April 30, 2019; date of current version March 6, 2020. This work was supported in part by a 3M Non-Tenured Faculty Award to Prof. V. Sze, in part by GE Global Research through MIT's Medical Electronic Device Realization Center, and in part by SenseTime through MIT's Quest for Intelligence. *(Corresponding author: Thomas Heldt.)*

H.-Y. Lai and G. Saavedra-Peña are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: hsinyul@mit.edu; gladynel@mit.edu).

C. G. Sodini is with the Department of Electrical Engineering and Computer Science, Institute for Medical Engineering and Science, and the Microsystems Technology Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: sodini@mit.edu).

V. Sze is with the Department of Electrical Engineering and Computer Science and the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: sze@mit.edu).

T. Heldt is with the Department of Electrical Engineering and Computer Science, Institute for Medical Engineering and Science, and the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: thomas@mit.edu).

Digital Object Identifier 10.1109/JBHI.2019.2913846

shifts in gaze (so-called saccades or saccadic eye movements)—whether spontaneous, volitional, or reflexive— have been of particular interest, especially in response to suddenly appearing visual stimuli [6], [14], [15]. Such visual reaction tasks require attention to and continual analysis and evaluation of the environment as well as appropriate decision-making and execution of oculomotor responses once a stimulus is registered. This stimulus-response paradigm therefore probes cognitive and oculomotor function, either or both of which can be impaired in neurocognitive diseases [6]. Saccadic eye movements also provide for a very rich set of features to analyze. A commonly studied feature is saccade latency, which is the time elapsed between the appearance of a visual stimulus and the beginning of the eye movement either toward (pro-saccade) or away from (anti-saccade) the stimulus [13]. An increase in saccade latency has been reported in Parkinson's disease [16]–[18], dementia with Lewy bodies [16], Huntington's disease [19], and Alzheimer's disease [14], [20]–[23]. Evidence that analysis of saccadic eye movement patterns may indeed be informative and add diagnostic value has also been reported for the differential diagnosis between Parkinson's disease and atypical parkinsonian syndromes [24], [25]. Additionally, differences in saccade latency and peak saccade velocity between horizontal and vertical shifts in gaze have shown to be particularly prominent in patients with progressive supranuclear palsy [14]. The error rates and types of errors committed in visual reaction tasks can also provide important disease-related information [6]. In Huntington's disease, for example, anti-saccade error rate has been found to increase with increased predicted (pre-Huntington's) or actual (Huntington's) disease burden [26].

Even though eye movement patterns provide useful quantitative information about a patients disease state, clinical studies of eye movement disturbances in neurodegenerative diseases have largely been based on cohort studies with comparatively small numbers of patients. This might, in part, be related to the need for special eye tracking equipment and a controlled environment within which to conduct eye tracking studies. Consequently, these studies require patients to visit the clinic or laboratory to participate in the measurement sessions. An alternative to this approach could be afforded by performing eye movement tracking and analysis at the convenience of the patient on consumer-grade electronic devices such as cell phones, tablets or laptop computers with user-facing cameras. In fact, the use of such "digital biomarkers" has recently attracted significant attention in neurology [27]–[29].

Here, we demonstrate that measurements of saccade latency can be made robustly using smartphone cameras. We propose a model-based approach to saccade-onset detection that allows for automated flagging and rejection of eye-movement traces that might be of questionable quality. We evaluate the resulting saccade latency measurements under a variety of environmental conditions and assess the intra- and inter-subject variability in saccade latency in self-reported healthy subjects. Finally, we determine the re-test variability of cell-phone based saccade latency measurements. The work opens the possibility for broader eye-movement measurements to be conducted on consumer-grade devices thus enabling tracking of such digital biomarkers on a much finer timescale (e.g. daily) than is currently possible with laboratory-based eye-movement assessment and thereby potentially aiding the characterization of disease progression and quantification of patient state. A preliminary version of this work has been reported in [30] and [31].

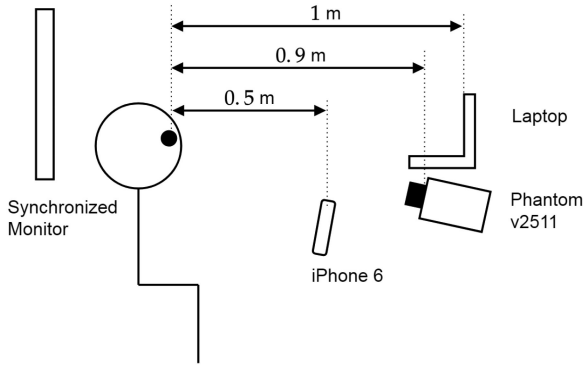

Fig. 1. Diagram of the video recording set-up. A subject is seated facing an iPhone, (in some experiments) a high-speed camera, and a laptop displaying the visual stimulus task. A synchronized monitor behind the subject also displays the visual stimulus task so the cameras capture the eye movements and the visual task simultaneously.

II. Materials

A. Video Recordings

Video recording of volunteers was approved by MIT's Committee on the Use of Humans as Experimental Subjects, and informed consent was obtained from each participant prior to recording. Subjects were seated centrally in front of a laptop at a distance of about 1 m, with their chin placed comfortably on a soft chinrest to minimize head movements (Fig. 1). The sequence of visual stimuli were presented on the laptop screen. A second monitor was placed behind the subject's head, facing and mirroring the laptop screen. An iPhone 6 was placed centrally between the subject and the laptop screen at a distance of about 0.5 m from the subject and with the rear-facing (non-selfie) camera facing the subject. The laptop position was chosen to generate eye movements of 10° amplitude, and the camera position was chosen to capture the subject's face and the mirrored screen during the task, thus capturing the eye movement and the visual stimulus sequence in the same recording. Video recordings were made in slow-motion mode, resulting in recordings at 240 frames per second (fps) and a resolution of 1280×720 pixels. In a subset of recordings, we additionally and simultaneously collected reference videos with a high-speed camera (Phantom v2511) at 500 fps and a resolution of 1280×720 pixels (see Table I). The distance from the high-speed camera to the subject was about 0.9 m; the camera lenses focused on the subject's eyes. Most recordings were acquired under fluorescent lighting. To understand the robustness of the recordings to realistic variations in ambient conditions, we collected a separate set of recordings while varying the lighting conditions with the help of LED panels, and subjects were recorded with and without glasses. In this work, all video recordings were processed offline.

TABLE I
CAMERA & RECORDING SPECIFICATIONS

	Frame Rate	Resolution	ISO	Pixel Size	Shutter Type	Shutter Speed	Cost
iPhone 6	240 fps	1 280×720	32-160	1.5 μm	Rolling	0.125-333.33 ms	$200-400
Phantom v2511	500 fps	1 280×720	6 400-32 000	28 μm	Global	1.9 ms	~$150 000

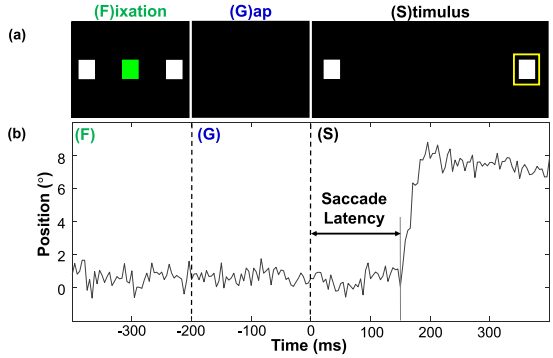

Fig. 2. (a) Example of the visual tracking task during a saccade-latency measurement. The tasks consisting of a fixation period (F), a gap (G), and the appearance of the stimulus (S). Only the final 200 ms of the fixation period are shown. (b) The corresponding horizontal eye movement trace.

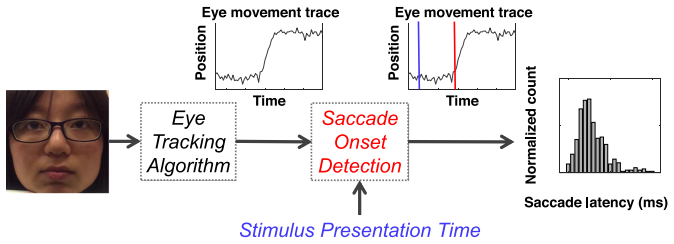

Fig. 3. Pipeline for automated saccade-latency measurement, consisting of eye tracking and saccade-onset detection. The time difference between the stimulus presentation time (blue line) and the saccade onset (red line) is the saccade latency.

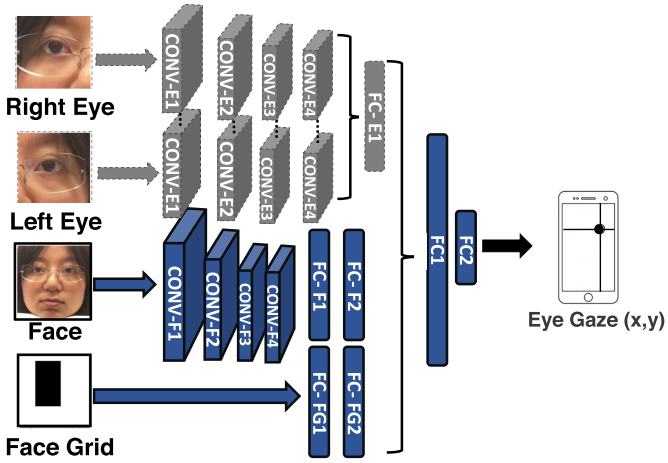

Fig. 4. Convolutional neural network architecture used by iTracker and iTracker-face [33]. iTracker processes the face grid and the eye and face layers *(gray and blue)*, while iTracker-face only processes the face layers *(blue)*. See [33] for details.

B. Task Design

We used the Psychophysics Toolbox 3 for Matlab [32] to implement the visual fixation/stimulus task presented to participating subjects on the laptop screen. A single saccade task started with a fixation period in which three squares were presented on the screen, arranged horizontally, against a black background, a green square at the center of the laptop screen and two white squares arranged at a horizontal distance on either side (Fig. 2a). Subjects were asked to fix their gaze on the green square. After 1000 ms of fixation, all three squares disappeared. Following a 200 ms gap, the two lateral squares reappeared in their original position, with one of them bounded by a yellow square (the stimulus). Subjects were tasked with moving their eyes to – and subsequently keeping their gaze fixed on – the stimulus (Fig. 2b). After the stimulus disappeared, subjects returned their gaze back to the centrally located green square. This task was repeated 40 times per trial, with a total of 20 stimuli appearing on the right and 20 on the left in randomized order. Each recording session consisted of three such trials conducted in close succession, resulting in 120 saccade tasks per session and taking about ten minutes to complete (including breaks between trials).

III. METHODS

The two principal steps in determining saccade latency are (1) *eye tracking* to extract the eye position from each frame in a video sequence, and (2) *saccade-onset detection* to determine when the eyes begin to move (Fig. 3). In this section, we discuss how each step is performed within our signal-processing pipeline.

A. Eye Tracking Algorithm

Several algorithms have been proposed to estimate gaze on portable devices [33]–[35]. Here, we first discuss our use of iTracker [33], a pre-trained convolutional neural network (CNN) previously designed for gaze estimation on smartphones, for tracking the eye position as a function of time. We then propose iTracker-face, a subset of the iTracker neural net, for eye tracking and saccade-onset detection.

To estimate where a user is looking on a screen, iTracker was trained on static images taken with the front-facing (selfie) camera of an iPhone or iPad. These images were collected through an iOS application named GazeCapture, which includes built-in iOS face and eye detectors. The inputs to iTracker include a face grid that indicates the location of the face within the image, a cropped image of the face and cropped images of the right and left eye (Fig. 4), where the cropped face and eyes were determined by the iOS detectors and were resized to 224 × 224 pixels.

Since we did not collect our data through an iOS application, we manually annotated six anatomical landmarks on the first

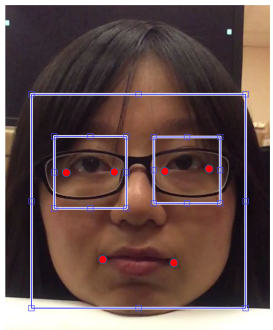

Fig. 5. Manual eye crops and face crops for input to iTracker. The corners of the eyes and the mouth are manually determined on the first frame. The bounding boxes show the regions of eye and face crops derived from these fiducial markers.

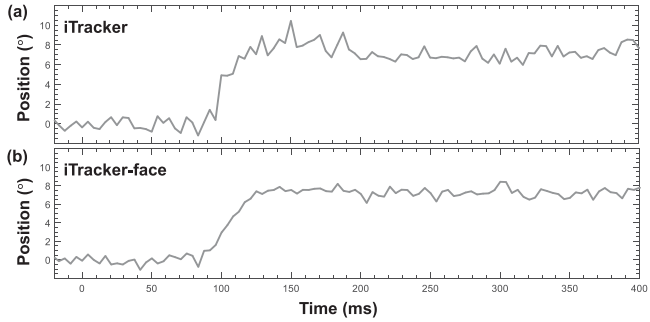

Fig. 6. The same sample eye-movement trace from (a) iTracker and (b) iTracker-face.

frame of each video clip: the two corners of each eye and the two corners of the mouth. To crop each eye region, in accordance with [36], we determined the midpoints of the inner and outer corners of each eye and surrounded these midpoints with squares of width 1.5 times the distance between the corners (Fig. 5). We also computed the centroid of the six annotated landmarks and determined the face-crop region likewise as the square of width 1.5 times the largest distance of any two of the six landmarks, centered at the centroid location. All images are fed into iTracker at a resolution of 224×224 pixels, which means they undergo resizing from the original resolution. The eye crops are upsampled, while the face crop is downsampled with an anti-alias filter, using the `imresize` function in Matlab. We then apply iTracker to each frame in the video sequence, and the x-coordinate of the estimated gaze location over time is taken as the horizontal eye-movement trace.

While iTracker is designed to operate on video sequences of 30 fps, a temporal resolution above 50 fps is required for clinical applications [37]. Thus for this work, we used the rear-facing camera of the phone in slow-motion mode, which results in a frame rate of 240 fps and corresponding temporal resolution of approximately 4 ms. However, the higher frame rate also results in poorer image quality compared to 30 fps due to the reduction in exposure time. Recordings taken at 240 fps are dimmer than recordings taken at 30 fps. (This is not an issue with high-end image sensors such as those found in the Phantom high-speed camera. Phantom v2511 for example supports larger ISO (see Table I). Although there may be a trade-off between ISO and digital noise, Phantom v2511 also has a larger pixel size, which allows Phantom to produce less-noisier images even at higher ISO.) We discovered that in some challenging scenarios (e.g. the illumination was low or the subject was wearing glasses), the variations in the output of iTracker can be so large that the saccade onset becomes ambiguous.

To further understand the source of the variations, we tested the output of iTracker when fixing the face grid input and two of the other three inputs (left eye crop, right eye crop, and face crop) to be the first frame of the video. We discovered that the variations in the output will be the smallest when we only changed the input to the face layers. Since the receptive field in the cropped eye only contains parts of the eye, one potential explanation for the observation could be that the eye layers may be trained to learn detailed features in the eyes to fine-tune the gaze estimation. On the contrary, the receptive field in the cropped face may contain a full eye. That is, the face layers may be trained to learn more global features in the eyes. When the image becomes blurrier, the detailed eye features will be replaced by noise, which causes the eye layers more sensitive to noise than the face layers.

To address the comparatively low image quality at high frame rate, we propose the *iTracker-face* algorithm, for which we only use the face-related convolutional layers of iTracker (Fig. 4 blue layers). Although this choice does degrade the accuracy of the gaze estimation as discussed in [33], our objective is to determine if the gaze changes. Fig. 6 shows a sample eye-position trace using the iTracker and iTracker-face algorithms. In our application, iTracker-face generally has higher signal-to-noise ratio than iTracker.

B. Modeling Horizontal Eye-Movement Traces

To calculate saccade latency, it is necessary to determine the onset of the eye movement toward the target. In prior work, the saccade onset has commonly been defined as an increase in eye velocity above a predefined threshold [20], [37], such as 30 °/s, where the velocity is commonly determined through numerical differentiation and subsequent filtering of the raw eye-position tracing [38]. Such saccade-onset determination requires accurate measurement of gaze and is prone to significant error at low sampling rates [39].

Here, we instead propose to model the eye-position trace during a saccade task as a hyperbolic tangent of the form

$$\tilde{x}(t) = A + B \cdot \tanh\left(\frac{t - C}{D}\right)$$

and fit the model to the the eye-position tracing from 100 ms before to 500 ms after the stimulus presentation (Fig. 7). The fitting was performed using the nonlinear least-squares solver `lsqcurvefit` in Matlab to estimate the model parameters A, B, C, D. Using these optimal model parameters, we

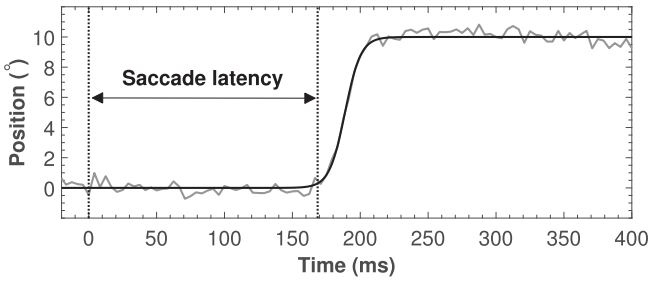

Fig. 7. Eye position as estimated by the iTracker-face algorithm (gray) and hyperbolic tangent fit (black). The dashed line at 0 s indicates the moment of stimulus presentation. The saccade onset is determined by an increase in saccade amplitude above 3% of the target saccade amplitude.

Fig. 8. A sample frame from each video taken under four distinct lighting conditions. From left to right, the pictures are arranged from the highest illuminance (278 Lux) to the lowest (26 Lux).

determine the saccade onset as the time when the best-fit solution exceeds 3% of the maximal saccade amplitude, which is independent of the velocity of the saccade.

In addition to generating well-behaved velocity tracings, this model-based approach has the benefit of providing a goodness-of-fit metric on the basis of which the reliability of saccade tracings can be evaluated in an automated manner, as the normalized root-mean-squared error (NRMSE) between the model fit and the eye-position trace quantifies the residual discrepancy between the two. Here, the normalization was done to the saccade amplitude (10° in our experiments). Measurements contaminated by excessive noise, artifact, or eye movements in the wrong direction typically result in a high NRMSE value while reliable measurements result in a low NRMSE. Thresholding the NRMSE allows for automated rejection of recordings in which the saccade onsets might have been erroneously detected or the measurements are subject to excessive variability, noise or artifact. After evaluating the sensitivity and specificity of saccadic eye-movement traces across a range of candidate NRMSE thresholds, we selected the NRMSE threshold and included in our analysis traces for which the optimal model fit resulted in a NRMSE$<$ 0.1 (see Section IV-B for more details).

IV. Algorithm Evaluations

A. Robustness of Eye Tracking Algorithms

To determine the robustness of iTracker and iTracker-face under a variety of environmental conditions that may be encountered outside the well-controlled clinical setting for eye-movement measurements, we compared the performance of the algorithms on video sequences of subjects with and without glasses and under various ambient lighting conditions. Two illumination-adjustable LED panel lights were used to vary the illumination during the recording sessions. In total, four distinct lighting conditions were tested: (1) room light switched on in addition to the panel lights set to high (278 Lux); (2) room light switched on without additional lighting support from the LED panels (220 Lux); (3) room light switched off and the panel lights set to medium (54 Lux); and (4) room lights switched off and the panel lights set to low (26 Lux). Illuminance was measured at the participant's face using an LT40 LED Light Meter (Extech Instruments). Figure 8 shows how the lighting conditions affect image brightness. Five subjects contributed 120 saccade tasks under each of the four lighting conditions with and without glasses, for a total of eight test conditions per subject.

The video sequences were processed with both iTracker and iTracker-face, and the 9,600 resultant eye-movement traces were each reviewed by two annotators. Each annotator independently determined if a trace represented a horizontal saccade movement and had sufficiently high signal-to-noise to allow for credible saccade-onset determination. Traces that met these criteria were labeled 'good'; all other traces were labeled 'bad'. Traces labelled as 'bad' were typically interrupted by blinks, initially directed toward the opposite direction of stimulus presentation, or had a low signal-to-noise ratio. To assess the annotator agreement, we computed both the accuracy (fraction of annotations in which both annotators agreed) and Cohen's kappa coefficient (κ). The algorithm with the highest fraction of 'good' saccade traces, as judged by both annotators, across the different environmental conditions was deemed the more robust algorithm.

Figure 9 reports the inter-rater annotation accuracy, broken down by 'agreed good' and 'agreed bad', for both algorithms and each of the eight environmental conditions tested. The average annotation accuracy was 94.1% for eye-movement traces generated by iTracker-face and 86.8% for iTracker, with corresponding Cohen's κ values of 0.802 and 0.730, respectively. These results indicate excellent inter-rater agreement for the overall annotation task, which means that their judgment can be used as a benchmark. Their annotations also reveal that important trends exist between algorithms and across environmental conditions. The inter-rater agreement is lower when participants wear glasses and tends to decline with decreasing illuminance. For example, at the lowest illuminance level (26 Lux) and with participants wearing glasses, the annotators agreed in their label of 'good' in over 40% of the traces generated by iTracker-face. In contrast, their agreement of what constitutes a good saccade trace was less than 8% of the traces generated by iTracker. Obviously poor illumination conditions result in image sequences with lower contrast which makes it harder to detect eye features and subtle eye movements. A closer inspection of the video sequences also revealed that glasses, especially those with dark rims, tend to cast shadows that can obscure the eye regions.

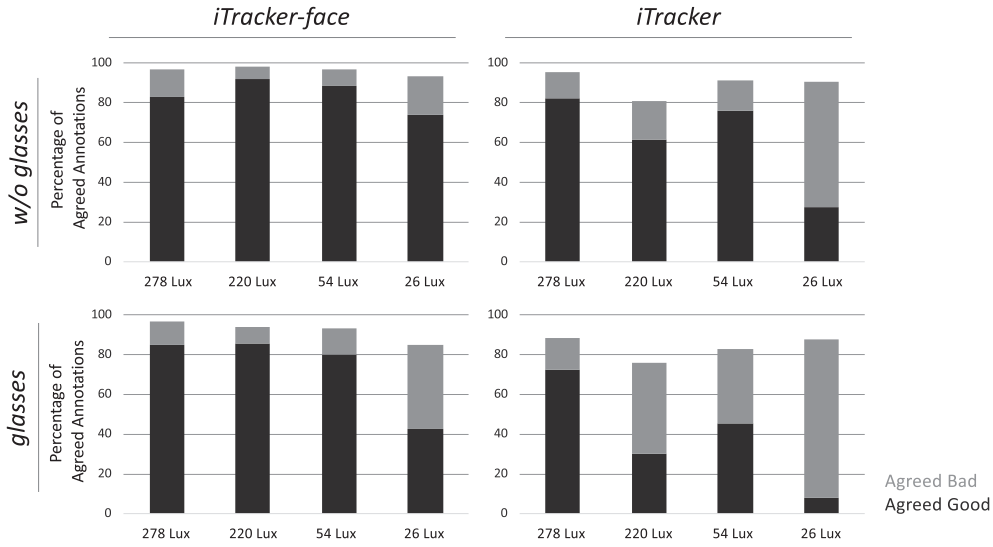

Fig. 9. Annotation accuracy broken down for each of the eight environmental conditions tested per algorithm. The accuracy (or percentage of agreed annotations) is additionally broken down into the fraction of agreed-good and agreed-bad eye-movement traces between two annotators.

Additionally, some glasses have lenses with high reflectivity that make the eyes even less visible and therefore difficult to track.

Across all eight conditions tested, the average fraction of traces judged as good by both annotators was consistently and significantly higher for traces generated by iTracker-face (78.9%) than for those generated by iTracker (50.7%). We conclude from this analysis that across all environmental conditions tested, iTracker-face is the more robust algorithm of the two and therefore formed the basis of all subsequent results reported here.

B. Automation of Saccade-Onset Detection

Annotation of the 9600 eye-movement traces took each annotator about 12 hours to complete. Since our goal is to leverage smartphones to make eye-movement recordings and analyses widely available and ubiquitous, visual inspection of individual tracings is not an option. Having identified iTracker-face as the more robust of the two algorithms for iPhone-based eye-movement tracking, we applied the tanh model to the resultant eye-position traces to estimate saccade onset. To evaluate the usefulness of the NRMSE as an automated metric to flag bad saccades, we used the expert-annotator labels as the ground truth for all iTracker-face derived traces described in the previous section and swept the NRMSE threshold to generate a receiver-operating characteristic (ROC) curve. By separately considering each annotator's judgment as the ground truth, we obtained two ROC curves (Fig. 10), one for each annotator, and generated associated 95% confidence intervals (CI) for the true positive rate by stratified bootstrapping over 2000 replicates at fixed false positive rate [40]. The two resultant ROCs tracked each other closely and achieved an area under the curve (AUC) of 0.923 (95% CI: 0.913 – 0.932) and 0.933 (95% CI: 0.923 – 0.943), respectively. If we consider all traces with a NRMSE$<$ 0.1 as 'good' saccades, we achieve average true positive rates of 0.87 and 0.86 and average false positive rates of 0.20 and 0.16 for the first and second annotator, respectively. In the following, we selected an NRMSE of 0.1 as the threshold.

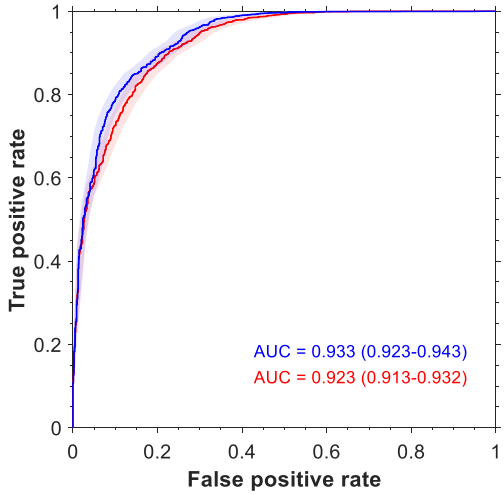

Fig. 10. Performance of model-based fitting in classifying saccades. The adjudications of two annotators were taken as the ground truth, with the solid lines being the corresponding mean ROC curves. The shaded areas indicate the confidence intervals for the true positive rate. The parentheses mark the 95% confidence intervals for the areas under the curves.

C. Comparison Across Cameras

To verify that recordings from consumer-grade cameras can lead to similar saccade-latency statistics as those obtained from recordings of high-end, research-grade cameras, we took simultaneous recordings on four subjects using a low-cost, consumer-grade camera (iPhone 6) and a research-grade camera (Phantom v2511, see Table I for their specifications).

Fig. 11 shows the resulting saccade-latency distributions obtained using the iTracker-face algorithm and the model-based

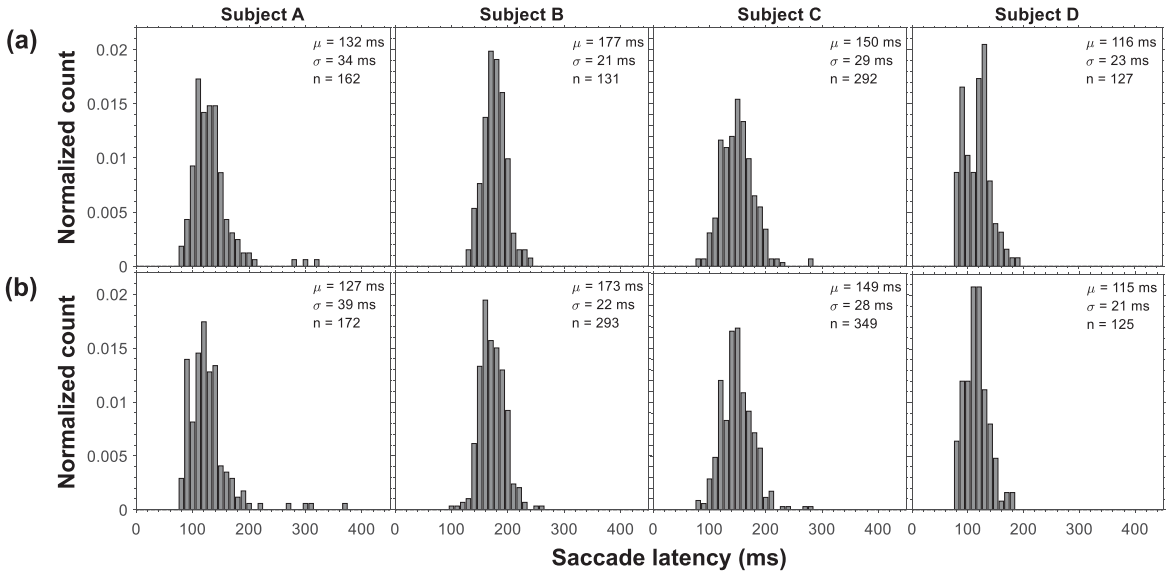

Fig. 11. Saccade-latency distributions from four subjects obtained from video recordings using (a) the iPhone 6 and (b) a Phantom v2511 high-speed camera.

onset detection. The inclusion of the high-speed camera in the recording set-up resulted in increased distances between the subject and the cameras, as well as between the subject and the laptop's screen. The increased distances result in a smaller horizontal eye movement, which in turn produce slightly noisier, but acceptable, eye movement traces. Fig. 11 demonstrates that the distributions from both cameras are consistent, with negligible differences in the mean saccade-latency values and associated standard deviations between the two recording systems.

D. Face-Crop Automation

To fully automate the signal-processing pipeline of Fig. 3, we replaced the manual face annotation and cropping (Fig. 5) with an automated face-detection step. As mentioned in Section II, with the head supported by the chinrest, we can expect the position of the face to remain relatively stable throughout a sequence of saccade tasks and the manually determined face region to remain valid throughout the subsequent frames of a video recording. To automate the face-region determination, we used the Viola-Jones face detector [41] and evaluated the changes in the estimated saccade latencies after this automation on 158 sessions of recordings. The mean absolute differences in the mean per-session saccade latencies with an NRMSE$<$ 0.1 was 1.10 ms with an associated standard deviation of 1.24 ms (Fig. 12). We therefore conclude that automating the face-detection step does not materially affect the saccade-latency determination in normal subjects. This result may be understood by considering that the convolutional layers in iTracker are trained to properly adjust gaze estimation under translation and scaling differences in the cropped face. As a result, the shape of the resulting eye-movement traces are hardly changed given slight differences in the cropped regions of the face.

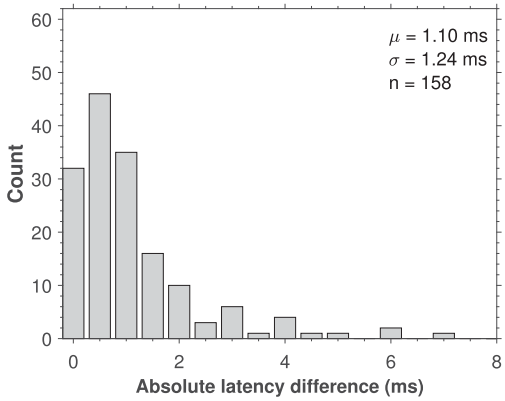

Fig. 12. The absolute difference in mean saccade latencies between face crop based on manual face annotation and automated face detection using the Viola-Jones algorithm [41].

E. Chinrest Dependence

Ideally, we would like to enable eye-movement capture and analysis without the need for restraining the head. Without the chinrest in place, the assumption of limited head movement throughout the (approximately) two-minute 40-saccade sequence is bound to be violated. However, the assumption might still be reasonable over the course of a single saccadic eye movement, of which we typically analyze 600 ms (from 100 ms before till 500 ms after stimulus presentation). To test this hypothesis, we conducted two sessions of video recordings in four subjects each with and without the participants' heads resting on the chinrest (16 sessions in total). We applied the Viola-Jones face detector to the first frame of each individual saccade tracing and used the detected face region from the first frame and applied it to every subsequent frame. If there had been any significant head movements within a single saccade trial,

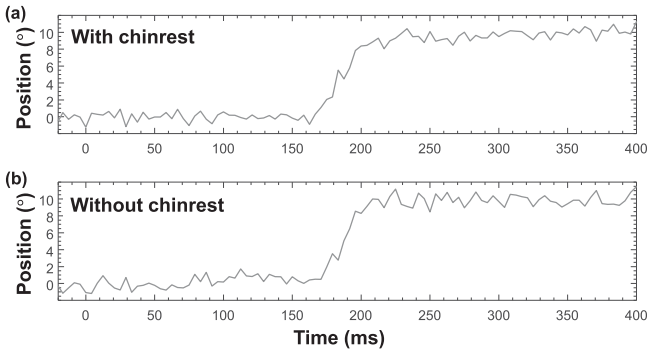

Fig. 13. Two examples of saccadic eye-movement traces in the same subject. (a) Recording with chinrest, and (b) recording without chinrest. They have a comparable signal-to-noise level.

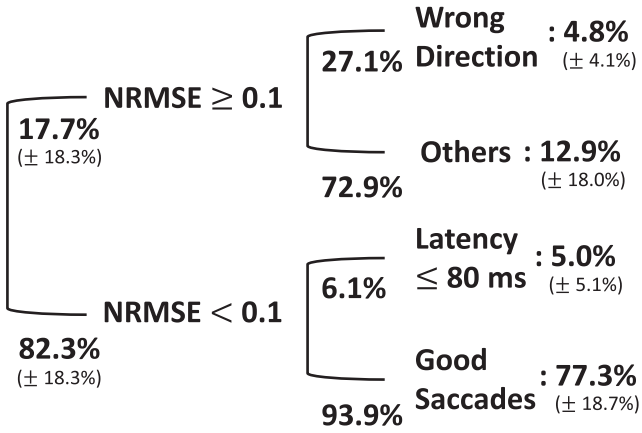

Fig. 14. Breakdown of over 19,000 saccade measurements by saccades initiated into the wrong direction, noisy saccades, potentially predictive/express saccades, and good saccades.

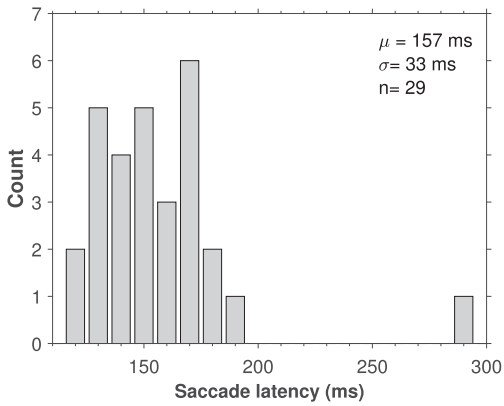

Fig. 15. Distribution of the mean saccade latencies from 29 self-reported healthy individuals, including one subject whose mean saccade latency is 290 ms.

we would have expected the tanh model to no longer attain low NRMSE fits. When the Viola-Jones face detector was applied to iTracker-face derived eye-movement traces on recordings obtained with and without chinrest, most of the traces have comparable signal-to-noise (Fig. 13). After confirming that the null hypothesis of normally distributed mean saccade latency cannot be rejected at the 0.05 level (using the Anderson-Darling test), we performed a formal analysis of variance (ANOVA) to assess whether a significant difference existed between mean saccade latencies measured with and without chinrest. The ANOVA null hypothesis of a significant difference was rejected ($p = 0.59$). We therefore conclude that in our cohort of self-reported healthy volunteers, the chinrest is not essential to obtaining recordings of sufficient quality for saccade-onset detection and saccade-latency determination.

The selection of iTracker-face to generate the eye-movement tracings, the NRMSE threshold value of 0.1 to select traces for inclusion in our analysis, and the Viola-Jones algorithm for automated face detection on the first frame of each saccade task video sequence completes the automation of the saccade-latency determination pipeline of Fig. 3. In the next section, we apply this pipeline to determine the intra- and inter-subject variability in saccade-latency measurements obtained from video sequences of self-reported healthy subjects, and explore the statistical modeling of the saccade-latency distributions.

V. Data Analysis

A. Saccade-Latency Determination in Healthy Individuals

We recorded 19 200 saccadic eye movements across 160 experimental sessions in 29 self-reported healthy subjects (20 males, 9 females; median age: 27 years; age range: 22–64 years), including five or more repeat recording sessions in a subset of eleven subjects. In two recording sessions, the Viola-Jones algorithm failed to detect the face of the subject, so the results presented here are based on 158 experimental sessions in 29 subjects.

Applying the NRMSE<0.1 criterion to accept saccadic eye movements for analysis resulted in an average retention rate of 82.3% (100 of the 120 saccadic eye movements per experiment). Consistent with prior work [16], [37], [42],we further excluded from our analysis saccade latencies of 80 ms or less. This censoring excludes anticipatory eye movements but also removes possible express saccades [13], [43]. With this additional exclusion criterion in place, the average fraction of good saccades per recording session was 77% with an associated standard deviation of $\pm$ 19% (see Fig. 14). Of the 17.7% of eye-movement recordings rejected from analysis due to NRMSE$\geq$ 0.1, 27.1% were initiated into the wrong direction and represent an error rate that by itself might carry pathognomonic information [6]. The remaining 72.9% of the rejected eye movement had generally low signal-to-noise. Thus, overall, only about 13% of all saccadic eye movements were rejected because of excessive noise.

When we aggregated the saccade latency measurements greater than 80 ms and NRMSE$<$ 0.1 for each subject, the mean latencies across the 29 subjects typically ranged from 120 ms to 200 ms (Fig. 15), with one subject having a mean saccade latency of 290 ms. (Review of the latter subject's video sequences, eye-movement traces, and health questionnaire did

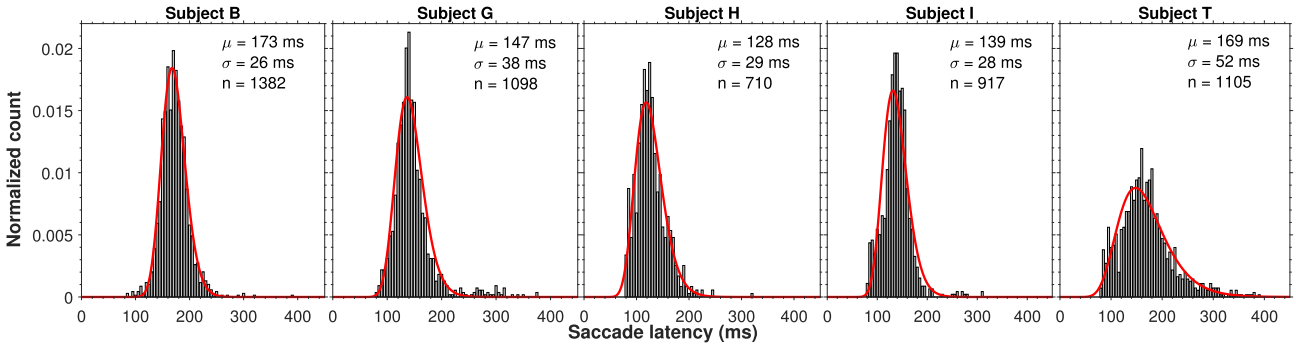

Fig. 16. Saccade latency distributions for five self-reported healthy individuals. μ is the sample mean, σ is the associated sample standard deviation, and n is the total number of observations. Saccade latencies below 80 ms were censored. The estimated log-normal probability density functions are shown in red.

not provide a credible reason to exclude this subject from our analysis.) While it is common practice in clinical studies to only report the population mean or median saccade latency, such aggregation results in loss of information encoded in each subject's full saccade latency distribution.

Fig. 16 shows normalized saccade-latency distributions for five subjects, selected to illustrate the range of intra- and inter-subject variation among our study cohort. The distributions show variable degrees dispersion and skewness, with some subjects having a significant fraction of latencies above 200 ms.

It has been suggested that reaction times follow log-normal distributions [44]. We tested this hypothesis on our recordings by fitting a log-normal distribution to the saccade latency distributions of the individual recording sessions, and also to the saccade latency distribution of each subject for which we aggregated each subject's measurements across recording sessions. The log-normal distributions were truncated at 80 ms to reflect the censoring we imposed on the minimum saccade latency. The Kolmogorov-Smirnov test was used with the significance level set to 0.05 to test the null hypothesis that the saccade-latency distributions can be described by a truncated log-normal distribution. Of the 158 individual saccade-latency distributions (one for each recording session) across all subjects, 155 (or 98.1%) distributions were not significantly different from a log-normal distribution ($p < 0.05$). When the data from across different recordings sessions were aggregated into a single distribution for each subject, 26 out of the 29 (89.7%) distributions were not significantly different from a log-normal distribution ($p < 0.05$).

B. Test-Retest Reliability

If the subject condition is stable (healthy subjects, for example), we want our saccade-latency measurement to be consistent across sessions. To assess the test-retest reliability of our approach to saccade-latency determination, we computed the intraclass correlation coefficient (ICC) of the per-session mean saccade latency in eleven subjects that participated in at least five repeat recording sessions. As in the previous section, individual saccadic eye movements were included in the analysis if the associated NRMSE < 0.1 and the measured saccade latency exceeded 80 ms.

As suggested in [45], we used a repeated-measure, two-way ANOVA approach in which subject identity and session number were used as categorical variables, and the outcome variable was the mean (per-session) saccade latency. The effect of repeat experimental session turned out to be non-significant ($p = 0.78$), suggesting that no significant trend existed across sessions that ought to be accounted for. Using the psych library in R [46], we computed the ICC (ICC 3,1 in the Shrout and Fleiss [47] convention) of the mean saccade latency for each of the five sessions in the eleven subjects. The resultant ICC value was 0.76 (95% CI: 0.55-0.92), generally indicating good [48] to excellent [49] test-retest reliability.

To put this ICC value into further context, it is informative to compare it to the values reported in the literature for pro-saccade tests using specialized eye tracking equipment. Using the Eye trac Model 210 (ASL, Waltham, MA, USA), Roy-Byrne *et al.* [50] report ICC values between 0.61 and 0.75 for mean latency for a visually guided saccade task in healthy subjects. Blekher and co-workers [51] likewise used an infrared illumination based eye tracking system (EyelinkII, SR Research Ltd., Ontario, Canada) in a case-control study of subjects at risk of Huntington's disease. For the control arm of the study, the authors report an ICC of 0.71 (95% CI: 0.55-0.97) for a visually guided gap pro-saccades task. The test-retest reliability of our approach to saccade-latency determination using a consumer-grade camera therefore compares very favorably to the reliability reported using specialized eye tracking equipment.

C. Sample-Size Considerations

We initially designed our experimental session, consisting of three consecutive 40-saccade trials, on the basis of a trade-off between session duration and number of saccade latency measurements acquired. Each 40-saccade trial takes about two minutes to perform. Each experimental session takes between six and ten minutes, depending on how much time a subject would like to take between consecutive trials. Assuming that the saccade latency of healthy subjects is sampled from static underlying distributions, like the ones in Fig. 16 for example, we can conduct a bootstrapping exercise to determine how many individual saccade latency measurements ought to be taken per

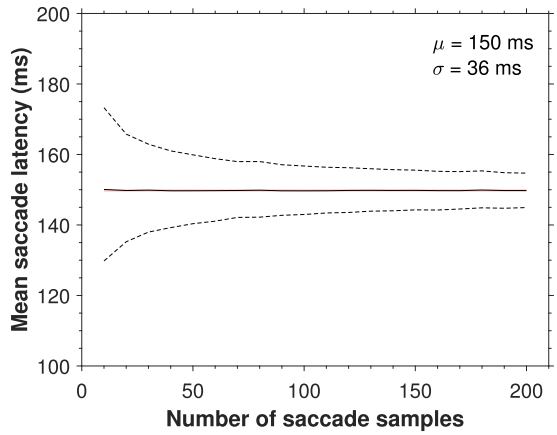

Fig. 17. Bootstrap determination of mean saccade latency (solid black line) and associated 95% CI of the mean (dashed lines) as a function of number of samples drawn from a subject's full saccade latency distribution. The ground truth mean saccade latency is shown in red. The mean and 95% CI at each sample number are based on 4000 bootstrap realizations.

session to obtain a reasonably accurate estimate of mean saccade latency. We conducted such bootstrapping exercise on the aggregated saccade latency distributions in each of the ten subjects for which we had at least ten recording sessions. Fig. 17 shows a representative example in which the mean saccade latency and associated 95% confidence interval were determined as a function of number of samples drawn from the subject's full saccade latency distribution.

On average, single experimental session with about 50 good saccadic eye movement measurements allows for accurate and precise (to within less than 10 ms in 95% of the cases) determination of mean saccade latency. Such precision may be required to differentiate normal saccade latency from the increased latencies reported in Alzheimer's disease [37], for example. Taking into account our result that on average only about 77% of the tracking tasks result in good saccadic eye movement measurements, we arrive at around 65 individual tracking tasks to include in a single experimental session, or about half of the number of saccadic eye movements currently included in our session design.

Reducing the number of visual tracking tasks per session to 65 also reduces the total recording length per session to slightly more than three minutes, which is less than the average length of a typical Youtube video [52]. This recording length is not an onerous imposition on a subject's time and is eminently compatible with subjects providing one or two such recordings per day thus aiding our goal of bringing high-accuracy determination and longitudinal tracking of saccade latency to a broad patient population through the use of smartphone technology.

For disease tracking, another consideration is how many daily recordings should be aggregated to balance averaging of random variations with detecting neurocognitive decline. Among the dementias, for example, the prion diseases have one of the fastest rates of decline, with the timespan from initial disease diagnosis to death sometimes only covering a few months to a year [53]. Aggregating seven consecutive daily recordings of 65 saccade measurements each, for example, and advancing the averaging window one day at a time in a sliding manner, would result in around 350 good saccade measurements to determine a personalized saccade latency distribution while maintaining a temporal resolution that would allow for disease tracking of even the fastest neurodegenerative diseases. Longer averaging times can be considered for more slowly progressing neurodegenerative diseases to allow for more averaging of day-to-day variation in eye movement features.

VI. Discussion

The successful execution of an eye movement relies on a complex interplay of cognitive and motor function. Neurodegenerative diseases affect the neural circuits responsible for these functions, altering several eye movement features during neurocognitive decline. In this work, we developed and validated an approach for assessing one such feature, saccade latency, without the need for specialized equipment (such as infrared illumination, chinrest and research-grade cameras) or specific environmental conditions. This approach enables measurement and analysis of saccade latency outside of the clinical environment, hence, paving the way for large-scale data collection.

A. Method Development

Several technological challenges needed to be overcome to allow for repeat saccade latency measurements outside a specialized clinical environment. Among these technological challenges were the reliability on infrared (IR) light to estimate the position of the eye and the use of research-grade cameras that yield distinct images of the eyes. Here, we demonstrated the feasibility of deriving saccade latency from consumer-grade cameras, such as smartphones, along with a model-based approach to determining the onset of the saccade that provides a metric for automated rejection of eye-movement traces of poor quality. To extract the position of the eye from each frame in a video sequence, we proposed iTracker-face, a modified version of a deep convolutional neural network for gaze estimation on smartphones that does not rely on IR illumination. In our application, iTracker-face is more robust to lower image quality than iTracker, providing eye-movement traces with a higher signal-to-noise ratio. Once the eye-movement traces are extracted with iTracker-face, our eye movement model is fitted to the individual traces to determine the onset of the eye movement toward the target. This model-based approach has the added benefit of providing a goodness-of-fit metric that allows for automated rejection of unreliable data, an instrumental contribution toward making saccade latency determination broadly available as large cohorts of patients and self-reported healthy subjects start recording saccadic eye movements on a continuous basis.

B. Method Evaluation

Because the environmental conditions outside of a typical clinical setting are variable, the evaluation of the robustness of our eye tracking algorithm is paramount and strengthens our ability to measure saccade latency in complex real-world

scenarios. Our robustness evaluation shows that iTracker-face was consistently and significantly more robust than iTracker across all testing conditions, as ascertained by two annotators that manually reviewed 9600 eye-movement traces. Because the agreement between annotators was high (as given by the accuracy and Cohens kappa coefficient), their annotations were used to determine an optimal threshold value for the NRMSE that automatically eliminates eye-movement traces that provide unreliable saccade latency estimates. Our evaluation of the sensitivity and specificity of this approach suggests very high sensitivity and specificity for automated signal quality determination compared against human annotators, and in a variety of environmental conditions that are expected to be encountered in everyday recordings.

One important contribution to this field is our demonstrated ability to obtain essentially the same recording quality with and without chinrest support during the video recordings. A formal ANOVA test confirmed that there are no significant differences between saccade latencies measured with and without a chinrest. The removal of the chinrest further enhances the flexibility of our system and eliminates the need for specialized equipment to measure saccade latency. While this result might hold more generally for self-reported healthy subjects, further image processing techniques might be required for certain neurodegenerative diseases, such as Parkinson's disease, that are known to lead to tremors. Our decision to include the Viola-Jones face detector to our signal-processing pipeline might help mitigate this effect, but this remains to be seen. Finally, we evaluated the test-retest reliability of our proposed approach to determining saccade latency and compared it to the reliability reported using specialized eye tracking equipment, showing that our system achieves comparable results. A high reliability demonstrates that our low-cost approach to saccade-latency determination yields consistent results over time in self-reported healthy subjects – a desired attribute for a system built to track the progression of neurodegenerative diseases.

C. Intra- and Inter-Subject Variability

In order to use saccade latency to track neurodegenerative disease progression, it is important to understand the intra- and inter-subject variability among self-reported healthy subjects to put into context the changes seen in patients with neurodegenerative disease. In our work, we measured more than 19 200 saccade latencies in 29 subjects (Fig. 15), a significantly larger number compared to the values reported in clinical studies, ranging from 8 – 30 saccade latencies [14], [21], [37], [54]. These recordings were enabled by the accessible nature of our measurement system that allows for 120 saccade measurements in less than ten minutes. Considering the criteria to reject certain saccade latencies, we are still able to retain, on average, a sizable amount (77%) of data per session across subjects. With this amount of data, we observed that the intra- and inter-subject variability in saccade latency are quite substantial, with mean saccade latencies ranging from 120 ms to 290 ms (Fig. 15) and standard deviations from 26 ms to 52 ms (Fig. 16). Prior clinical studies comparing saccade latency measurements from patients with neurodegenerative diseases to those obtained in age-matched health control subjects have reported differences in group means or medians between 16 ms and 200 ms [16], [37]. The approach to saccade latency measurements developed here allow us to resolve such differences.

The rich information regarding the distinctive shape and parameters of the individual distributions is lost when saccade latency values are pooled. As seen in Fig. 16, some individuals have a tendency to make more saccades with shorter latencies and others to make more saccades with longer latencies. In combining all the data into a single distribution, these individual characteristics – that have been linked to specific brain pathologies [55], [56] – are lost. Saccade latency intra-subject variability is also lost when data is pooled. If instead the information regarding this variability were preserved, it could be used as a feature to assess the cognitive state of a subject. For example, some studies suggest that intra-subject variability is larger in some conditions compared to normal subjects [42], [57]. Our accessible, low-cost measurement system enables widespread data collection and hence avoids having to combine data from different subjects, allowing us to preserve the distinctive information in each individual saccade latency distribution (Fig. 16).

In addition to the large intra- and inter-subject variability, we observed that the saccade latency distribution of the majority of the subjects may be modeled as a log-normal distribution. This observation is consistent with [58], in which neural mechanisms are discussed that might give rise to log-normally distributed reaction times. It might therefore be sufficient to characterize individual saccade-latency distributions using the two parameters of a log-normal distribution ($\log -\mu$ and $\log -\sigma$) and analyze how these parameters change through time. Lastly, we performed a bootstrapping experiment to determine what is the minimum number of saccade latencies that are needed to obtain a reasonable estimate of mean saccade latency and discovered that around 65 measurements allow us to estimate the mean saccade latency with a within-10-ms precision. This suggests that we can minimize the recording time in a data collection session to about four minutes and still have sufficient precision to distinguish between healthy subjects and patients. Four minutes per day is a negligible amount of time, which we believe will incentivize frequent measurements and enable longitudinal tracking of saccade latency.

For disease tracking, another consideration is how many recordings should be aggregated to balance averaging of random variations with detecting neurocognitive decline. Unfortunately, there are few studies that track the longitudinal changes in saccade latency among patients [19], [59], especially within the same cohort. Because the data in these studies was collected in clinical environments and the analyses usually involved manual removal of outliers, longitudinal measurements are sparse (typically with a $>=$ 6-month interval), which limits the types of diseases they can track. For example, among the dementias, the prion diseases have one of the fastest rates of decline, with the timespan from initial disease diagnosis to death sometimes only covering a few months to a year [53]. Current methods cannot render a sufficiently timely assessment on patient state. With our approach, however, around 350 good saccades can be

measured over a week with a less-than-4-minute recording each day. These saccades may provide an adequate characterization of a personalized saccade latency distribution. By analyzing how this distribution changes over weeks, we may be able to track the progression of prion diseases. Longer averaging times can be considered for more slowly progressing neurodegenerative diseases to allow for more averaging of day-to-day variation in eye movement features.

D. Limitations and Future Work

The limitations of this work include the fact that our current recording setup requires a laptop that shows the visual task and an additional display to calculate saccade latency. Currently, we are implementing the visual task on the iPhone itself, therefore minimizing the required equipment to this single device and reducing saccade latency delays introduced by the two-display set-up. For patient privacy reasons, our long-term goal is to perform all the data analysis directly on the iPhone, which requires us to build an energy-efficient version of iTracker-face that runs smoothly on devices that are power and memory constrained.

Another venue of future exploration is our ability to measure a richer set of eye movement features, such as, for example, the gain and the peak velocity of a saccade. These features have proven to be significantly different between healthy subjects and patients afflicted with neurodegenerative diseases [6]. To measure the gain, we need to confirm the accuracy of the estimated gaze position in iTracker-face. Accurate velocity measurements depend on the accuracy in measuring the gain. Incorporating these and other eye movement measurements into our system and signal processing pipeline will enhance our ability to track neurocognitive decline and help with the differential diagnosis of neurodegenerative diseases.

VII. Conclusion

Our work here presents a method to measure saccade latency outside of the clinical environment using a consumer-grade camera. A thorough algorithm evaluation showed that iTracker-face, along with the tanh model for saccade-onset determination, is robust to varying recording conditions, allows for automated outlier rejection, and produces saccade latency distributions that are very similar to those obtained from a high-end, high-speed reference camera. Furthermore, our implementation of the tanh model allows for automated rejection of bad saccades and therefore enables efficient large-scale data analysis. Because of this efficiency, we collected over 19 000 saccade latency measurements across 29 self-reported healthy volunteers and observed that their saccade-latency distributions have distinctive shapes, with different means and standard deviations. A deeper understanding of these differences is essential to put into perspective the saccade-latency changes seen in patients with neurocognitive disease. An evaluation of the test-retest reliability of our system showed that our approach is capable of determining consistent saccade latency values over time in self-reported healthy subjects. These contributions pave the way to expanding saccade latency measurements to a broad population for tracking of neurologic and neurodegenerative disease progression.

Acknowledgment

The authors would like to thank Prof. Lydia Bourouiba of MIT's Fluid Dynamics in Disease Transmission Laboratory and Dr. James Bales of MIT's Edgerton Center for their assistance with the high-speed videography.

References

[1] B. S. O. Diniz, M. S. Yassuda, P. V. Nunes, M. Radanovic, and O. V. Forlenza, "Mini-mental state examination performance in mild cognitive impairment subtypes," *Int. Psychogeriatrics*, vol. 19, no. 4, pp. 647–656, 2007.

[2] S. Hoops *et al.*, "Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease," *Neurology*, vol. 73, no. 21, pp. 1738–1745, 2009.

[3] A. J. Mitchell, "A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment," *J. Psychiatric Res.*, vol. 43, no. 4, pp. 411–431, 2009.

[4] H. Posner *et al.*, "Outcomes assessment in clinical trials of Alzheimer's disease and its precursors: Readying for short-term and long-term clinical trial needs," *Innov. Clin. Neurosci.*, vol. 14, no. 1/2, pp. 22–29, 2017.

[5] P. D. Harvey *et al.*, "Performance-based and observational assessments in clinical trials across the Alzheimer's disease spectrum," *Innov. Clin. Neurosci.*, vol. 14, no. 1/2, pp. 30–39, 2017.

[6] T. J. Anderson and M. R. MacAskill, "Eye movements in patients with neurodegenerative disorders," *Nature Rev. Neurol.*, vol. 9, pp. 74–85, 2013.

[7] J. Hlavnička, R. Čmejla, T. Tykalová, K. Šonka, E. Růžička, and J. Rusz, "Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behavior disorder," *Sci. Rep.*, vol. 7, no. 12, pp. 1–13, 2017.

[8] J. Hanuška *et al.*, "Eye movements in idiopathic rapid eye movement sleep behavior disorder: High antisaccade error rate reflects prefrontal cortex dysfunction," *J. Sleep Res.*, 2018, Art. no. e12742.

[9] J. G. Holden *et al.*, "Prodromal Alzheimer's disease demonstrates increased errors at a simple and automated anti-saccade task," *J. Alzheimer's Disease*, vol. 65, pp. 1209–1223, 2018.

[10] R. J. K. Jacob and K. S. Karn, "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises," in *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, J. Hyona, R. Radach, and H. Deubel, Eds., Oxford, U.K.: Elsevier, 2003, pp. 547–470.

[11] P. M. Corcoran, F. Nanu, S. Petrescu, and P. Bigioi, "Real-time eye gaze tracking for gaming design and consumer electronics systems," *IEEE Trans. Consum. Electron.*, vol. 58, no. 2, pp. 347–355, May 2012.

[12] P. Majaranta and A. Bulling, "Eye tracking and eye-based human–computer interaction," in *Advances in Physiological Computing*, S. H. Fairclough and K. Gilleade, Eds., London, U.K.: Springer, 2014, pp. 39–65.

[13] R. J. Leigh and D. S. Zee, "The saccadic system," in *The Neurology of Eye Movements*. Oxford, U.K.: Oxford Univ. Press, 2015, ch. 4, pp. 169–288.

[14] S. Garbutt *et al.*, "Oculomotor function in frontotemporal lobar degeneration, related disorders and Alzheimer's disease," *Brain*, vol. 131, no. 5, pp. 1268–1281, 2008.

[15] R. J. Molitor, P. C. Ko, and B. A. Ally, "Eye movements in Alzheimer's disease," *J. Alzheimer's Disease*, vol. 44, no. 1, pp. 1–12, 2015.

[16] U. P. Mosimann, R. M. Müri, D. J. Burn, J. Felblinger, J. T. O'Brien, and I. G. McKeith, "Saccadic eye movement changes in Parkinson's disease dementia and dementia with Lewy bodies," *Brain*, vol. 128, no. 6, pp. 1267–1276, 2005.

[17] R. Perneczky, B. C. P. Ghosh, L. Hughes, R. H. S. Carpenter, R. A. Barker, and J. B. Rowe, "Saccadic latency in Parkinson's disease correlates with executive function and brain atrophy, but not motor severity," *Neurobiol. Disease*, vol. 43, no. 1, pp. 79–85, 2011.

[18] J. M. Chambers and T. J. Prescott, "Response time for visually guided saccades in persons with Parkinson's disease: A meta-analytic review," *Neuropsychologica*, vol. 48, no. 4, pp. 887–889, 2010.

[19] C. A. Antoniades, Z. Xu, S. L. Mason, R. H. S. Carpenter, and R. A. Barker, "Huntington's disease: Changes in saccades and hand-tapping over 3 years," *J. Neurol.*, vol. 257, no. 11, pp. 1890–1898, 2010.

[20] R. Shafiq-Antonacci, P. Maruff, C. Masters, and J. Currie, "Spectrum of saccade system function in Alzheimer's disease," *Arch. Neurol.*, vol. 60, pp. 1275–1278, 2003.

[21] A. L. Boxer *et al.*, "Saccade abnormalities in autopsy-confirmed frontotemporal lobar degeneration and Alzheimer's disease," *Arch. Neurol.*, vol. 69, no. 4, pp. 509–517, 2012.

[22] T. J. Crawford, A. Devereaux, S. Hingham, and C. Kelley, "The disengagement of visual attention in Alzheimer's disease: A longitudinal eye tracking study," *Frontiers Aging Neurosci.*, vol. 7, no. 118, pp. 1–10, 2015.

[23] J. K. H. Lim *et al.*, "The eye as a biomarker for Alzheimer's disease," *Frontiers Neurosci.*, vol. 10, no. 536, pp. 1–14, 2016.

[24] S. Rivaud-Péchoux, M. Vidailhet, J. Brandel, and B. Gaymard, "Mixing pro- and antisaccades in patients with parkinsonian syndromes," *Brain*, vol. 130, no. 1, pp. 256–264, 2006.

[25] C. Bonnet *et al.*, "Eye movements in ephedrone-induced parkinsonism," *PloS One*, vol. 9, no. 8, pp. 1–8, 2014.

[26] S. J. Tabrizi *et al.*, "Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: Cross-sectional analysis of baseline data," *Lancet Neurol.*, vol. 8, no. 9, pp. 791–801, 2009.

[27] E. R. Dorsey, S. Papapetropoulos, M. Xiong, and K. Kieburtz, "The first frontier: Digital biomarkers for neurodegenerative diseases," *Digit. Biomarkers*, vol. 1, pp. 6–13, 2017.

[28] E. R. Dorsey, A. M. Glidden, M. R. Holloway, G. L. Birbeck, and L. H. Schwamm, "Teleneurology and mobile technologies: The future of neurological care," *Nature Rev. Neurol.*, vol. 14, no. 5, pp. 285–297, 2018.

[29] National Academies of Sciences, Engineering, and Medicine, *Harnessing Mobile Devices for Nervous System Disorders: Proceedings of a Workshop*. Washington, DC, USA: Nat. Acad. Press, 2018.

[30] H.-Y. Lai, G. Saavedra-Peña, C. G. Sodini, T. Heldt, and V. Sze, "Enabling saccade latency measurements with consumer-grade cameras," in *Proc. IEEE Int. Conf. Image Process.*, 2018, pp. 3169–3173.

[31] G. Saavedra-Peña, H.-Y. Lai, V. Sze, and T. Heldt, "Determination of saccade latency distributions using video recordings from consumer-grade devices," in *Proc. IEEE Eng. Med. Biol. Conf.*, 2018, pp. 953–956.

[32] "Psychophysics Toolbox 3." [Online]. Available: http://psychtoolbox.org/. Accessed on: Feb. 9, 2018.

[33] K. Krafka *et al.*, "Eye tracking for everyone," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2176–2184.

[34] Q. Huang, A. Veeraraghavan, and A. Sabharwal, "Tabletgaze: Dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets," *J. Mach. Vis. Appl.* vol. 28, no. 5-6, pp. 445–461, 2017.

[35] E. Wood and A. Bulling, "EyeTab: Model-based gaze estimation on unmodified tablet computers," in *Proc. Symp. Eye Tracking Res. Appl.*, 2014, pp. 207–210.

[36] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "It's written all over your face: Full-face appearance-based gaze estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 2299–2308.

[37] T. J. Crawford *et al.*, "Inhibitory control of saccadic eye movements and cognitive impairment in Alzheimer's disease," *Biol. Psychiatry*, vol. 57, no. 9, pp. 1052–1060, 2005.

[38] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer, "Eye-tracker hardware and its properties," in *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford, U.K.: Oxford Univ. Press, 2011, ch. 2, pp. 48–49.

[39] D. J. Mack, S. Belfanti, and U. Schwarz, "The effect of sampling rate and lowpass filters on saccades—A modeling approach," *Behav. Res.*, vol. 49, no. 6, pp. 2146–2162, 2017.

[40] "Package pROC." [Online]. Available: https://cran.r-project.org/web/packages/pROC/pROC.pdf. Accessed on: Dec. 12, 2018.

[41] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2001, vol. 1, pp. I-511–I-518.

[42] T. Karantinos *et al.*, "Increased intra-subject reaction time variability in the volitional control of movement in schizophrenia," *Psychiatry Res.*, vol. 215, no. 1, pp. 26–32, 2014.

[43] P. Knox, F. D. A. Wolohan, and M. S. Helmy, "Express saccades in distinct populations: East, west, and in-between," *Exp. Brain Res.*, vol. 235, no. 12, pp. 3733–3742, 2017.

[44] W. J. van der Linden, "A lognormal model for response times on test items," *J. Educ. Behav. Statist.*, vol. 31, no. 2, pp. 181–204, 2006.

[45] J. P. Weir, "Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM," *J. Strength Conditioning Res.*, vol. 19, no. 1, pp. 231–240, 2005.

[46] "psych: Procedures for psychological, psychometric, and personality research." [Online]. Available: https://cran.r-project.org/web/packages/psych/index.html. Accessed on: Apr. 1, 2019.

[47] P. E. Shrout and J. L Fleiss, "Intraclass correlations: Uses in assessing rater reliability," *Psychol. Bull.*, vol. 86, no. 2, pp. 420–428, 1997.

[48] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropractic Med.*, vol. 15, no. 2, pp. 155–163, 2016.

[49] D. Cicchetti, "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology," *Psychol. Assessment*, vol. 6, no. 4, pp. 284–290, 1994.

[50] P. Roy-Byrne, A. Radant, D. Wingerson, and D. S. Cowley, "Human oculomotor function: Reliability and diurnal variation," *Biol. Psychiatry*, vol. 38, no. 2, pp. 92–97, 1995.

[51] T. Blekher *et al.*, "Test-retest reliability of saccadic measures in subjects at risk for Huntington disease," *Investigative Ophthalmol. Vis. Sci.*, vol. 50, no. 12, pp. 5707–5711, 2009.

[52] "The best video length for different videos on YouTube." [Online]. Available: https://www.minimatters.com/youtube-best-video-length. Accessed on: Jun. 1, 2019.

[53] K. A. Josephs *et al.*, "Rapidly progressive neurodegenerative dementias," *Arch. Neurol.*, vol. 66, no. 2, pp. 201–207, 2009.

[54] H. W. Heuer *et al.*, "Antisaccade task reflects cortical involvement in mild cognitive impairment," *Neurology*, vol. 81, no. 14, pp. 1235–1243, 2013.

[55] S. Rivaud, R. M. Müri, B. Gaymard, A. I. Vermersch, and C. Pierrot-Deseilligny, "Eye movement disorders after frontal eye field lesions in humans," *Exp. Brain Res.*, vol. 102, no. 1, pp. 110–120, 1994.

[56] C. Pierrot-Deseilligny, R. M. Müri, C. J. Ploner, B. Gaymard, S. Demeret, and S. Rivaud-Pechoux, "Decisional role of the dorsolateral prefrontal cortex in ocular motor behaviour," *Brain*, vol. 126, no. 6, pp. 1460–1473, 2003.

[57] C. Theleritis, I. Evdokimidis, and N. Smyrnis, "Variability in the decision process leading to saccades: A specific marker for schizophrenia?" *Psychophysiology*, vol. 51, no. 4, pp. 327–336, 2014.

[58] R. Ulrich and J. Miller, "Information processing models generating lognormally distributed reaction times," *J. Math. Psychol.*, vol. 37, no. 4, pp. 513–525, 1993.

[59] M. Proudfoot *et al.*, "Eye tracking in amyotrophic lateral sclerosis: A longitudinal study of saccadic and cognitive tasks," *Amyotrophic Lateral Sclerosis Frontotemporal Degeneration*, vol. 17, no. 1/2, pp. 101–111, 2015.