Saccade Latency Determination Using Video Recordings from Consumer-Grade Devices

by

Gladynel Saavedra-Peña

Submitted to
the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

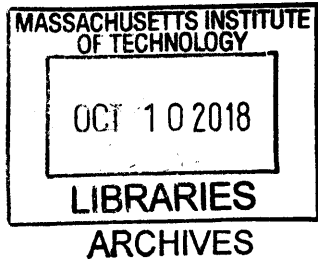MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2018

Author . . . . **Signature redacted** . . . .
Department of Electrical Engineering and Computer Science

Signature redacted August 31, 2018

Certified by . . . .
Thomas Heldt
Associate Professor of Electrical and Biomedical Engineering

Signature redacted Thesis Supervisor

Certified by . . . .
Vivienne Sze
Associate Professor of Electrical Engineering and Computer Science

Signature redacted Thesis Supervisor

Accepted by . . . .
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Saccade Latency Determination Using Video Recordings from Consumer-Grade Devices

by

Gladynel Saavedra-Peña

Abstract

Accurate quantification of neurodegenerative disease progression and neurocognitive decline is an ongoing challenge that muddles efforts to understand and treat these conditions. The golden standard to track neurodegenerative disease progression is a neurocognitive assessment administered through a series of forms and questionnaires, both of which suffer from high retest variability. Thus, there is a need for a validated biomarker that can accurately track neurocognitive decline or disease progression unobtrusively. Clinical studies have shown that saccade latency – a measure of eye movement reaction time – can be significantly different between healthy subjects and neurodegenerative disease patients. Hence, we propose and test a system that enables measurement and tracking of saccade latency outside of the clinical environment, without relying on supplemental or special-purpose illumination. iTracker-face, a variant of a deep-learning approach that estimates gaze on mobile devices, was tailored to measure saccade latency on video sequences obtained with an iPhone 6. Additionally, our model-based approach for saccade latency estimation provides a means to automatically detect eye position traces that should not be included in further analysis (AUC = 0.92). We recorded over 20,000 saccade latencies in 30 healthy subjects and observed that there is a substantial amount of intra- and inter- subject variability in the mean saccade latency (μ) and associated standard deviation (σ). Among the subjects with the most saccade latency measurements, μ ranged from 128 to 173 ms, while σ varied from 24 to 53 ms. Our results suggest that accurate and robust saccade latency determination is feasible using consumer-grade cameras and might therefore enable unobtrusive tracking of neurodegenerative disease progression.

Thesis Supervisor: Thomas Heldt
Title: Associate Professor of Electrical and Biomedical Engineering

Thesis Supervisor: Vivienne Sze
Title: Associate Professor of Electrical Engineering and Computer Science

Acknowledgments

This thesis is a result of the guidance and support of many wonderful individuals. Firstly, I would like to thank my thesis advisors, Thomas Heldt and Vivienne Sze, who needless to say were crucial to the success of this work. Thomas and Vivienne, thank you for your invaluable guidance these past two years and for teaching by example how to become an inquisitive, passionate, and well-rounded researcher. Your unwavering dedication and belief in this project has been a buttress in difficult times and I could not be more grateful for that. Many thanks as well to Prof. Charles Sodini, whose commitment to this project is evidenced by his constant presence in countless meetings and conversations. Your wise inputs have been key to the success of this project; thank you for always steering us in the correct direction.

Many thanks as well to the individuals and organizations that have supported this work financially: the Irwin Mark Jacobs and Joan Klein Jacobs Presidential Fellowship, the Alfred P. Sloan UCEM Scholars Award, and the 3M Non-Tenured Faculty Award.

My lab-mates in the INCCI and EEMS groups have been a source of companionship and support. Thank you for your insightful comments and for helping shape this project in positive, meaningful ways. Of my lab-mates, I am beholden to Hsin-Yu Lai. Hsin-Yu, I am incredibly blessed to have you as my research partner and friend. I look forward to learning and growing beside you for the rest of this journey and beyond.

From my MIT family, a special thanks to the ladies of the EECS graduate office. Your guidance and support have made the transition to MIT much smoother. I am deeply indebted to Leslie Kolodziejski, who always find time in her busy schedule to listen and care. To the mentors and friends from my UCEM family: thank you for being there and understanding what it is like for us at MIT.

To my teachers and mentors through elementary school up to college: I could not have done this without your dedication and support. Todd Coleman, Kedar Shah, Eduardo Juan, and Baldomero Llorens have believed in my abilities even when I doubted them most. Thank you for encouraging me to do, be, and believe my best.

To my beautiful, strong, and inspiring group of friends: Valeria, Keishly, Erika, Adriana, Nancy, Wilmarie, and Gabriela. Thank your listening and cheering me on when I need it the most. You all inspire me with your strength, courage, and hope. I miss having all of you within arm's reach, but am happy we find ways to keep each other in our lives. *¡Las adoro!*

To the most important people in my life, my family, who do not tire of loving. *Mami,* thank you for your constant encouragement and for leaving everything behind when I needed you most. Your pure, selfless love and tireless character make me aspire to be more like you. *Te amo incansablemente y gracias por enseñarme con tu ejemplo a ser una mejor mujer.* *Papi,* I can never repay you the time you have sacrificed to ensure we live a better life. Thank you for listening enthusiastically, giving your best, and loving me. Rocío, you are my better half. Throughout the entirety of my life you have taught me courage, resilience, and determination. I love you with every ounce in my heart and am thankful for every day I get to have you in my life. Millito, *eres la luz y alegría en mis días.* Thank you for always making me a part of your life and never drifting away. I love you deeply and miss you everyday. Ata, *no tengo palabras para agradecerte todo lo que haces por mí y por nuestra familia. Me has dado el mejor ejemplo de amor y lucha; me has acercado más a Dios. Gracias por nunca desistir y por darnos lo mejor de tu vida para nuestro bienestar. Te amo y te amaré siempre.*

Lastly, my most genuine and heartfelt gratitude to my partner in life, Rafael. You have been by my side since day one in this graduate school journey and have supported me in innumerable ways. You have patiently listened to my rants, encouraged my ideas, and always have words of support. Thank you for being my love, my best friend, and for filling my days with happiness. My life is better since you are in it. Many thanks as well to your wonderful family, who have welcomed me into their lives lovingly. *Te amo con todas mis fuerzas.*

A todos los boricuas en la diáspora: esto también es para ustedes.

Contents

Chapter 1

Introduction

Quantitative and accurate tracking of neurocognitive decline remains an ongoing challenge. The clinical gold standard for assessing cognitive function relies on an ensemble of neurocognitive and neuropsychological tests, that take a significant amount of time to administer, require a trained provider, and suffer from high retest variability [21]. Previous studies have shown that the Mini-Mental State Examination (MMSE), a universally used cognitive screening test, lacks the sensitivity to accurately discriminate between healthy subjects and patients with early dementia or Parkinson's Disease (PD) [19, 28]. Furthermore, repeat assessments using the gold standard are qualitative, sparse, and suffer from inter-observer variability. Thus, there is no objective metric or biomarker for accurate tracking of neurocognitive decline or neurodegenerative disease progression. This technology gap is particularly limiting in Alzheimer's Disease (AD), in which expensive and invasive neuroimaging studies are performed to assess a patient's response to candidate treatments.

One way to address the lack of an objective, quantitative, and accurate metric to track neurocognitive decline is by monitoring the changes in a set of digital biomarkers that correlate with disease progression [21]. Digital biomarkers are features of physiological variables measured through portable platforms, such as laptops and smart phones. Contrary to current diagnostic methods, data from consumer devices can be collected unobtrusively and repeatedly. This allows for data averaging, reducing random variability in observations and enabling longitudinal feature tracking for mon-

itoring of disease progression or quantification of response to treatment. One example of a prominent digital biomarker is saccade latency – the time delay between the appearance of a visual stimulus and the eye movement towards the stimulus. Previous studies have shown significant differences in saccade latency between normal subjects and patients with AD [54], as well as other neurodegenerative diseases [5, 47].

Typically, saccade latency is recorded under controlled conditions using high-cost, special purpose capital equipment, such as the Fourward Technologies Dual Purkinje eye-tracker [10, 23, 25] or the Applied Science Laboratories eye-tracker [68]. Recording set-ups rely on infrared (IR) lighting, one or more high-speed cameras (1-2 kHz), and a chin or forehead rest, to perform gaze estimation and subsequently measure saccade latency. These requirements make it difficult to measure saccade latency outside of dedicated environments (e.g. clinical), more so to acquire repeat observations for longitudinal tracking. At the same time, there has been a significant effort to enable gaze estimation on mobile devices for multimedia purposes within other research domains, such as computer vision and human-computer interfaces [22, 29, 32, 62]. Accurate eye-tracking on mobile devices is challenging given the absence of an IR light source and high-quality video cameras, adverse lighting conditions, limited computational resources, and an unstable tracking device. While these algorithms are optimized to maximize gaze estimation accuracy, this metric does not translate into accuracy of saccade onset detection. Accurate and precise saccade onset detection is crucial for calculating saccade latency, as saccade latency is the time difference between the presentation of a visual stimulus and saccade onset.

1.1 Summary of contributions

The overarching goal of this work is to develop and evaluate a system that enables measurement and tracking of saccade latency outside of the clinical environment. A significant step toward this goal is to exchange the recording device from a clinical-grade camera to a consumer-grade camera. In our video-recording set-up, subjects were asked to move their gaze toward a visual stimulus while they were recorded

with the rear-facing camera of an iPhone 6 in slow-motion mode. After acquiring the video-recordings, an eye-tracking algorithm was employed to extract an eye-position estimate from each video frame. The iTracker algorithm [32], a state-of-the-art convolutional neural network that estimates a user's gaze on a mobile platform's screen, was considered in this work. A modified version of iTracker that only processes facial information (iTracker-face) was more robust under challenging environmental conditions. After an eye-position estimate was obtained, a hyperbolic tangent model was fit unto the horizontal eye movement trace to determine the onset of the eye movement. This model-based technique differs from the traditional numerical differentiation approach [27], which amplifies high-frequency noise in the eye-position data. Once the onset of the eye movement was estimated, saccade latency was calculated by measuring the time delay between the presentation of the visual stimulus and the eye-movement onset.

In total, we recorded over 20,000 saccades in 30 healthy subjects with our measurement system. With this amount of data, we observed that the intra-subject variability in saccade latencies can be quite substantial across healthy subjects. Each individual has a saccade latency distribution with distinctive shape and parameters. This result suggests that the common approach of analyzing saccade latency in case-control studies by pooling data from different subjects may obscure important intra-subject variation that might be important to detect. Moreover, we observed that the saccade latency distribution of the majority of subjects is well fit by a log-normal distribution. Among the healthy subjects with the most saccade latency measurements ($785 < n < 2,940$), seven out of ten distributions conformed to a log-normal distribution ($p < 0.05$). These same ten subjects were recorded on ten separate occasions; the behavior of their saccade latency sample mean and standard deviation was analyzed throughout these recording sessions. It was observed that although the variation in the sample mean and standard deviation is significant, the log-normal fitting parameters seem to be stationary across time for the majority of subjects.

These initial contributions are a substantial effort towards recording saccade latency from a broad population, paving the way for understanding the behavior of this

11

measurement in the general public. However, a deeper understanding of saccade latency is necessary to put into perspective the saccade latency changes seen in patients with neurocognitive decline.

1.2 Thesis structure

In Chapter 2, an overview of eye movements, relevant neurophysiology, and saccade latency behavior is provided. Chapter 3 describes current and state-of-the-art eye-tracking systems in multiple environments, drawing comparisons between clinical and multimedia environments. Subsequently, the current data collection procedure and a description of the eye-tracking algorithms for measuring saccade latency are discussed in Chapter 4. Next, existing models for saccade waveforms as well as our own model are described in Chapter 5. Chapter 6 explores in detail the properties of the saccade latency distributions across healthy subjects. A discussion of the results presented in Chapter 6, a conclusion of the thesis, and future work are presented in Chapter 7.

Chapter 2

Background

This chapter lays the groundwork for understanding the following chapters. Each section in the chapter covers a variety of topics ranging from a discussion on functional classes of eye movements to an overview of saccade latency behavior in normal, healthy subjects and patients with neurodegenerative diseases.

2.1 Eye movements

Eye movements are a rich source of information for clinicians and scientists, serving as a tool for investigating brain function and understanding brain and motor diseases. Most of the appeal of studying eye movements stems from the fact that they are readily accessible for observation, measurement, and examination. Furthermore, abnormalities in eye movement patterns can mostly be attributed to a specific pathophysiology or anatomical localization [36]. Hence, observed abnormalities can be linked to specific pathophysiologies.

Eye movements are broadly categorized into two main types: (1) eye movements that stabilize gaze and (2) eye movements that shift gaze. Both types of eye movements share the same purpose, which is to steadily place an image of an object on the region of the retina where visual acuity is highest, the fovea [36]. Eye movements can be further divided into seven main classes: (1) vestibular, (2) optokinetic, (3) visual fixation, (4) smooth pursuit, (5) nystagmus quick phases, (6) vergence, and (7) sac-

cades [36]. Each class has a main function in achieving the primary goal of positioning an image on the center of the retina, i.e. fovea. Vestibular eye movements steadily hold an object's image on the retina during brief head movement while optokinetic eye movements do the same but during sustained head movement. Visual fixation holds the image of a static object on the fovea and smooth pursuit movements hold the image of a small moving object. Nystagmus quick phases reset the eyes during a prolonged rotation and redirect the gaze toward the new visual scene, while vergence moves the eyes in opposing directions to hold the visual image on each eye's fovea. Lastly, saccades shift the line of sight to bring the image of the object of interest onto the center of the retina. For each of the seven functional classes of eye movements, there is an anatomical circuit that controls its proper execution. Here, we will only review visual fixation and saccades.

Eye movement measurement units

Briefly, we discuss the unit for measuring eye movements. Eye movements are often measured in visual degrees or degrees, which are a function of the viewing distance d and the units of displacement z spanned by the visual angle:

$$\tan(\theta) = \frac{z}{d} \tag{1}$$

Fig. 2-1 illustrates this relationship, where a displacement of the line of sight from point a to point b at a distance d from the laptop screen is equivalent to θ degrees. The displacement on the screen z and the distance d should be in the same measurement units (e.g., mm, cm).

Visual Fixation

As previously mentioned, visual fixation is the act of holding an object's image steady upon the fovea. Physiologically, this means that the extraocular muscles work jointly to prevent the elastic and viscous forces of the eye from swaying the eyeball to a different orientation [36]. Although the term implies otherwise, the eyes are in constant
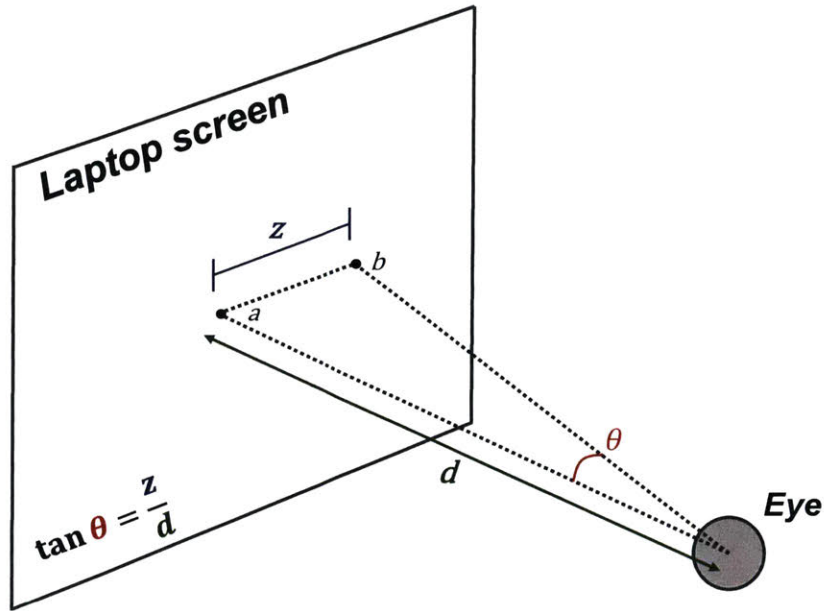
14

Figure 2-1: Visual degrees are a function of the viewing distance and the displacement spanned by the visual angle. When the line of sight is displaced from point a to point b at a distance d from the screen, the eye has moved θ degrees. Figure modified from [27].

movement during visual fixation. These eye movements are very small in amplitude and mostly function to prevent the visual response to a static image from fading away. This fading occurs because the visual system's response desensitizes to unchanging visual stimuli.

Saccades

Saccades are quick eye movements that shift the line of sight between two fixed visual points with the goal of placing the image of interest onto the fovea. Saccadic eye movements can be hierarchically categorized from the most primal to the most complex [35]. With this hierarchy in mind, the classification of saccadic eye movements is as follows [35]:

- **Quick phases:** Generated during vestibular or optokinetic stimulation to hold an image on the retina during head motion.

- **Scanning saccades:** Performed consecutively with the purpose of extracting

new information from a visual scene.

- **Spontaneous saccades:** Performed at random without any type of stimuli.

- **Express saccades:** Typically occur when a subject performs a visual task called the gap-stimulus task; are of very short latency.

- **Reflexive saccades:** Generated toward novel, unexpected stimuli. The stimuli does not need to be visual, but can be auditory or tactile.

- **Volitional saccades:.** Performed by choice as part of purposeful behavior. These saccades can be further classified into *predictive*, *memory-guided*, *anti-saccades*, and *to command*.
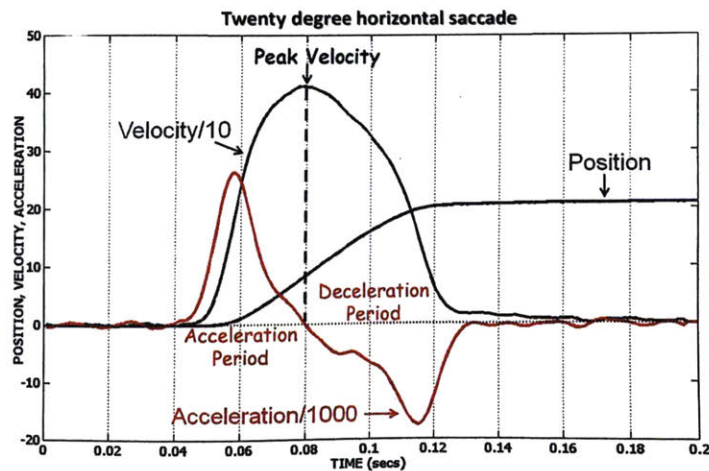

Figure 2-2: A representative record of saccade position, velocity, and acceleration. The x-axis is time in seconds and the y-axis is the position, velocity, and acceleration of the eye movement in degrees, degrees per second, and degrees per second squared. The velocity and acceleration of the saccade are scaled so they can be superimposed. Figure reproduced from [35]

A schematic of a saccade is presented in Fig. 2-2. The saccade itself is illustrated by the curve labeled 'position', while the sacadde's computed velocity and acceleration curves are labeled correspondingly. Notice that there is an acceleration period at the beginning of a saccade that extends until peak velocity is reached. This is immediately followed by a deceleration period that lasts until the saccade ends. The beginning and
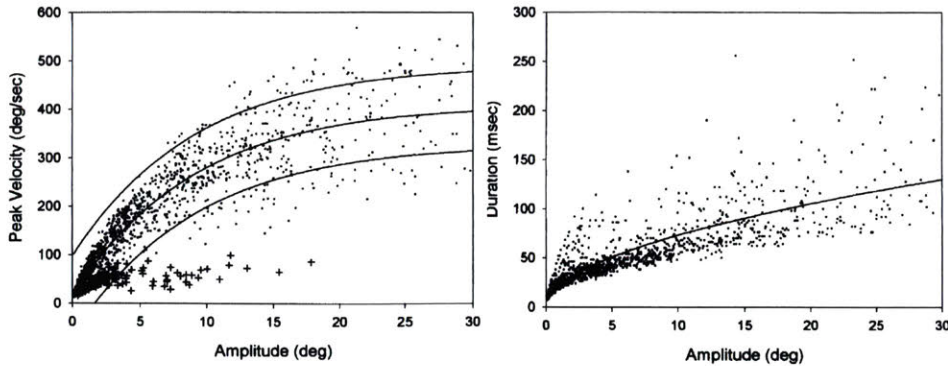
Figure 2-3: Main sequence relationship plots depicting saccade peak velocity and duration as a function of saccade amplitude. The dots in both plots are saccades from 10 normal subjects, while the crosses *(left)* are saccades from a patient with a disease. The data are fit with exponential and power equations, where the 5% and 95% prediction intervals are shown for the plot on the left. Figure reproduced from [34].

end of a saccade are commonly determined by fixed thresholds, such as 30 degrees per second (deg/s), on the saccade velocity profile.

The relationship between the size of a saccade, its speed, and its duration is remarkably consistent [35], as illustrated by Fig. 2-3. The left plot shows that the relationship between peak velocity and amplitude is fairly linear for saccades that are smaller than about 20 degrees. In other words, the larger the saccade amplitude the higher its peak velocity. Above an amplitude of 20 degrees, the peak velocities saturate progressively until they reach a value of about 500 deg/s. One equation that is commonly used to describe this relationship is:

$$Peak\ velocity = V_{max}(1 - e^{-Amplitude/C}) \tag{2}$$

where V_{max} is the asymptotic peak velocity and C is a constant [35].

The right plot in Fig. 2-3 shows the relationship between saccade duration and amplitude. This relationship is somewhat linear for saccade amplitudes ranging from 1 to 10 degrees, but there is substantial variability in the duration values. However, in general, the larger the saccade the longer its duration.

It is important to note that neither saccade velocity nor duration can be purpose-

17

fully controlled. In fact, these variables have a substantial amount of variability, even within individuals, on a day to day basis. However, the main sequence relationships are consistent enough that they allow for the definition of a range of velocity and duration values for normal saccades. Any substantial deviation from this range indicates abnormal saccades, which in turn might indicate abnormal neurophysiology (Fig. 2-3).

2.2   Neurophysiology of saccades
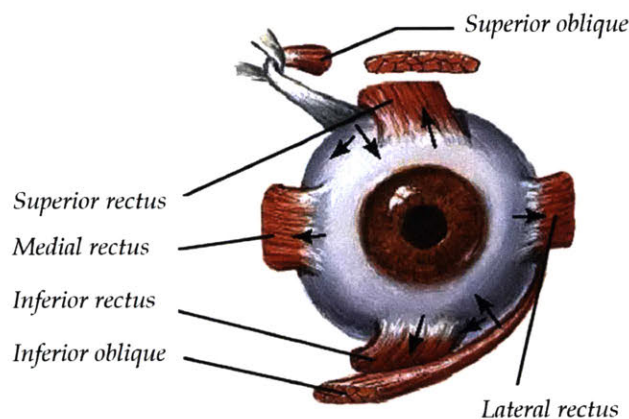
Figure 2-4: The six extraocular muscles of the left eye. The superior rectus, medial rectus, inferior rectus, and inferior oblique are innervated by Cranial Nerve III. The superior oblique is innervated by Cranial Nerve IV, while the lateral rectus is innervated by Cranial Nerve VI. Image modified from [2].

Generating an eye movement entails overcoming the viscous and elastic forces placed upon the eye by the surrounding tissue [36]. A rapid contraction of the extraocular muscles drives eye movement by overcoming these forces. Specifically, eye movements are controlled by six extraocular muscles: the lateral rectus, medial rectus, inferior rectus, superior oblique, superior rectus, and inferior oblique (Fig. 2-4). Neural recordings have shown that the motoneurons that innervate these extraocular muscles (Cranial Nerves III, IV, and VI) have a pulse-step discharge during saccadic eye movements [36, 42].
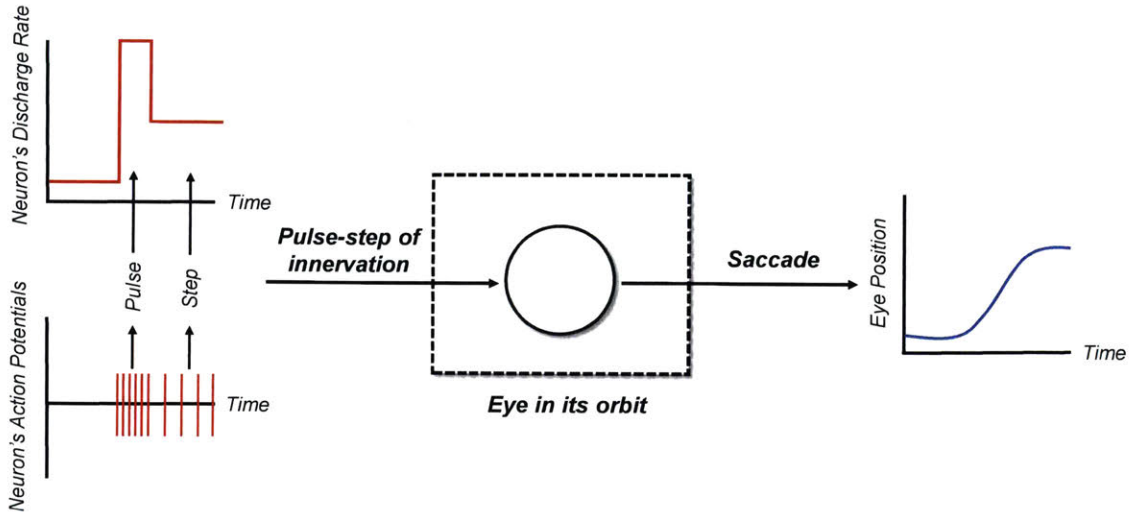
2.2.1 The pulse-step of innervation

Figure 2-5: The pulse-step of innervation. On the left is the neural input signal, the pulse-step of innervation, sent to the extraocular agonist muscles to produce a saccade. The vertical lines indicate the occurrence of an action potential in an ocular motoneuron, while the graph above them is the neuron's discharge rate. To the right is a plot of the resulting eye position across time. Image modified from [36].

During a saccade, the motoneurons that innervate the extraocular agonist muscles discharge a pulse-step signal that drives the movement of the eye toward the target (Fig. 2-5). As the name implies, the signal has two components: the pulse of innervation and the step of innervation. The pulse of innervation is a burst of neural activity that encodes the velocity and amplitude of the saccade [42]. Specifically, the frequency of the pulse component is correlated with the eye velocity, while the total number of spikes is related to the amplitude. The step of innervation is a tonic discharge that correlates with eye position and keeps the eye fixed at the end of a saccade. Chiefly, the step of innervation maintains the steady contraction of the extraocular agonist muscles and prevents the eye from rolling back to its central position. When the pulse-step signal is working properly, the eye moves quickly to its new position and stays there [36]. On the contrary, if one of the two components is faulty, saccadic performance is inappropriate. Without the pulse of innervation the eye movement would be too slow, while the absence of the step of innervation would result in an inability to keep the eye steady at its new position, causing the eye to

drift back to center.

It is important to note that both the pulse of innervation (velocity command) and the step of innervation (position command) are velocity-encoded, i.e. the neural traffic is proportional to velocity [36]. This might seem counter intuitive considering the pulse-step signal encodes eye position, but neurophysiological evidence suggests that the step of innervation is generated by a neural network within the brainstem and cerebellum that mathematically integrates the velocity command with respect to time. This neural network is referred to as the *neural integrator for eye movements* and when faulty, it fails to produce the step of innervation that holds the eye steady in its new position. A diagram summarizing the neural integration process is shown in Fig. 2-6.
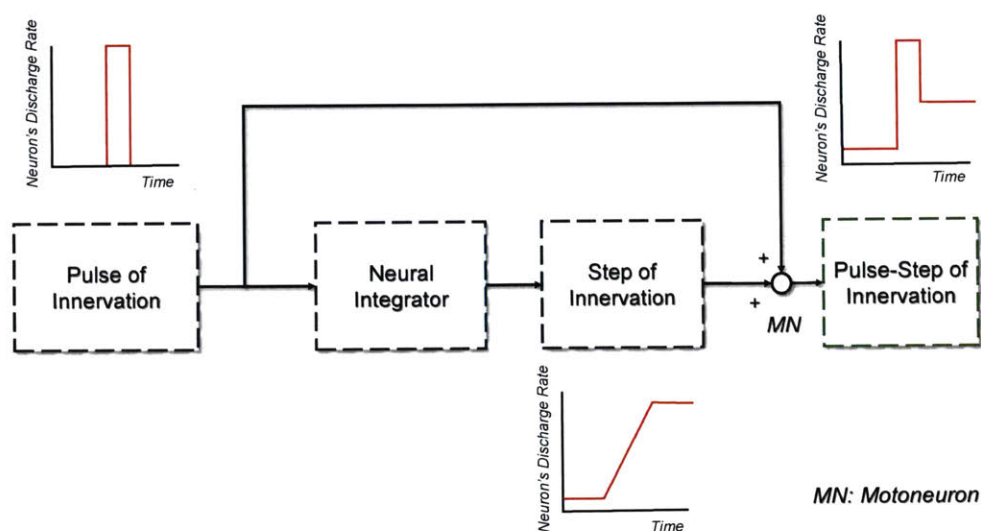
Figure 2-6: A simplified diagram depicting the neural integration process. The pulse of innervation (velocity command) is integrated by the neural integrator to produce the step of innervation (position command). The pulse and step are combined at the motoneuron to produce the pulse-step of innervation. Image modified from [36].

2.2.2 An overview of relevant brain areas

Experiments in human and animal subjects have aided researchers in identifying the brain areas involved in the generation of saccades. The major structures participating

Figure 2-7: Midsagittal section of the brain. The lobes of the cerebral cortex, the brainstem, the cerebellum, and the superior colliculus are labeled. The basal ganglia are not visible from this view. Image adapted from [15].

in this process are the cerebral cortex, the brainstem, the cerebellum, the superior colliculus and the basal ganglia [42]. The cerebral cortex is the convoluted surface of the cerebral hemispheres and it is divided into four lobes: (1) the frontal lobe, (2) the parietal lobe, (3) the temporal lobe, and (4) the occipital lob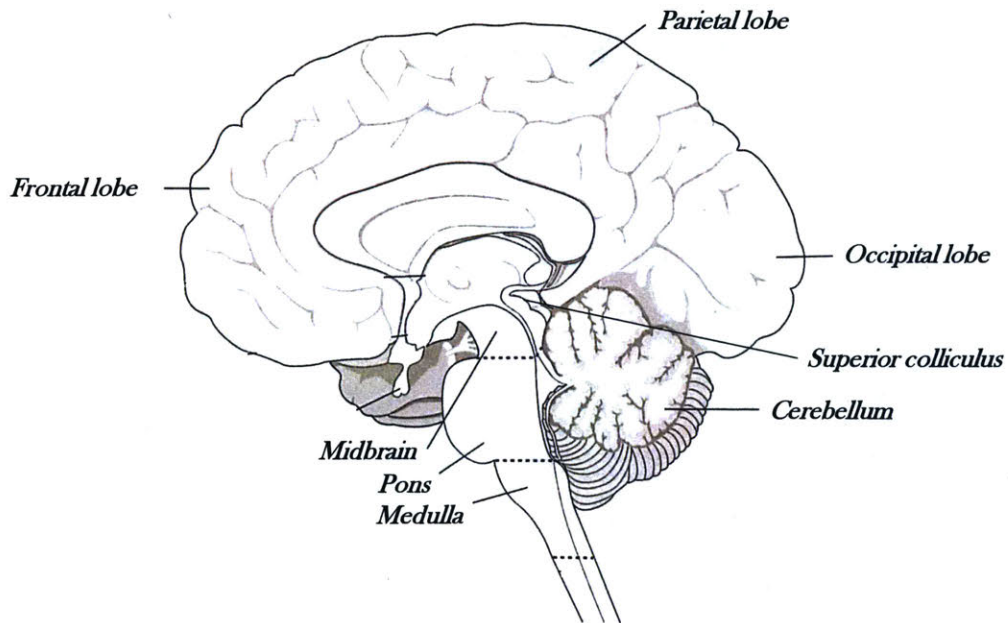e [15]. The main function of the cerebral cortex is to receive and process sensory information, as well as to integrate motor functions. The brainstem consists of three distinct structures: the medulla, the pons, and the midbrain. The medulla regulates breathing and blood pressure, and with the pons maintains balance and posture. The midbrain participates in the control of eye movements. The cerebellum's main functions are coordinating, planning, and executing movement, especially head and eye movements. Damage to the cerebellum results in a lack of coordination, such as delayed onset of movement or poor execution of a sequence of movements [15]. The superior colliculus is a specialized, laminated structure with an important role in the interplay between visual fixation and saccades [44]. Lastly, the basal ganglia are a set of nuclei deep

within the brain whose role is to aid in the planning and execution of smooth movements [15]. Parkinson's Disease and Huntington's Disease are known, for example, to specifically target the basal ganglia [15]. An illustration of the brain with some of these structures is seen on Fig. 2-7.

2.2.3 Saccade neural circuitry

The areas of the brain involved in the generation of saccades and their interconnections are shown in Fig. 2-8. Although there are multiple areas within the major structures of the brain that are involved in this process, the following discussion will focus on the main areas i.e., the brainstem, cerebellum, superior colliculus, and cerebral cortex.
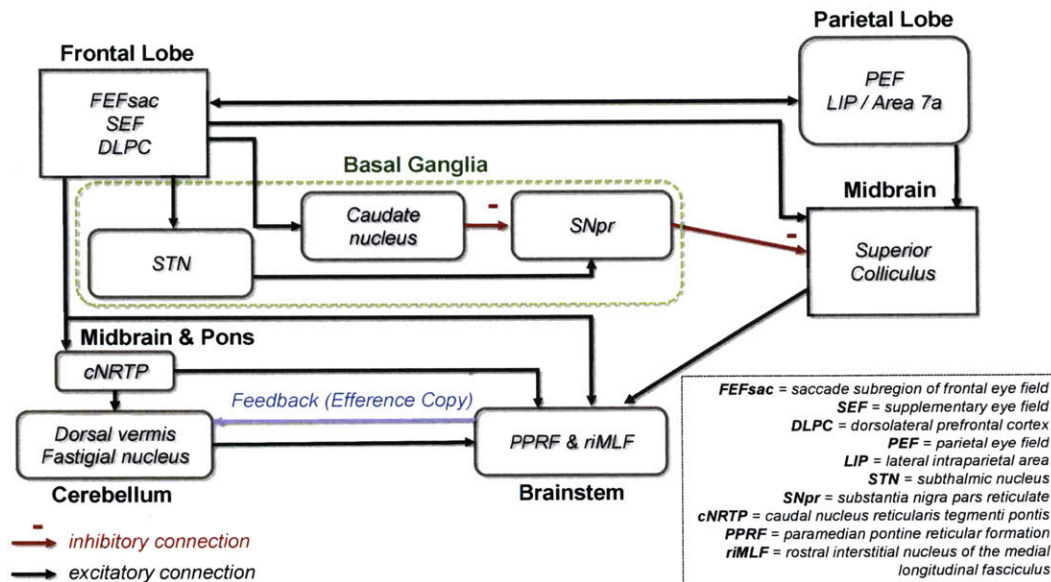
Figure 2-8: The areas of the brain involved in generating a saccade and the connections among them. Major brain structures in this diagram are highlighted in bold (e.g., the frontal and parietal lobes, the basal ganglia, etc.). Image modified from [61].

2.2.4 The role of the brainstem

The motoneurons that innervate the extraocular muscles and discharge the pulse-step of innervation receive their signals from premotor burst neurons located in the

brainstem, specifically the mesencephalic, pontine, and medullary regions [34, 42]. The pulse is generated by premotor burst neurons within the saccadic burst generator circuit in the brainstem reticular formation [42], while the step is the result of integrating the pulse within the brainstem and cerebellum [34].

There are two types of premotor burst neurons: excitatory burst neurons (EBN) and inhibitory burst neurons (IBN) [34]. The EBN and IBN for horizontal saccades are located in the paramedian pontine reticular formation (PPRF), while for vertical saccades they are located in the rostral interstitial nucleus of the medial longitudinal fasciculus (riMLF) [42]. While the anatomical locations of the burst neurons for horizontal and vertical saccades differ, their function is the same. To initiate and sustain a saccade, the EBN excite the motoneurons that drive saccades in the desired direction. Concurrently, the EBN excite the IBN, inhibiting the motoneurons that innervate the antagonizing muscle.

The neural activity of the EBN is gated by long-lead burst neurons (LLBN) and omnipause neurons (OPN) located in the brainstem. The LLBN project to the EBN and IBN and show a buildup of activity before discharging a high frequency burst to the EBN (Fig. 2-9) [42]. Both the EBN and IBN are inhibited by OPN, which are active during visual fixations and silent during saccades. The potent inhibition of the OPN is demonstrated by the fact that electrical stimulation of the OPN will stop a saccade during execution [34]. Hence before generating a saccade, OPN are silenced, followed by activation of the corresponding EBN and IBN by the LLBN [42]. After the saccade has finalized, the OPN resume their discharge and inhibit the EBN and IBN.

There are several inputs to the brainstem premotor circuitry coming from the frontal cortex, the superior colliculus, and the cerebellum. The sections that follow discuss these inputs in detail.

Lesions to the brainstem

Diseases that affect the brainstem often cause slow or absent saccades because premotor burst neurons play a major role in determining the speed and duration of
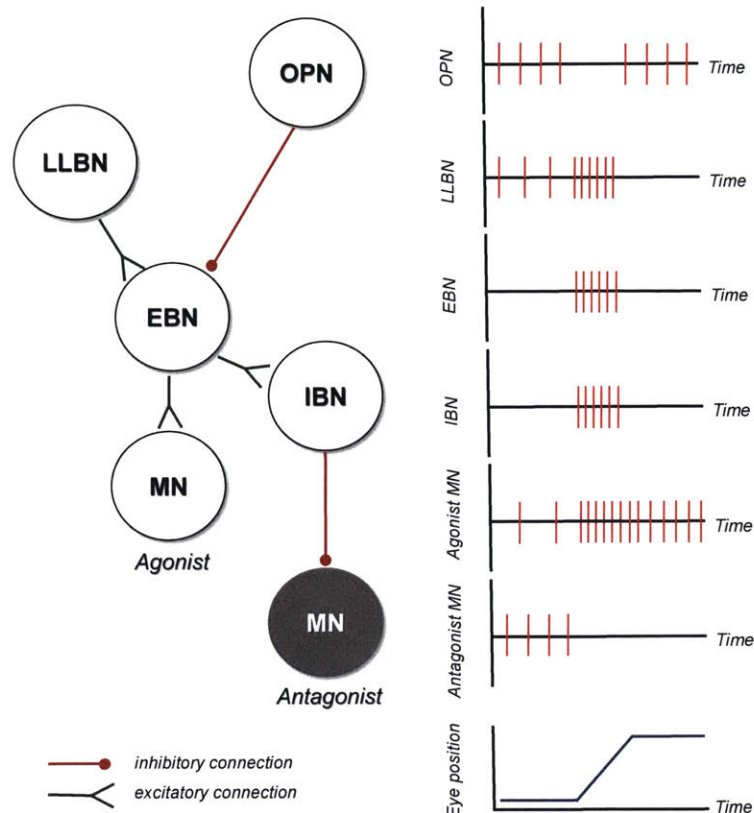
Figure 2-9: The neurons in the brainstem premotor circuit. *(left)* The excitatory and inhibitory connections between neurons. The antagonist motoneuron is colored gray. *(right)* The vertical lines on the time axis indicate the occurrence of an action potential in the specified neuron. The bottom plot is the resulting saccade. EBN, excitatory burst neuron; IBN, inhibitory burst neuron; LLBN, long lead burst neuron; MN, motoneuron; OPN, omnipause neuron. Image modified from [42].

saccades [34]. Furthermore, diseases that affect the midbrain, such as progressive supra-nuclear palsy (PSP) or Niemann-Pick type C disease, cause slowing of vertical saccades. On the other hand, OPN dysfunction causes excessive saccadic oscillations because of an inability to effectively inhibit EBN and IBN.

2.2.5 The role of the cerebellum

The dorsal vermis and the fastigial nucleus within the cerebellum project to both excitatory and inhibitory burst neurons in the brainstem, playing a key role in main-

taining the accuracy of saccades [34]. The dorsal vermis receives input from the tegmental pontine reticular nucleus within the pons and the midbrain (Fig. 2-8). The cerebellum encodes the size of saccades and repairs saccadic inaccuracies [61]. A common belief is that the cerebellum monitors a saccadic command (also called an *efference copy*) from the brainstem to determine when the eye is placed on the desired target. Once it is, the cerebellum terminates the eye movement [34].

Lesions to the cerebellum

Selective inactivation of one fastigial nucleus causes ipsilateral (on the same side of the body) saccades to overshoot their targets, while contralateral saccades are undershot [34]. However in clinical environments, lesions to the fastigial nucleus are bilateral, resulting in saccades that overshoot targets on both sides. Similarly, when the efference copy process is compromised, patients are prone to overshoot saccades. An interesting observation is that inactivation of the dorsal vermis and the fastigial nucleus results in an inability to adapt saccades to new visual demands [34]. When this pathway is damaged by disease, patients show a poor performance in tasks that require generating a sequence of saccades.

2.2.6 The role of the superior colliculus

The superior colliculus (SC) plays an important role in determining saccadic initiation, direction, and size by providing signals to the brainstem that specify where and when a saccade will occur [34, 42]. The SC is a laminated structure within the midbrain with multiple layers; the superficial layers receive direct input from the retina and other visual areas, while the intermediate layers serve as a motor map [34, 42]. The neurons in the superficial layers are organized into a topographic map of the contralateral visual hemifield, while the intermediate layers contain neurons whose discharges are correlated with saccades and visual fixations [44]. The neurons that discharge before and during saccades are the sacccade-related neurons, while those discharging during visual fixation (and pausing during saccades) are the fixation-

related neurons. The fixation-related neurons are localized in the rostro-lateral end of the SC and send excitatory connections to the OPN in the brainstem premotor circuit [34, 44]. The saccade-related neurons have been shown to project to the EBN in the brainstem premotor ciruit and are distributed throughout the intermediate layers of the SC [42, 44].

The motor map in the intermediate layers of the SC is in polar coordinates and each point in this map encodes a unique direction and amplitude [34, 44]. The smallest saccades are elicited near the rostral end of the SC and the largest near the caudal end [44]. Upward saccades are represented toward the middle (medial) and downward saccades near the side (lateral). The fixation-related neurons are localized in the extreme rostral pole of the SC; electrical stimulation of this site suppresses saccades [34]. Fig. 2-10 illustrates this motor map and the location of each stimulation site.

Higher brain centers, such as the parietal eye field (PEF) and the frontal eye field (FEF), have direct projections to the SC (Fig. 2-8). The FEF projects to the SC through the basal ganglia via direct and indirect pathways [42]. The direct pathway (FEF → Caudate nucleus → SNpr → SC) leads to the disinhibition of the SC, enabling the generation of saccades. The indirect pathway (FEF → STN → SNpr → SC) through the basal ganglia inhibits the SC. It follows that inhibitory inputs to the SC must be suppressed before driving a saccade.

Lesions to the superior colliculus

Lesions to the SC in humans are not common, but researchers report that unilateral lesions result in an inability to initiate saccades at short latencies [34]. PSP is a rare example of a neurodegenerative disorder that affects the SC, causing an initial slowing of vertical eye movements. Still, the SC plays a critical role in the generation of saccades, demonstrated by the fact that inactivation of neurons in the SC blocks the effect of stimulating higher-order cortical areas, such as the FEF [34].
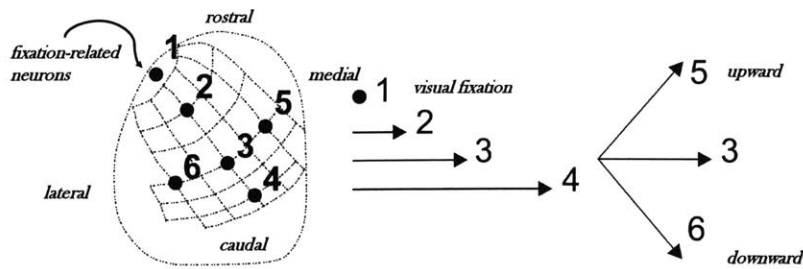
Figure 2-10: The motor map in the superior colliculus (SC). Each coordinate in the motor map corresponds to a specific saccadic direction and amplitude. The numbered vectors represent the elicited eye movements from stimulating at each site. Image modified from [44].

2.2.7 The role of cortical and subcortical areas

In humans, the cortical and subcortical areas that contribute to the generation of saccades include the FEF, the supplementary eye field (SEF), the dorsolateral prefrontal cortex (DLPC), the PEF, and the lateral intraparietal area (LIP) (Fig. 2-8). These areas project to the SC through direct and indirect pathways (through the basal ganglia) [34]. Projections from the frontal cortical areas (FEF, SEF, DLPC) also reach the brainstem premotor circuit and the cerebellum. Both cortical and subcortical areas are spatially coded, very much like the SC. For example, studies have shown that micro-stimulating distinct sites on the FEF results in a saccade with a specific direction and amplitude [13].

Lesions to cortical and subcortical areas

Different classes of saccades are driven by distinct cortical areas [42]. For example, the PEF is important in the triggering of reflexive, visually guided saccades [34]. In humans, lesions to the PEF result in an increased latency of these saccades [12, 48]. Similarly, there is abundant evidence that the FEF plays a critical role in generating voluntary, purposeful (volitional) saccades [50]. Studies have shown that lesions to the FEF cause increased latency of visually guided saccades, especially during the overlap paradigm (see Section 2.5) [12, 51].

27

An important part of successful saccadic behavior is the suppression of reflexive eye movements toward irrelevant stimuli [34]. An example of a visual task that tests this behavior is the anti-saccade paradigm, where subjects are required to suppress a saccade towards a stimulus and instead generate a voluntary saccade of equal size in the opposite direction. The FEF is bilaterally activated during pro-saccades and anti-saccades in this test paradigm [14, 18]. Results show that patients with lesions in the FEF had increased latencies on correct anti-saccades [49]. However, patients with lesions in the DLPC had a higher percentage of performance errors in this task.

2.3 Saccade latency

Saccade latency is the time delay between the presentation of a visual stimulus and the start of an eye movement (saccade) toward the stimulus (Fig. 2-11). Typically, the beginning of a saccade is identified by computing and imposing a threshold on the velocity signal, such as 30 deg/s [35]. Saccade latency has been studied extensively as it reflects the state of many core cognitive functions, such as sensory information processing, motor command programming and decision making [35].
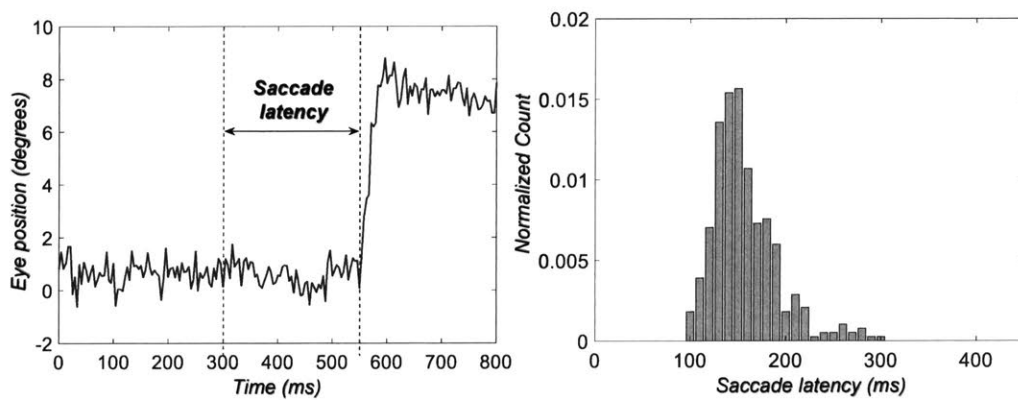

Figure 2-11: Saccade latency. *(left)* Tracing of the horizontal eye position across time. The first vertical line is the time a stimulus is presented, while the second is the moment when the eye starts to move toward the stimulus. The time difference between the two events is the saccade latency, in this case 250 ms. *(right)* Saccade latency distribution for a single subject.

Typical saccade latency values for normal, healthy subjects are around 200 ms, but the values are substantially variable within and across individuals [35, 45]. Furthermore, saccade latency variability is influenced by the properties of the stimulus, such as the modality (e.g., visual, auditory, target size) and the temporal properties of its presentation [35]. This variability is seen clearly when saccade latency values are presented as a distribution (Fig. 2-11). Independently of the modality of the stimulus, such distributions tend to be right-skewed (Fig. 2-11) [45]. Despite this significant variability and deviation from normality, the mean (μ) and standard deviation (σ) of the distribution are the only summary statistics commonly reported in clinical studies.

2.3.1 Saccade latency in the young and elderly

Aside from the properties of the stimulus and the attentive state of the subject, age plays a well documented role in the latency of saccades [35]. The consensus among studies is that saccade latency is larger in elderly subjects (> 60 years of age) compared to younger subjects (20 to 30 years of age) [9, 43, 59, 60, 64]. In other words, the older the subject, the longer it takes them to react to visual stimuli. Munoz et al. [43] measured saccade latency in 168 healthy subjects, whose ages ranged from 5 to 79 years and found that saccade latency was larger in elderly subjects (60 to 79 years of age) compared to young adults (20 to 30 years of age). Mean saccade latency values for young adults were around 220 ms, while they were around 290 ms for the elderly (Fig. 2-12). Warren and co-workers [59] found a similar trend, in that the mean saccade latency in young subjects ranged from 282 to 388 ms, and from 361 to 479 ms in older subjects (depending on the size of the saccade). The reported measure of variability was the standard error of measurement (SEM), with values ranging from 11 to 13 ms in young adults and 14 to 16 ms in older subjects. A separate study [9] with 145 healthy individuals (19 to 82 years of age) divided the subjects into smaller age groups and found that the largest increase in mean saccade latency occurs between the age brackets of 40 to 49 years and 50 to 59 years (from 177 ± 28 ms to 192 ± 35 ms for horizontal saccades to the right). Although there

is disagreement on the exact difference in mean saccade latency between age groups, the consensus is that elderly subjects have larger saccade latency values than young adults.
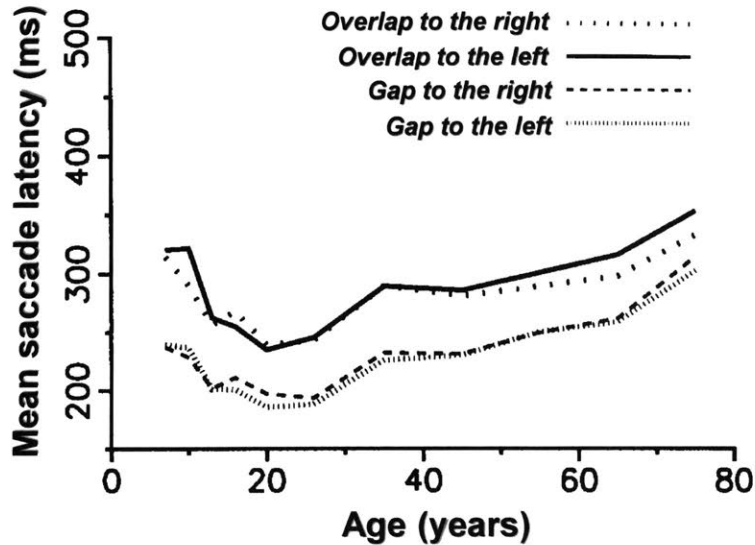
Figure 2-12: Age related changes in mean saccade latency [43]. The mean saccade latencies measured in two distinct visual tasks, the gap and overlap (see Section 2.5), are shown separately. Saccades generated to the left or right are also shown apart.

2.4 Saccade latency in neurodegenerative diseases

The nature of the eyes has made ocular biomarkers an attractive option for tracking neurodegenerative disease progression. Multiple measures of eye movement, such as velocity and accuracy, have been studied as potential indicators of neurodegenerative disease [4, 5, 54]. Among them, saccade latency is frequently examined as it reflects the state of many core cognitive functions, such as sensory information processing, motor command programming and decision making [35].

Saccade latency has been analyzed in the context of multiple neurodegenerative disorders, such as Alzheimer's Disease (AD) [10, 16, 38, 54, 65], Parkinson's Disease (PD) [40, 47], Huntington's Disease (HD) [5] and schizophrenia [31, 56]. In

the majority of these studies, the finding is that saccade latency is larger in the patient population compared to age-matched, healthy controls. One study [54] with 245 healthy volunteers and 35 AD patients found that patients had longer and more variable saccade latencies compared to normal, healthy controls in saccadic eye movements toward a target. Specifically, it reported a mean saccade latency difference of 60 ms between groups and a standard deviation of 40 ms and 90 ms respectively for healthy controls and patients. A separate study reported a smaller mean saccade latency difference of 16 ms between groups [16].

The multiple mean saccade latency differences may arise for a variety of reasons, such as the properties of the visual task performed in the study, the number of participants, and the distribution of disease severity across patients. Another distinct parameter of saccade latency is its variability, which some studies report is larger in patients with neurodegenerative disease [31, 54, 56]. One specific example are schizophrenic patients, which have been found to have a significantly larger intra-subject saccade latency variability [31, 56]. Hence, the general agreement among these studies is that saccade latency is larger and more variable in these patient populations compared to age-matched healthy controls.

Unfortunately, there are few studies that examine the longitudinal behavior of saccade latency in the context of neurodegenerative diseases and disease progression. A rare exception is [5], where saccade latency was measured in a cohort of normal, healthy subjects and HD patients over the course of three years. This study reports an annual increase in saccade latency in the patient group and no significant changes in the healthy control group, indicating the potential use of this measurement as a biomarker for tracking neurodegenerative disease progression.

A separate body of work focused on correlating saccade latency performance with brain pathology [10, 47]. One study found that all autopsy-confirmed AD patients showed a significant increase in saccade latency compared to other patient groups ante-mortem [10]. A separate study analyzed brain MRI images in eighteen PD patients and observed that the mean and variability of saccade latency was inversely correlated with grey-matter volume in distinct brain regions related to saccadic con-

trol [47]. Although there is a growing effort to understand brain degeneration and its effect on saccade latency, more work is needed to fully understand the intricacies of this relationship.

Apart from saccade latency, other saccade variables, such as the proportion of saccades initiated toward the incorrect direction, have also been found to be significantly different between healthy controls and patient groups [4, 16, 54]. Due to the inherent variability in these measurements, the performance of some individual patients places them within the normal range and vice versa. Nonetheless, some researchers [54] point to the importance of monitoring the longitudinal behavior of saccade variables, as they could deteriorate progressively within individual patients and serve as indicators of disease progression.

2.5   Visual tasks to measure saccade latency

A crucial step toward acquiring saccade latency is the design and implementation of visual tasks that allow for measuring this feature. There exist a variety of such visual tasks in the clinical literature, where the most common paradigms are the step-jump, overlap and gap tasks. All paradigms include a fixation light, usually displayed at the center of a screen, and a target, located to the side of the fixation light (Fig. 2-13). In the step-jump paradigm, the target is turned on as the fixation light is turned off (Fig. 2-14 A), while in the overlap paradigm the fixation light stays on throughout the task (Fig. 2-14 B). In the gap paradigm, the fixation light is turned off before the target is shown, producing a gap period where the screen is completely dark (Fig. 2-14 C). The gap period is usually 100 ms to 400 ms long. In all tasks, the subjects are asked to fix their gaze on the fixation light until the target appears. When it does, they should direct their gaze toward the target.

The temporal properties of the visual task have been shown to influence saccade latency [35]. Saccade latencies are shorter when the gap paradigm is used compared to saccade latencies in the overlap and step-jump paradigms [11, 20, 53]. The explanation for this phenomenons is that the gap period serves as a mechanism to release visual

attention and disinhibit the saccade-generating neural circuits, leading to a reduction in saccade latency [20]. A gap period of approximately 200 ms yielded the shortest saccade latencies [11, 20, 53]. In contrast, the step-jump and overlap paradigms engage visual attention and inhibit the saccadic neural circuit, resulting in larger latencies compared to the gap paradigm [53].
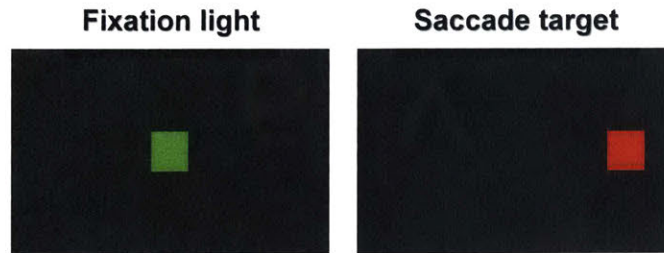
Figure 2-13: A fixation light (green square) and a saccade target (red square) against a black background. In this example, the fixation light is at the center of the screen and the saccade target is to the right of it.
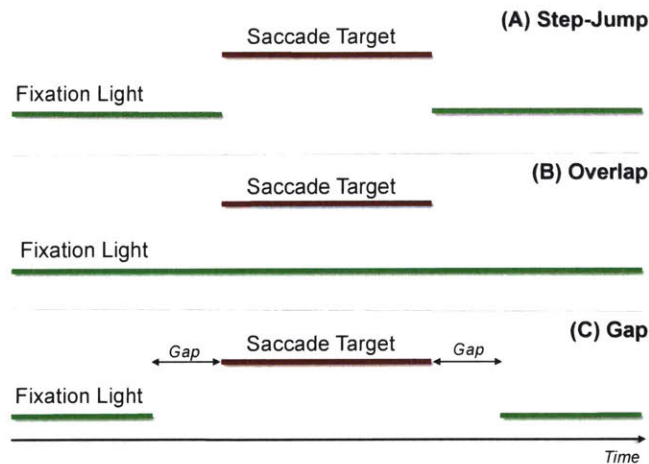
Figure 2-14: The time sequence of three paradigms used in the laboratory to measure saccade latency. The green and red lines indicate when the fixation and targets are shown on the screen. All trials start with the subject viewing a fixation light (green) at the center of the screen. Once the target (red) appears on the periphery, the subject must direct their gaze toward it. Modified from [35].

Chapter 3

Eye-tracking techniques

Accurate detection and tracking of the eyes has long been of interest for many applications, such as clinical medicine, cognitive state analysis, and human-computer interfaces [24]. To that end, a variety of eye-tracking techniques have been developed throughout scientific history, the earliest being contact lenses and electro-oculography systems [27]. However, this chapter will focus on discussing current eye-tracking technology, namely, video-based systems (video-oculography) that track eye movements in a video sequence.

The goal of a video-based eye-tracking system is to identify and localize the eyes in an image sequence and track their movements to estimate the direction of gaze. Gaze represents the focus of attention of a person and has been variably defined as a person's line of sight, a set of 3-D visual axes projecting from the eyes, or a 2-D coordinate on a computer's screen [66]. The basic components of a video-based eye-tracker are illustrated in Fig. 3-1. Facial and ocular information is captured through an image using one or multiple video-cameras. The image is analyzed by an eye-detection procedure (e.g. the Viola-Jones algorithm [58]) that outputs the location of the eye within the image (*eye location* in Fig. 3-1). Subsequently, the eye location information is used in the intended application or to initialize an eye-tracking algorithm. If the ultimate goal of this procedure is to estimate gaze, the initialized eye-tracking algorithm analyzes the subsequent images and outputs information obtained from the eye region (*eye data* in Fig. 3-1). This information could be the position

of the eye across frames or a set of extracted eye features. The eye information is combined with an estimated head pose to produce a final gaze estimate. In the end, the estimated gaze coordinates are utilized by a gaze-based application [24].
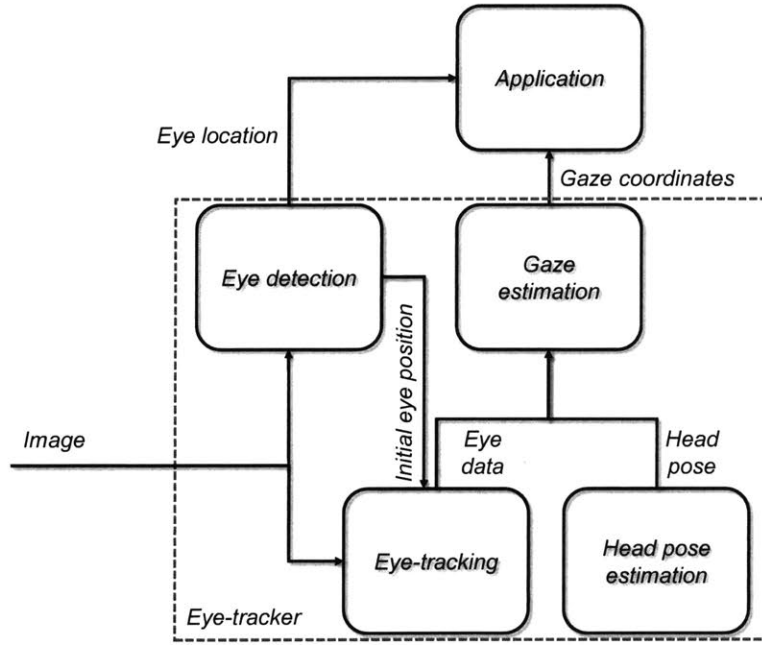
Figure 3-1: The basic components of a video-based eye-tracker. An eye-tracking system is typically composed of three processes: (1) eye-detection, (2) eye-tracking, and (3) gaze estimation. Image modified from [24].

Classification of eye-tracking and gaze estimation techniques

Eye-detection and tracking techniques can be broadly categorized into *shape-based*, *appearance-based*, and *hybrid* methods [24]. *Shape-based* techniques assume a predetermined model of the shape of the eye and use local point features (e.g. pupil and iris) or contours to model the eyes. *Appearance-based* techniques track the eyes directly using their photo-metric appearance as characterized by their color distribution or a filter response [24]. Conceptually, *appearance-based* techniques perform a template matching procedure and locate the eyes using a similarity measure. Lastly, *hybrid* methods benefit from the respective advantages of both approaches by combining them.

Gaze estimation aims to determine the direction of gaze by establishing the relationship between the image data and gaze direction [24]. The majority of gaze trackers use *feature-based* and *appearance-based* techniques. *Feature-based* techniques extract local features of the eye, such as the pupil or the corneal reflection, to estimate gaze. Because these features are easy to identify and can be formally related to gaze, *feature-based* methods are the most widely used [24]. *Appearance-based* methods do not extract eye features, but directly map the content of the image to gaze direction. These methods can be more robust than *feature-based* techniques, as eye feature extraction is prone to errors [24]. Examples of *appearance-based* methods are regression functions, Gaussian processes, and deep-learning techniques [24].

Challenges in eye-tracking and gaze estimation

Eye-detection and tracking remains an active area of research because it poses many unique challenges. One of them is the fact that the shape of the eye changes when viewed from different angles (Fig. 3-2). From some of these views, the eyelids appear straight and the iris circular, while in others the eyelids appear curved and the iris elliptical. Furthermore, eyelids are particularly challenging because they can partially or completely occlude the eyes. Another challenge is the variability in the appearance of the eyes across different subjects. Some of these variations are influenced or exacerbated by ethnicity, color texture, and light conditions [24].
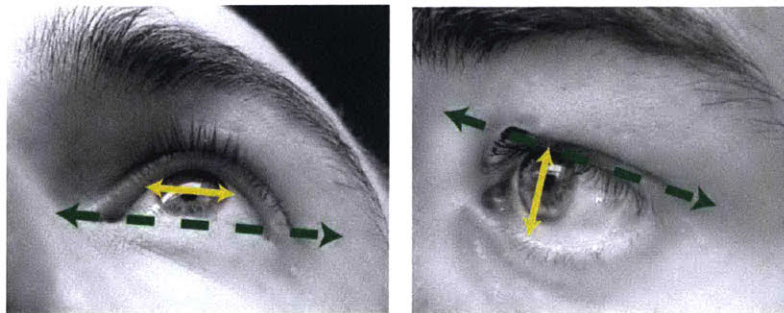

Figure 3-2: Changes in the shape of the eye when viewed from different angles. The dashed green arrows indicate when the eyelid appears straight, while the solid yellow arrows show the major axis of the iris ellipse. Image taken from [24].

In gaze estimation, head movements can introduce significant inaccuracies to the final gaze estimate [24]. This implies that the position of the head should be taken into account when estimating gaze, as a subject's gaze is determined by head pose and eyeball orientation. A typical approach for addressing this issue is to implicitly incorporate head pose information through a calibration procedure. However, the calibration becomes invalid if head pose varies considerably, in which case calibration can result in erroneous gaze estimates.

As described in the previous paragraphs, there exists a variety of video-based eye-tracking techniques. The following sections will discuss the most widely used techniques in clinical set-ups and mobile platforms.

3.1   Eye-tracking in clinical environments

Within the clinical field, a sizable portion of eye-trackers are video-based trackers that use *feature-based* techniques to estimate gaze [10, 23, 25, 59, 68]. These eye-trackers utilize active lighting schemes to enhance the lighting conditions and highlight features of the eye within the image [24]. The active light source is typically near infrared (IR) with a wavelength between 780 to 880 nm. The position of the IR light source determines whether the pupil appears darker or brighter than its surrounding eye structures in the resulting image. If the light source is on the same axis as the optical axis of the camera the pupil appears bright, while if it is off-axis the pupil appears dark (Fig. 3-3). The pupil appears bright when the light source is on-axis because the light reflects off the posterior part of the eye back to the camera [24].

The pupil center-corneal reflection technique

The primary method for estimating gaze using active lighting schemes is based on pupil and corneal reflection tracking [27]. The corneal reflection, usually referred to as the *glint*, is a small reflection on the cornea produced by an IR light source and utilized by video-based eye-trackers to estimate gaze (Fig. 3-3). Specifically, the length and direction of the vector between the center of the pupil and the glint, the pupil center-
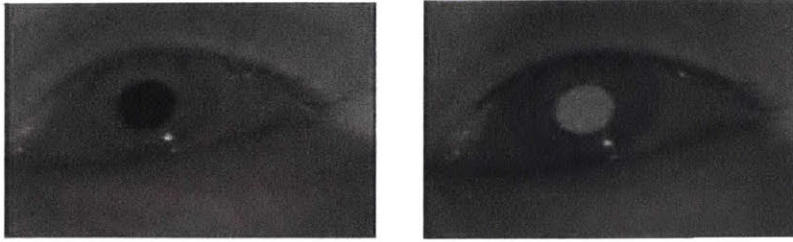
38

Figure 3-3: A set of IR-illuminated eye images. *(left)* Dark-pupil image versus *(right)* bright-pupil image. The images are created by varying the location of the light source relative to the optical axis of the camera. Notice the distinctiveness of the pupil and the small, bright spot below the pupil, i.e. the corneal reflection. Image taken from [24]

corneal reflection (PCCR) vector, are matched to gaze position through a calibration procedure, thus enabling gaze estimation. The assumption behind this technique is that if the relative position of the camera, illumination source, and head remain fixed, the glint is stationary [24]. This method consists of two major components: the extraction of the PCCR vector and the mapping of gaze through calibration. The position of the pupil and corneal reflection are typically extracted using *feature-* or *model-based* techniques [27]. *Feature-based* techniques use image thresholds and gradients to detect the pupil and the glint, while *model-based* solutions fit a model of the eye, such as an ellipse, onto the image. Once the pupil center and the glint are localized, a vector is traced between the two to create the PCCR vector (Fig. 3-4). During the calibration procedure, the subject directs their gaze toward a set of predefined calibration points displayed on a screen while minimizing head motions [27, 67]. A PCCR vector is recorded at each calibration point with known screen coordinates, allowing the calculation of a function that maps any screen coordinate to a PCCR vector. For example, assuming the PCCR vector v is represented as (v_x, v_y) and the screen coordinate is represented as $(x_{gaze,}, y_{gaze})$, gaze mapping functions can be modeled by the following set of nonlinear equations [67]:

$$x_{gaze} = a_0 + a_1 * v_x + a_2 * v_y + a_3 * v_x * v_y \tag{1}$$

$$y_{gaze} = b_0 + b_1 * v_x + b_2 * v_y + b_3 * v_y^2 \tag{2}$$

The coefficients a_0, a_1, a_2, a_3 and b_0, b_1, b_2, b_3 are estimated during calibration.
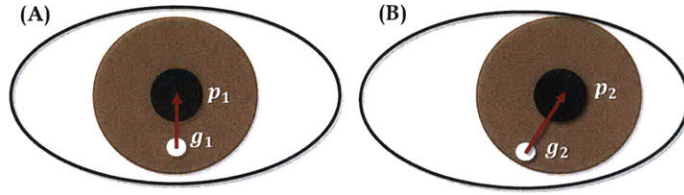
Figure 3-4: The pupil center-corneal reflection (PCCR) vector. Panel (A) shows the PCCR vector at one gaze location, while Panel (B) shows the vector at a different location. Assuming the relative position of the camera, illumination source, and head remain fixed, notice how the vectors are different when the subject views two distinct locations. Image modified from [67].

Challenges in clinical set-ups

Infrared video-based eye-trackers require distinct bright- or dark-pupil effects and a single corneal reflection to work well, making them sensitive to external illumination sources and eye glasses [24]. Another weakness of these systems is their vulnerability to partial or complete occlusions of the pupil or glint by the eyelashes or eyelids [27]. These occlusions might result in imprecise data for some parts of the visual field. Furthermore, IR video-based eye-trackers that use the PCCR technique are highly sensitive to head movement. The accuracy of the gaze estimation drops substantially once the subject moves their head away from the original position in which the system was calibrated [67]. As a result, these systems utilize specialized equipment, such as a chin rest, to restrict head motion. The constraints on head movement can be relaxed by using more cameras or more IR sources, but this requires additional equipment [27].

In addition to their dependency on IR illumination, another limiting factor of these systems is the high sampling frequency of the video-cameras [23, 25]. Conventionally, the lower end of high-speed systems is around 250 frames per second (fps), while the higher end is about 2,000 fps [27]. To measure saccade latency accurately, it is necessary to sample the saccade as closely as possible to the onset of the movement. This implies that the sampling frequency f_s introduces an error of $1000/f_s$ milliseconds to the latency measurement. Therefore, as the sampling frequency increases the measurement error decreases. Of course, higher sampling rates require specialized

and often expensive equipment, which is why the cost of these clinical systems is around $30,000 [1].

3.2   Algorithms to detect fixations and saccades

Event-detection algorithms analyze gaze-position data and classify all visual events as visual fixations or saccades. These algorithms are broadly categorized into dispersion-based algorithms and velocity-based algorithms [26]. The two types of event-detection algorithms are discussed in the following sections.

Velocity-based algorithms

Velocity-based algorithms use eye-velocity criteria to identify saccades within the gaze position data. Saccades are detected by specifying a fixed eye-velocity threshold and assuming that any motion above this threshold is a saccade [26]. Any data excluded by this criteria are assumed to be visual fixations. Eye-velocity is calculated by numerically differentiating the eye position curve [27]. As expected, numerical differentiation amplifies high-frequency noise and requires further processing to enhance the velocity curve (Fig. 3-5).

In practice, there is a lack of consistency in the specification of velocity thresholds, evidenced by the use of different thresholds that have varied from 20 deg/s to 75 deg/s [6, 30, 55]. Furthermore, the choice of threshold is influenced by the level of noise in the gaze position data - the higher the noise, the higher the threshold. Some studies have suggested data driven thresholds that adapt to the noise level and eliminate the need to manually specify a velocity threshold [46]. From this disscusion, it follows that the calculation of saccade latency is highly sensitive to the choice of velocity threshold. Recall that the beginning of a saccade is identified as when the velocity of said saccade exceeds a predefined threshold. Hence, higher thresholds result in larger saccade latency values compared to lower thresholds.

Although there are many implementations of velocity-based algorithms, the version by [55] is one of the most documented and widely used [27]. The goal of this
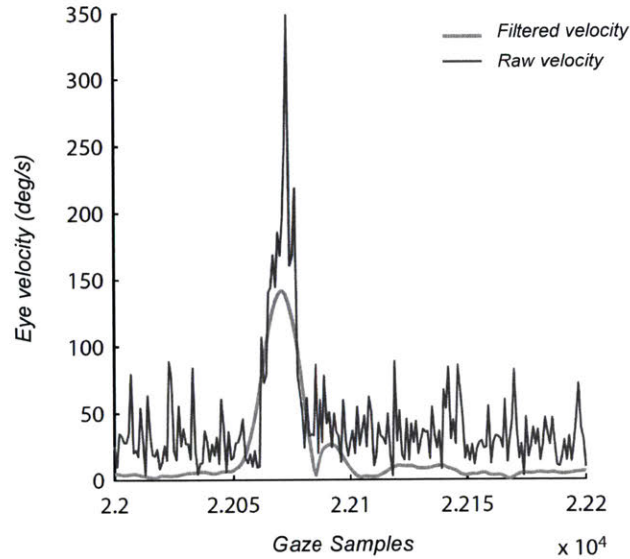
Figure 3-5: An unfiltered, noisy eye-velocity curve and its filtered counterpart. The raw velocity is calculated by numerically differentiating the eye position curve, while the filtered data was processed with a second order Savitzky-Golay filter. Image modified from [46].

two-pass algorithm is to detect saccades, their onsets, and their offsets by finding velocity peaks in the data that exceed a velocity threshold. Once the velocity peaks are identified, the data samples to the left and right of each peak are analyzed to determine the onset and offset of the saccade. By definition, visual fixations are everything that are not saccades.

Velocity-based algorithms are often used in high-speed eye-tracking systems. The main sequence relationship between saccade amplitude and duration (Fig. 2-3) establishes that most saccades last for around 20 to 80 ms. Hence, if the main goal is to analyze the properties of a saccade, a minimum sampling frequency of 200 Hz (4 - 16 samples per saccade) is required [26].

Dispersion-based algorithms

The most common event-detection algorithms are dispersion-based algorithms, which are generally applied to data acquired at low sampling rates (50 Hz or less) [26]. The main goal of dispersion-based algorithms is to identify visual fixations by finding gaze

position samples that are sufficiently close to one another for a minimum period of time. Among dispersion-based algorithms, the identification by dispersion threshold (I- DT) algorithm is the most widely used due to its simplicity, robustness, and speed [26]. The I-DT algorithm initializes a window of minimum length over the first few samples of gaze position data and calculates the dispersion within that window. Dispersion in this algorithm is defined as:

$$d = [max(x) - min(x)] + [max(y) - min(y)] \qquad (3)$$

where (x, y) represents the gaze samples within the window. Additional samples are added to the window until the dispersion is larger than a predefined threshold. Once the dispersion exceeds the threshold, the gaze position of the visual fixation is calculated as the centroid of the window samples. Lastly, the data samples within the window are removed and the process repeats until all samples are classified. As an example, if the dispersion threshold is 1 degree and the minimum window length is 100 ms, data samples that stay within a 1 degree diameter for at least 100 ms are considered visual fixations [26]. A pictorial representation of the I-DT algorithm is shown in Fig. 3-6.

Although the I-DT algorithm is implemented on a variety of commercial eye-tracking systems, the approach is very sensitive to noisy data and does a poor job at providing accurate temporal estimates of the onset and offset of events [46]. Moreover, the need and flexibility of specifying duration and dispersion thresholds can yield significant differences in the number and duration of fixations [7].

3.3   Eye-tracking on mobile platforms

In recent years, a small number of methods for estimating gaze on portable devices, such as tablets and smart-phones, have been proposed [29, 32, 62, 63]. Accurate eye-tracking in this environment is challenging given the absence of IR illumination and high-speed cameras, adverse lighting conditions, limited computational resources, and the instability of the tracking device. An analysis of existing literature shows that
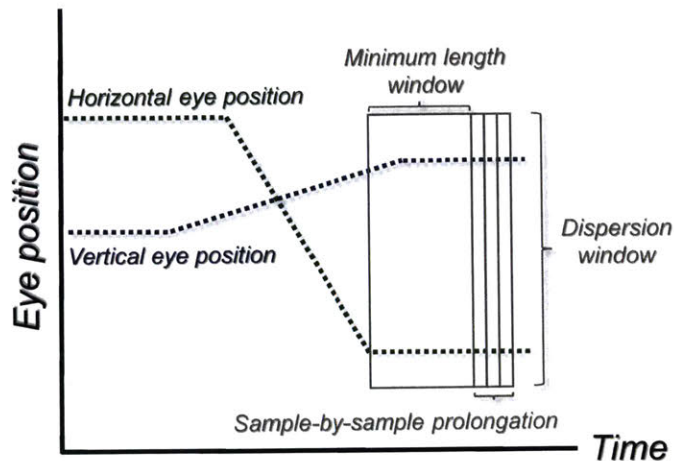
Figure 3-6: The identification by dispersion threshold (I-DT) algorithm. A temporal window of minimum duration is initialized over the gaze position data and expanded sample by sample until a dispersion threshold is exceeded. The centroid of the samples within the temporal window is the location of the visual fixation. Image modified from [46].

the most popular approaches to mobile gaze-tracking are *model-based* and *appearance-based*. The following sections discuss and compare both approaches.

Mobile gaze estimation with model-based techniques

EyeTab [62] is a *model-based* approach that fits a 2-D ellipse onto an eye image and back-projects it to estimate the optical axes. Firstly, EyeTab performs a rough estimate of the location of both eyes using Haar-like features [58]. The location estimate is refined using a shape-based approach that searches for the center of the eye under the assumption that the eye has a circular appearance. Once the region of interest is determined, a set of points along the limbus (the boundary between the white of the eye and the iris) is identified using the maxima of a radial derivative. The random sample consensus (RANSAC) method is subsequently used to eliminate outliers from the set of candidate feature points. After rejecting the outliers, the limbus is modeled as a 2-D ellipse. Lastly, the center of the 2-D ellipse is mapped to a 3-D position and the optical axis is determined. The complete EyeTab pipeline is

44

illustrated in Fig. 3-7.

Eight participants evaluated EyeTab on a tablet that yielded images at a resolution of 1920×1080 pixels. Image data was recorded while participants held the device and looked at nine discrete locations on the screen. The root-mean-square error (RMSE) between all true and estimated gaze points was reported as a measure of the gaze estimation accuracy. During evaluation, EyeTab produced gaze estimates at a speed of 12 fps with an accuracy of approximately 6.88 degrees (2.59 cm). These results allow for high-level interactions, but are far away from the performance of sophisticated eye-trackers (from 0.3 degrees to 2 degrees).
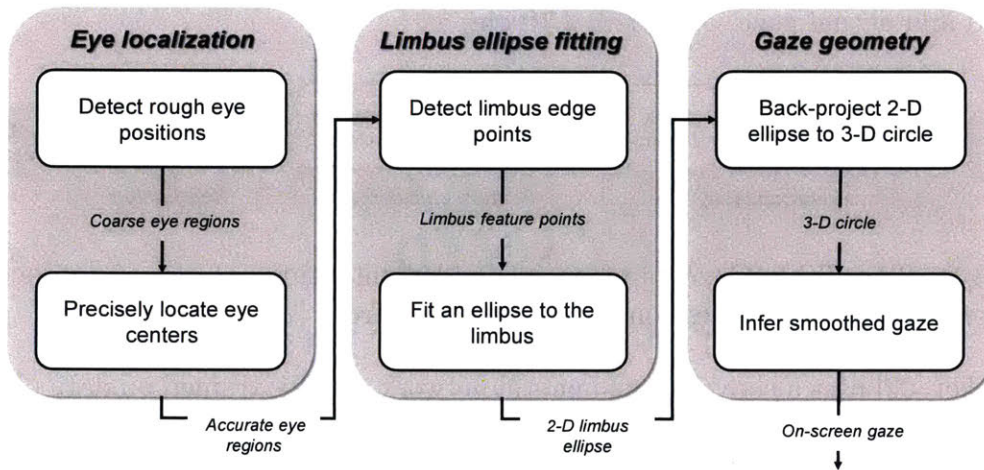
Figure 3-7: The EyeTab pipeline is divided into the stages of eye localization, limbus ellipse modeling, and gaze estimation. Image modified from [62].

Mobile gaze estimation with appearance-based techniques

TabletGaze [29] and iTracker [32] are *appearance-based* methods that use large datasets to train algorithms that estimate gaze on mobile devices. Neither of these techniques uses IR sources or chin rests, resulting in less constrained recording environments compared to those in the clinical literature. TabletGaze [29] aims to design a mobile gaze-tracker that works in unconstrained environments and does not require calibration. To achieve this goal, researchers collected 816 videos (captured with the tablet's front-facing camera) from 51 subjects and created the Rice TabletGaze dataset. During the data collection procedure, participants held the tablet in four different body

postures (standing, sitting, slouching or lying) and recorded four video sequences in each where they watched a dot across 35 different locations on the tablet's screen. In itself, the TabletGaze pipeline consists of three stages: preprocessing, feature extraction, and regression (Fig. 3-8). During preprocessing, a Haar feature detector localizes the eyes in the dataset images and crops them. Following the eye cropping, histogram of oriented gradients (HoG) features are extracted from the image and mapped to a lower dimensional space using Linear Discriminant Analysis (LDA). Finally, the extracted features are used by two Random Forests regression models (horizontal and vertical gaze locations) to predict the 2-D gaze position on the tablet's screen. This approach achieves a mean estimation error, a mean Euclidean distance between the predicted and actual gaze, of 3.17 ± 2.10 cm.
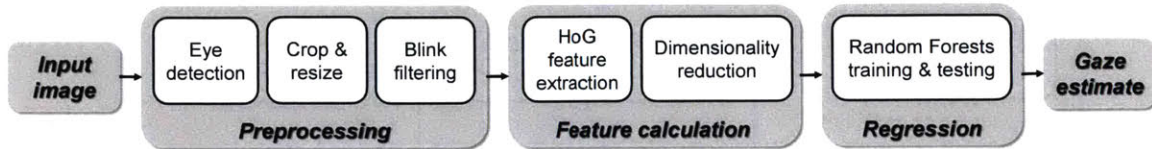
Figure 3-8: The TabletGaze pipeline is divided into three stages: preprocessing, feature extraction, and regression. Image modified from [29].

iTracker [32] uses a convolutional neural network (CNN), trained on data obtained from iPhones and iPads, to perform gaze estimation on mobile devices. Employing a similar approach, researchers used crowd-sourcing to record gaze data from 1,474 subjects while they followed a dot through random screen locations. During the data collection procedure, subjects were encouraged to change their head pose and phone distance to increase the dataset's variability. In total, the dataset consists of approximately 2.4 million frames that were used to design and train iTracker. iTracker takes as input the entire image, with the location of the face within the image, and a cropped image of each eye; the output is the distance in centimeters from the front-facing camera. During the evaluation phase, iTracker achieved a mean estimation error of 1.71 cm and 2.53 cm on iPhones and iPads respectively, attaining state-of-the-art performance. It remains to be seen whether both *model-based* and *feature-based* techniques can attain sufficiently precise saccade latency measurements for clinical applications.

Chapter 4

Data collection & eye-tracking algorithms for saccade latency

4.1 Data collection

The overarching goal of this work is to design and test a novel method that enables measurement and tracking of saccade latency outside of the clinical environment. Specifically, we aim to measure saccade latency on videos recorded with consumer-grade cameras. Acquiring accurate and precise data in this environment is challenging without the assistance of IR light or the availability of high-quality image sensors. Due to the lack of a publicly available saccade dataset, this project requires generating the videos on which the algorithms will be tested and that will serve as reference data in normal, healthy volunteers. To acquire this data, we implemented a visual task that allows us to measure saccade latency and recorded the saccadic eye movements of healthy individuals that performed this task. The data collection procedure was designed in collaboration with Hsin-Yu Lai.

4.1.1 Visual task design

The Psychophysics Toolbox 3 for Matlab [3] was used to implement and display the visual task. Of the previously discussed tasks (Section 2.5), we implemented a variant of the gap paradigm, in which the presentation of the visual fixation and the saccade

target are temporally separated by a period of time known as the gap. Specifically, an individual task started with a fixation period in which three squares were presented on the screen against a black background: a green square at the center of the laptop screen and two white squares arranged horizontally (Fig. 4-1). Subjects were asked to fix their gaze on the green square, which served as the visual fixation light. All three squares were displayed for 1.0 s, after which a gap period of 0.2 s ensued. Following the gap period, the two lateral squares reappeared in their original position but with one of them bounded by an additional yellow square (the saccade target). Subjects were tasked with moving their eyes to – and subsequently keeping their gaze fixed on – the saccade target (Fig. 4-1). After the saccade target disappeared, subjects returned their gaze to the centrally located green square. This task was repeated 40 times per trial, with 20 stimuli to the right and 20 stimuli to the left in a randomized order. Three such trials were conducted in one recording session, resulting in 120 saccade measurements per session.
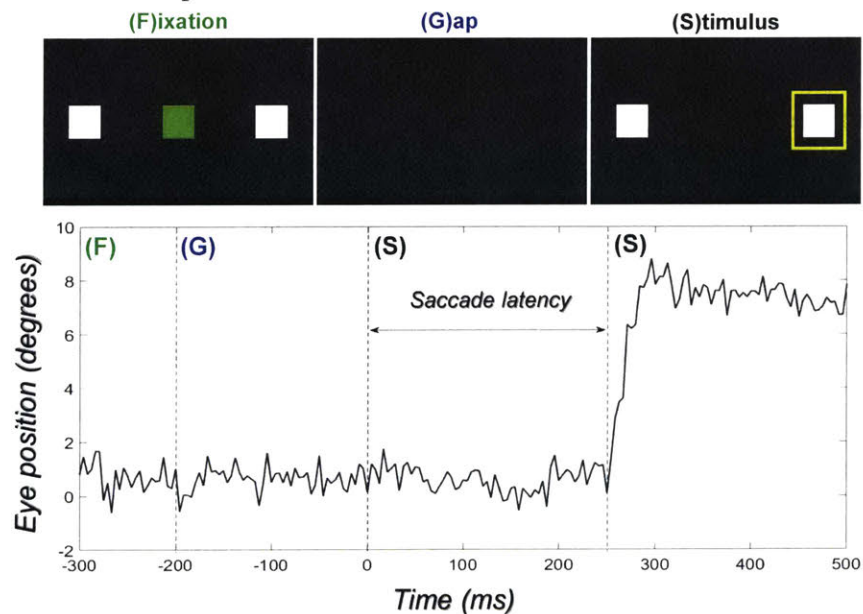

Figure 4-1: Top panel: An example of a single visual task during a saccade measurement. The three panels illustrate the laptop's display during each period of the task. Only the final 0.1 s of the 1.0 s fixation period is shown. Bottom panel: A corresponding saccadic eye movement during the visual task. The letters in bold indicate the fixation, gap, and stimulus periods.
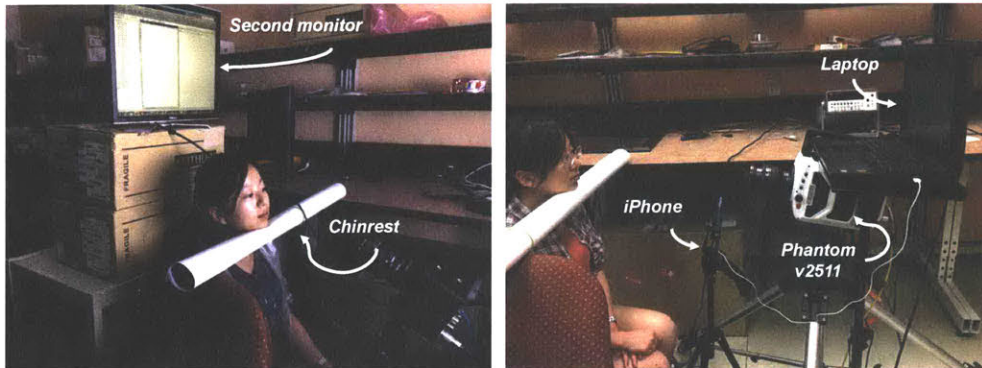
4.1.2 Video recordings

Figure 4-2: The video recording setup. On the left, the second monitor that mirrors the laptop's screen and the chinrest are labeled. To the right, the laptop, iPhone 6, and high-speed camera are identified.

Video recordings on volunteers were approved by MIT's Institutional Review Board, and informed consent was obtained from each participant prior to recording. Subjects were seated centrally in front of a laptop screen at a distance of about 0.8 meters, with their chin placed comfortably on a soft chin rest to minimize head movements. The laptop screen was used to display the visual tracking task. An iPhone 6 was centrally placed at a distance of 0.4 meters from the subject, and video recordings were made with the phone's rear-facing camera in slow-motion mode, resulting in frontal recordings at a frame rate of 240 fps and a full resolution of 1280×720 pixels. A second monitor was placed behind the subject's head mirroring the laptop's screen. 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 moment the visual stimulus appeared on the laptop screen.

In a subset of recordings, we simultaneously collected reference videos with a high-speed camera (Phantom v25-11 [41]) at 500 fps. The distance from the high-speed camera to the subject was about 0.6 meters; the camera lenses were focused on the subject's eyes. The laptop screen was displaced to accommodate for the high-speed camera's large size, increasing the distance between the subject and the screen from 0.8 meters to about 0.9 meters. The effect of this increased distance is that the

subject's horizontal eye movement is smaller during the visual task. The distance between the subject and the iPhone 6 was left unchanged (0.4 meters). The complete recording setup with the high-speed video camera is illustrated in Fig. 4-2. It follows that the typical set-up is the same without the inclusion of the high-speed video-camera.

Most recordings were acquired under the standard ambient lighting conditions of a laboratory environment. A separate set of recordings were taken under different illumination conditions with the assistance of two LED panel lights. In total, four distinct lighting scenarios were tested: (1) room light with the LED panel on a high setting, (2) room light, (3) the LED panel on a medium setting, and (4) the LED panel on a low setting (278 Lux, 220 Lux, 54 Lux, and 26 Lux respectively). The scenarios are organized from the best possible illumination condition to the poorest. Figure 4-3 shows how the lighting conditions affect the image brightness. Five subjects were recorded in each of the four conditions with and without glasses, for a total of eight (4×2) test cases.
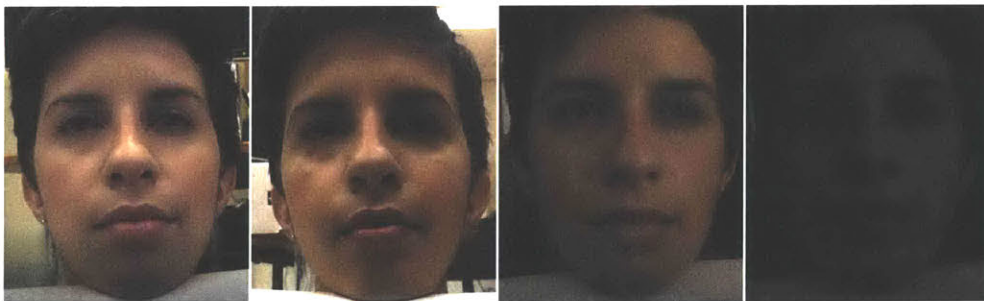

Figure 4-3: A sample frame from each video taken under four distinct lighting conditions. From left to right, the pictures are arranged from the best illumination to the worst.

4.2 Eye-tracking algorithms to measure saccade latency

The system pipeline consists of two main steps for acquiring saccade latency: (1) *eye-tracking* to extract the eye position from each frame, and (2) *saccade-onset detection* to determine when the eye begins to move. Fig. 4-4 shows the full pipeline, which processes the iPhone videos and produces the estimated saccade latency values. In this chapter, we discuss the process of developing the eye-tracking algorithm. The algorithm development was done in collaboration with Hsin-Yu Lai.
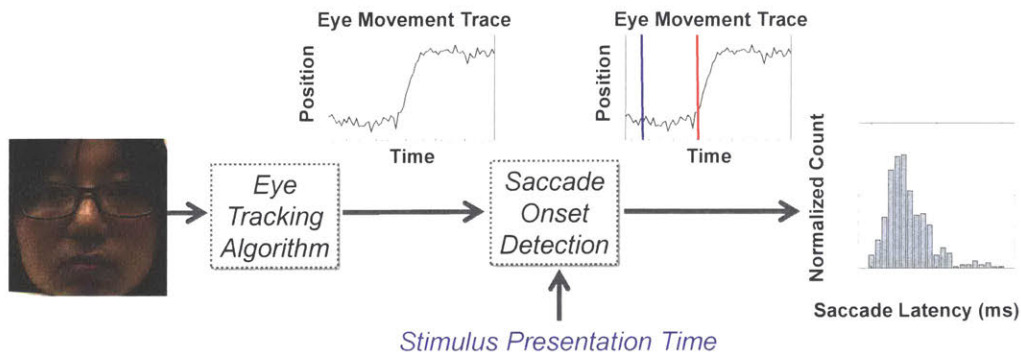
Figure 4-4: Pipeline for automated saccade-latency measurement, consisting of eye-tracking and saccade-onset detection. The eye-tracking algorithm takes as input the iPhone videos and outputs an eye position trace. This eye position trace is analyzed to calculate the onset of the saccade and provide a set of estimated saccade latency values, which are subsequently plotted as a histogram. On the diagram, the time difference between the stimulus presentation time (blue line) and the saccade onset (red line) is saccade latency.

The eye-tracking algorithm takes as input the recorded iPhone videos and outputs an eye movement trace that is presumably a saccade. The goal for developing this portion of the system was to extend the capabilities of preexisting gaze-estimation algorithms and tailor their performance to measure saccade latency. The standard performance metric of these algorithms – the accuracy of the gaze estimate – cannot be translated into the accuracy of saccade onset detection, since the former emphasizes the estimated position of the gaze while the latter requires a clear transition in the estimates whenever the gaze position changes. As a result, we discuss how we can use iTracker, the state-of-the-art mobile gaze estimator, on our data and then propose

modifications that are more suitable for saccade onset detection.

An alternative eye-tracking algorithm, the Starburst algorithm [37], was also considered in this work. Although not discussed in this chapter, details of the Starburst algorithm can be found in Appendix A.

4.2.1 The iTracker algorithm

iTracker is a CNN trained on images taken with the frontal camera of an iPhone or iPad that predicts where a user is looking at on a mobile screen [32]. The inputs of iTracker include a cropped left eye, a cropped right eye, a cropped face, and a face grid that indicates the location of the face within the frame. The eye and face crops are of size 224×224, while the face grid is of size 25×25. Face and eye detection is needed to acquire the corresponding crops. iTracker's architecture, as well as information on the size of its layers, is seen on Fig. 4-5.
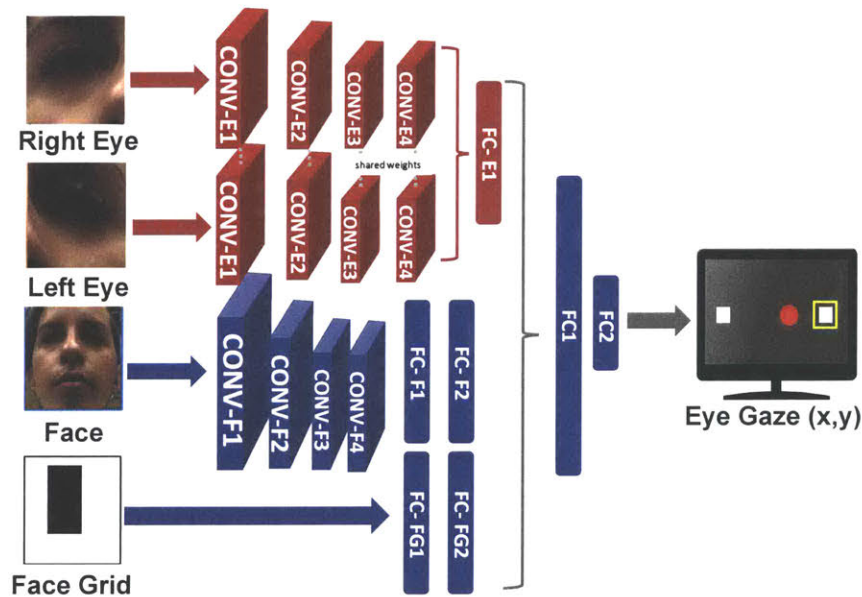
Figure 4-5: CNN architecture used by iTracker and iTracker-face. iTracker processes the eye and face layers *(red and blue)*, while iTracker-face *(blue)* only processes the face layers. CONV means convolutional layers with filter size/number of kernels: CONV-E1, CONV-F1: 11×11/96, CONV-E2, CONV-F2: 5×5/256, CONV-E3, CONV-F3: 3×3/384, CONV-E4, CONV-F4: 1×1/64. FC means fully-connected layers with sizes FC-E1: 128, FC-F1: 128, FC-F2: 64, FC-FG1: 256, FC-FG2: 128, FC1: 128, FC2: 2. Image is modified from [32].

The output of the neural network is the estimated gaze position on the screen relative to the device's front-facing camera in centimeters. Specifically, iTracker outputs the gaze estimate as a measure of the horizontal and vertical distance from the frontal camera.

GazeCapture: A large-scale dataset

A large-scale dataset called GazeCapture was built and utilized to train iTracker. The data for GazeCapture was collected through iPhones and iPads, while the subject recruitment was done through crowd-sourcing. Through the GazeCapture app, participants were shown a series of dots on the mobile device's screen while their gaze was recorded using the device's front-facing camera. To increase the variability in the dataset, subjects were asked to continuously move their head, change their distance to the device, and vary the device's orientation. As a result, GazeCapture contains 2,445,504 images from 1,474 subjects. To train iTracker, 1.5 million frames out of the 2.4 million were utilized; the remaining data were divided into train, validation, and test splits.

iTracker evaluation

During evaluation, the average Euclidean distance from the location of the true fixation (in centimeters) was reported as measure of accuracy. It was found that iTracker consistently outperforms the current baseline in gaze estimation with an average error of 1.71 cm and 2.53 cm in iPhones and iPads correspondingly. An ablation study that removed different components of the model revealed that removing the face grid decreased the accuracy of the gaze estimate the most, while removing the eye layers decreased the accuracy the least.

To apply iTracker on our data, we manually cropped each eye region by annotating the two corners of each eye on the first frame and surrounding the midpoint of these two corners with a square of a width 1.5 times the distance between the two corners. The face crop is determined by surrounding the center of six points – the four eye corners and the two mouth corners – with a square of a width 1.5 times the largest

distance between two points from the six annotated corners. All the images that are fed into iTracker have a resolution of 224×224 which means they undergo resizing from the original cropped images; the eye crops are upsampled, while the face crop is downsampled. Because the subjects only perform horizontal saccades during the visual task, the horizontal gaze estimate over time corresponds to the eye-movement trace.

While iTracker is designed to operate on video images taken at 30 fps, a temporal resolution below 20 ms is required for clinical applications [16]. Thus for this work, we use the rear facing camera at 240 fps, which has poorer image quality compared to 30 fps due to the reduction in exposure time as shown in Fig. 4-6.[1]

Figure 4-6: Frame rate impact on image quality. *(left)* 30 fps *(right)* 240 fps.

4.2.2 The iTracker-face algorithm

To address the low quality eye-crop image at high frame rate, we propose the iTracker-face algorithm, where we only use the face-related convolutional layers as shown in Fig. 4-5. Although this does degrade the accuracy of the gaze estimation as discussed in [32], our objective is to only determine if the gaze changes, and thus it does not significantly affect our results. In addition, eliminating the eye-related layers reduces the number of multiply-and-accumulate (MAC) operations in the neural-network by 66%, which translates to a significant reduction in computation. Specifically, the raw number of MACs in iTracker-face is about 700 million, while the original iTracker algorithm has 2.1 billion. The number of weights in iTracker-face is also considerably less than iTracker, reduced from 7.5 million to 3.2 million. Fig. 4-7 shows an example eye trace using the iTracker and iTracker-face algorithms. Note that iTracker-face has lower noise than iTracker.

---

[1]This is not an issue with high-end image sensors such as those found in the Phantom high-speed camera. However, this is one of the reasons the cost for those cameras are in the tens to hundreds of thousand dollars.
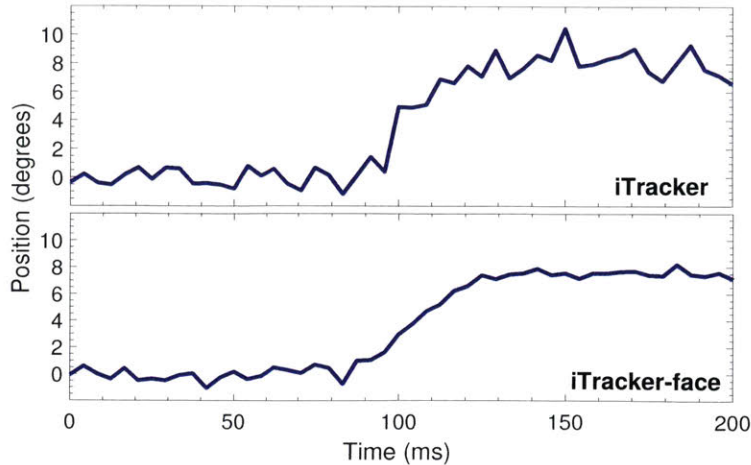
Figure 4-7: Sample traces for the two eye tracking algorithms.

4.2.3 Algorithm robustness evaluation

To determine the robustness of iTracker and iTracker-face in multiple environmental conditions, we compared their performance on video sequences of subjects with and without glasses and under various illumination conditions. As mentioned in Section 4.1.2, four distinct lighting scenarios were tested: (1) room light with the LED panel on a high setting, (2) room light, (3) the LED panel on a medium setting, and (4) the LED panel on a low setting (Fig. 4-3). Five subjects were recorded in each of the four conditions with and without glasses, for a total of eight (4 × 2) test cases.

All videos were processed with both iTracker and iTracker-face, generating a set of eye movement traces that were reviewed by two expert annotators. Each annotator separately determined if an individual eye movement trace represented a horizontal saccade and had a sufficiently high signal-to-noise ratio to allow for saccade-onset determination. Saccades that met these criteria were labeled 'good'; all other eye movement traces were labeled 'bad'. Some examples of 'bad' saccades are saccades interrupted by blinks, initially directed toward the opposite direction, or with a low signal-to-noise ratio (more in Section 5.2). The algorithm with the highest amount of 'good' saccades (according to the two annotators) was deemed the most robust. In total, each annotator labeled the same 9,600 saccades and spent approximately 10 - 15 hours each doing so.

Table 4.1: Percentage of 'agreed good' and 'agreed bad' traces among two expert annotators in different environmental conditions.

Illumination		High LED with Room Light		Room Light		Medium LED		Low LED	
Glasses		No	Yes	No	Yes	No	Yes	No	Yes
Agreed Good (%)	iTracker	82.5	72.2	64.3	30.2	75.8	45.2	27.8	7.8
	iTracker-face	83.2	84.5	92.2	85.8	88.2	80.3	74.2	42.5
Agreed Bad (%)	iTracker	12.5	16.0	19.3	45.5	15.3	37.3	62.7	79.7
	iTracker-face	13.7	11.8	5.8	8.0	8.3	12.7	19.3	42.3
Agreed Traces (%)	iTracker	95.0	88.2	83.7	75.7	91.2	82.5	90.5	87.5
	iTracker-face	96.8	96.3	98.0	93.8	96.5	93.0	93.5	84.8

Table 4.1 shows the percentage of traces both annotators agreed where 'good' or 'bad' for each illumination and glasses condition. As shown in Figure 4-7, the iTracker traces are noisier than the iTracker-face traces. Intuitively, it is easy to understand how a low signal-to-noise ratio makes the onset of a saccade ambiguous. Therefore, if the saccade traces for an illumination condition have a low signal-to-noise ratio, the percentage of 'agreed bad' traces and the disagreement between annotators increase. As shown in Table 4.1, the percentage of 'agreed good' traces for iTracker-face was higher than for iTracker in all four illumination conditions with and without glasses. Consistently, the percentage of 'agreed bad' traces in iTracker-face was lower than iTracker in all but one condition (high LED with room light without glasses). However, the percentages are comparable at 13.7% and 12.5% correspondingly (82 and 75 saccades respectively). This discrepancy is a result of recording in an optimum environmental condition, where it is possible to observe subtle eye movement dynamics in the saccade trace that increase the agreement between annotators.

In general, the level of agreement between annotators is higher in iTracker-face ('agreed traces' = 'agreed good' + 'agreed bad') in all but one condition (low LED with glasses). This last condition is the most challenging, producing traces with the lowest signal-to-noise ratio, and hence decreasing the agreement between annotators. However, the annotators agreed that there is a higher percentage of 'good' saccades in the iTracker-face traces compared to the iTracker traces in this condition. From

these results, we observe that because the traces produced by iTracker-face have a higher signal-to-noise ratio than the traces produced by iTracker, the percentage of 'good' traces and the agreement between annotators is higher. This supports the notion that iTracker-face is more robust to variations in the environmental conditions than iTracker.

Table 4.1 also shows that the performance of both algorithms deteriorates if the subject is wearing glasses. In the "high LED with room light" condition, iTracker-face's performance is consistent regardless of glasses, while iTracker performs worse with glasses. The basis of this observation is the fact that the percentage of 'agreed good' traces decreases for iTracker, while it is constant for iTracker-face. In the "room light" and "medium LED" conditions, the performance of iTracker-face worsened slightly for the "glasses" case, while it worsened significantly for iTracker. This is based on the observation that the percentage of 'agreed good' traces for iTracker decreased by at least 30%, while the percentage of 'agreed bad' increased by at least 20%. In comparison, the percentage of 'agreed good' traces for iTracker-face decreased by only 6-8%. In the "low LED" condition with and without glasses, the annotators agree that less than a third of the iTracker traces are 'good'. Although about 75% of the iTracker-face traces are labeled 'good' in the no glasses condition, less than half of the data is reliable in the glasses condition.

This analysis shows that the performance of both algorithms deteriorates when subjects are wearing glasses, especially in conditions with poor illumination. A closer inspection of the eye crops in these video recordings shows that heavy-rimmed glasses tend to cast shadows over the eye region, leading to a darker image. Certain eye features, such as the pupil or iris, are harder to discern in these conditions which might impact the quality of the gaze estimation. However, it is clear that iTracker-face is more robust to poor illumination and glasses than iTracker.

4.3 Summary

In this chapter, we introduced the data collection set-up (Fig. 4-2) and the automated pipeline for saccade latency measurement (Fig. 4-4). This discussion focused on *eye-tracking*, where the main goal is to extract an eye position estimate from each frame in a video sequence. The state-of-the-art mobile gaze-tracking algorithm, iTracker, was evaluated and modified to enable suitable saccade onset detection in our video recordings. iTracker is a CNN trained on images taken with the frontal camera of an iPhone or iPad that processes numerous inputs (a cropped left eye, a cropped right eye, a cropped face, and a face grid) to produce a gaze estimate (Fig. 4-5). The modified version of iTracker, iTracker-face, only analyzes the face crop and face grid, producing eye movement traces with a higher signal-to-noise ratio.

A robustness evaluation between iTracker and iTracker-face showed that iTracker-face is more robust than iTracker to challenging recording scenarios, such as low illumination and the presence of eye glasses. Although the performance of both algorithms progressively deteriorates with less illumination, iTracker-face still outperforms iTracker by a large margin in the worst illumination condition. When subjects are wearing glasses, the performance of both algorithms further deteriorates. A closer inspection of the eye images in these video recordings shows that the eye glasses cast shadows over the eye region, leading to a darker image. The eyes are harder to discern in these conditions, which might impact the resulting gaze estimation. This effect is exacerbated when the recording is taken in poor illumination conditions.

Chapter 5

Modeling the saccade waveform

As was mentioned in Section 4.2, the measurement system pipeline consists of two main stages: *eye-tracking* and *saccade-onset detection* (Fig. 4-4). Once the eye-tracking algorithm produces the eye movement trace, the saccade is analyzed to identify its onset. With prior knowledge of the moment in time when the saccade target appeared on the laptop screen, we are able to calculate the saccade latency.

By convention, the onset of a saccade is defined as the moment in time when the velocity of said saccade exceeds a predefined threshold (see Section 2.3). Although there is a lack of consistency in the choice of velocity threshold across studies (see Section 3.2), the methodology for obtaining eye velocity is consistent enough. Saccade velocity is typically calculated by numerically differentiating the eye position trace [27]. As expected, this method amplifies high-frequency noise in the gaze position data, which is why filtering of the data is required for subsequent analysis (Fig. 3-5). A wide range of filters are utilized across studies, but a body of work suggests that the parameters of a saccade, such as the peak velocity and duration, can vary as a function of filter type and threshold [30, 39]. For example, filters tend to "stretch" the saccade and lead to reductions in the peak velocity as well as extensions of the saccade duration [39]. Avoiding the use of filters might be permissible for recordings with very high SNR, as those obtained from high-end cameras and IR-based eye-tracking systems. However, given our use of consumer-grade electronics to record eye movements, point-by-point differentiation would substantially amplify

high-frequency noise in the eye-position data, resulting in higher-than-average noisy estimates of velocity and thus, saccade onset. To avoid these issues, we propose fitting a mathematical model to the raw eye position traces and differentiating *the model* instead of the raw data. The modeling work described in this chapter was mostly developed by Hsin-Yu Lai.

5.1 Previous modeling work

A recent, but small, amount of work has been published on developing models for saccadic waveforms [17, 39]. The motivation behind this work is to provide objective saccade traces, free from inherent instrumentation noise, by which to compare saccade detection algorithms, sampling rates, and filter choices. Equations for modeling the saccadic waveform range from exponential functions [17] to trigonometric functions [39]. Dai [17] proposed modeling the saccadic waveform $s(t)$ as the sum of a ramp function $f(t)$ and a shifted negated ramp function $-f(t - \tau)$, specifically:

$$s(t; V_{max}, C, \tau) = Cf(V_{max}t/C) - Cf(V_{max}(t - \tau)/C) \tag{1}$$

where,

$$f(t) = \begin{cases} t + 0.25e^{-2t}, & t \geq 0 \\ 0.25e^{2t}, & t \leq 0 \end{cases} \tag{2}$$

and V_{max} is the asymptotic peak velocity described by the main sequence (about 500 deg/s), while C is a constant that determines the proportionality between the saccade amplitude and the peak saccadic velocity. The parameter τ determines a point on the main sequence relationship curve, thus specifying the amplitude and peak velocity of a simulated saccade $s(t)$. The parameters V_{max} and C determine the main sequence relationship equation between saccade amplitude and peak velocity (see Section 2.1). Two additional parameters, t_0 and s_0, are added to the model to specify the saccade onset time and the initial position of the saccade,

$$S(t; V_{max}, C, \tau, t_0, s_0) = s(t - t_0; V_{max}, C, \tau) + s_0. \tag{3}$$

The proposed model was fit to a dataset of saccades from 24 healthy subjects, where an evaluation showed that it outperforms other models when fitting large-amplitude saccades (≥ 9 degrees). Furthermore, the proposed model has the added benefit of conforming to the main sequence that describes the relationship between saccade amplitude and peak velocity.

Separate work [39] proposed using a hyperbolic tangent to model saccadic waveforms:

$$p(t) = \int_{-\infty}^{t} v(\tau)d\tau = \frac{a_\infty(a)}{2}(\tanh(g(t)) + 1) \tag{4}$$

$$v(t) = v_{max}^*(a) \operatorname{sech}^2(g(t)) \tag{5}$$

$$g(t) = 2f(a)(t - \phi) \tag{6}$$

where $p(t)$ is the position profile of the saccade, $v(t)$ is the velocity profile, and $g(t)$ is a time scaling function. The variable a is the saccade amplitude, $a_\infty(a)$ is the asymptotic amplitude of the saccade for $t \to \infty$, and ϕ is the time of occurrence of the peak velocity $v_{max}^*(a)$. The function $f(a)$ is a function that determines the time scale of the saccade:

$$f(a) = \frac{v_{max}^*(a)}{a_\infty(a)} \tag{7}$$

A comparison between the model and several measured saccades showed that the simulated velocity profiles are symmetric, while this is generally not the case for the measured data. Real saccades seem to be skewed with longer deceleration than acceleration phases, which become more obvious at larger amplitudes [39]. However, the frequency content of the simulated saccades and the measured data is similar.

5.2   Our modeling approach

To calculate saccade latency, it is necessary to determine the onset of the eye movement toward the saccade target. The traditional approach of performing point-by-point differentiation on the eye position trace would substantially amplify high-frequency noise given our use of consumer-grade electronics to record eye movements.
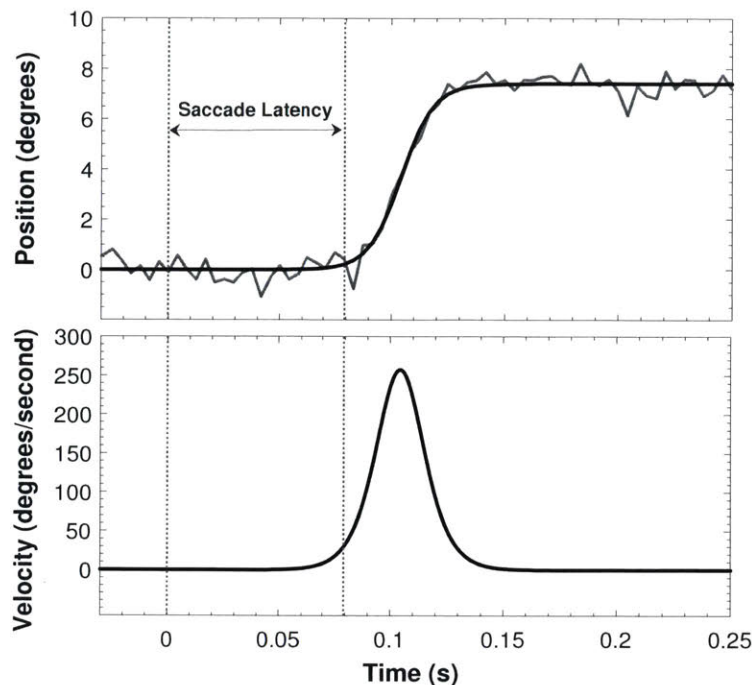
Figure 5-1: Top panel: Eye position as estimated by the modified iTracker algorithm (gray) and hyperbolic tangent fit (black). Bottom panel: Eye velocity obtained by differentiating the hyperbolic tangent fit. The dashed line at 0 s indicates the moment of stimulus presentation. The saccade onset is determined by an increase in eye velocity above 30 deg/s (second dotted line). The saccade latency is the time difference between the dotted lines.

Similar to [39], we used `lsqcurvefit` in Matlab to fit a hyperbolic tangent model of the form:

$$\tilde{x}(t) = E + F \cdot \tanh\left(\frac{t - G}{H}\right) \tag{8}$$

to the eye-position data during the saccade period (100 ms before the stimulus presentation and 500 ms after the stimulus presentation) and differentiate the resulting best-fit solution to obtain a smooth eye-velocity trace to threshold (Fig. 5-1). We identify the saccade onset by looking for the moment in time in which eye velocity exceeds 30 deg/s. The time delay between the saccade onset and the moment of stimulus presentation (which we gather from the video) is the final estimate of saccade latency.

In our tanh model, the parameters E, F, G and H are the four scale and translation parameters that define the hyperbolic tangent. When performing the tanh fit,

62

we assume that the subject always moves their gaze from the fixation light to the saccade target. Hence, the amplitude of the saccade (and the tanh) is always scaled to 10 degrees. In this sense, we do not perform gaze estimation during the saccade onset detection.

The NRMSE as a goodness-of-fit metric

This model-based approach has the added 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. The best-fit normalized root-mean-squared error (NRMSE) between the model and the eye-position trace quantifies the residual discrepancy between the two and is defined as,

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\tilde{x}_i - x_i)^2}{n}} \tag{9}$$

$$\text{NRMSE} = \frac{\text{RMSE}}{10 \text{ degrees}} \tag{10}$$

where $\tilde{x}$ is the tanh model, x is the raw data, and n is the number of individual gaze estimates in the raw data. The NRMSE is the RMSE normalized by the amplitude of the saccade, in our case, 10 degrees.
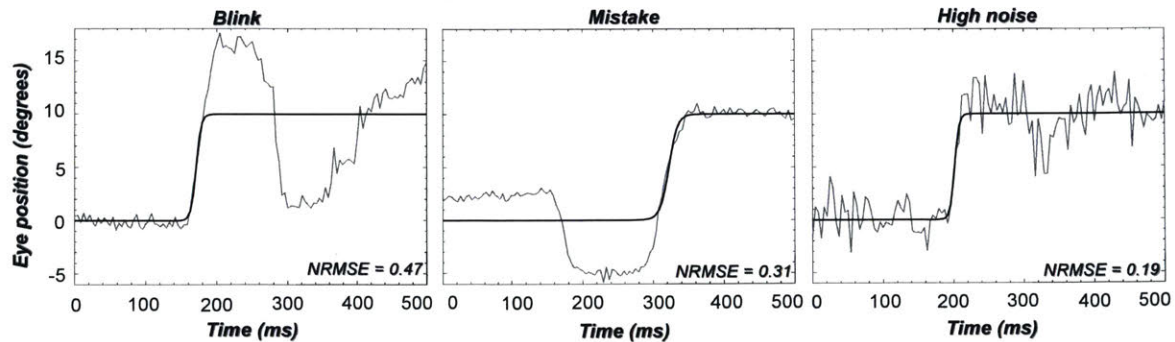

Figure 5-2: Examples of rejected saccades that were contaminated by blinks, initially directed toward the opposite direction (mistake), or had a low signal-to-noise ratio. The eye position as estimated by the modified iTracker algorithm is in gray, while the hyperbolic tangent fit is in black. The NRMSE between the gray and black lines is shown for reference.

63

Unreliable measurements or fits typically have a high NRMSE, while reliable measurements have a low NRMSE. Thresholding the NRMSE allows automated rejection of recordings in which the saccade onsets might have been erroneously detected. Rejected saccades include those contaminated by blinks or initially directed toward the opposite direction, as well as noisy eye traces (Fig. 5-2).

Evaluation of the NRMSE as a goodness-of-fit metric

To evaluate the effectiveness of the NRMSE as a goodness-of-fit metric, we used the 9,600 expert-annotated saccades as ground truth labels to generate a receiver operating characteristic (ROC) curve. The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) at various NRMSE thresholds, evaluating the ability of the metric to discriminate against 'good' and 'bad' saccades. In the end, the NRMSE threshold with the best trade-off between TPR and FPR is chosen. The area under the curve (AUC) of the ROC curve is typically reported as a measure of the effectiveness of the classifier. The value of this parameter ranges from zero to one, where one is a perfect classification and zero is at the other end of the spectrum.

In our experiments, there are two separate annotations of ground truth data made by two expert annotators who labeled the same set of 9,600 saccades. Thus, the labels assigned by each annotator are separately considered ground truth, yielding two ROC curves (red and blue) as shown in Fig. 5-3. Not only did the two ROC curves closely follow each other (showing consistency between the annotators), but they both attained an AUC of around 0.92. Moreover, the width of the double-sided 95% confidence interval for the AUC is 0.02, showing the high precision of the estimate. An extensive evaluation of the NRMSE thresholds showed that the best trade-off between the TPR and FPR occurs at an NRMSE value of 0.1, with an average TPR and FPR of 0.87 and 0.20 correspondingly for the first annotator, and 0.86 and 0.16 respectively for the second. Thus, all modeling results with an NRMSE higher than 0.1 are considered unreliable estimates of saccade latency and are removed from the final analysis.
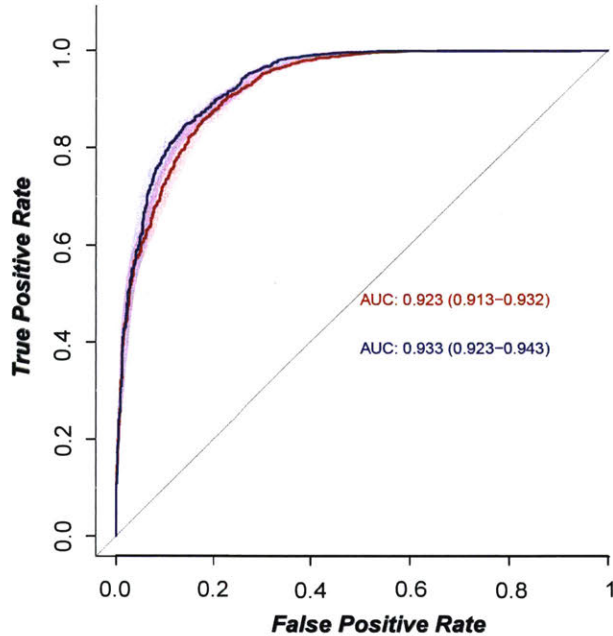
Figure 5-3: Two receiver operating characteristic (ROC) curves; one for each annotator. The area under the curve (AUC) for each annotator is also given with the 95% confidence interval. Image created by Hsin-Yu Lai.

As mentioned in Section 4.2.3, manual annotation of one recording session takes an expert annotator about one and a half hours to complete. Using the NRMSE as a goodness-of-fit metric speeds this process by automatizing the saccade latency measurement pipeline and leveraging the use of mobile phones to acquire a substantial number of saccade latency values.

5.3   Summary

Here we discussed a model-based technique to detect the onset of a saccade – the second step in the measurement system pipeline (Fig. 4-4). Saccade onset is typically calculated by numerical differentiation of the eye position curve, which yields the velocity profile of a saccade. Saccade onset is defined as the moment in time when the velocity exceeds a predefined threshold, such as 30 deg/s. To avoid amplification of high-frequency noise, we instead fit a hyperbolic tangent model to the raw eye position data and differentiate *the model* to obtain a velocity curve.

The NRMSE between the best-fit model and the eye-position trace is used as a goodness-of-fit metric to automatically discard unreliable saccade latency estimates. A high NRMSE indicates the measurement is unreliable, while reliable measurements have a low NRMSE. Two expert annotators reviewed 9,600 saccades to create ground truth labels and generate a ROC curve. The ROC curve is used to evaluate the classification performance of the NRMSE and choose a threshold on the NRMSE that is the best trade-off between the TPR and the FPR. In the end, an NRMSE of 0.1 provided the best results, with a TPR of about 0.86 and a FPR of at most 0.20. Hence, all modeling results with an NRMSE higher than 0.1 are not considered in the final analysis.

Chapter 6

Saccade latency distribution analysis

After completing the last step in the measurement system pipeline, *saccade onset detection*, the final output are the estimated saccade latency values (Fig. 4-4). Once a subject is recorded following the data collection protocol, we process the videos using the pipeline and attain (at most) 120 saccade latency values per recording session (Chapter 4). These values are analyzed as a distribution of saccade latencies for each individual participant (see Section 2.3). The mean (μ) and standard deviation (σ) of the control group and patient population distributions are commonly reported in clinical studies. Chapter 6 focuses on the analysis of these saccade latency distributions. This analysis was done in collaboration with Hsin-Yu Lai.

6.1 Comparison across cameras

Before analyzing the saccade latencies obtained with the iPhone camera, we compared this data with data recorded using a high-speed, research-grade camera. Our goal is to prove that the saccade latency distributions from the consumer-grade camera (iPhone 6, cost < \$1k, rolling shutter, 720p resolution, 240 fps) have a similar μ and σ to that of the research-grade camera (Phantom v25-11 [41], cost $\sim$ \$100k, global shutter, 720p resolution, 500 fps). For this comparison, we took simultaneous recordings

on four subjects using both cameras to verify that we can attain similar saccade latency distribution statistics. Including the high-speed camera in the recording set-up resulted in an increased distance between the subject and the laptop's screen. The increased distance results in smaller horizontal eye movements, which in turn produce noisier, but acceptable, eye movement traces. Fig. 6-1 shows the resulting saccade latency distributions obtained using the iTracker-face algorithm and the hyperbolic tangent model, demonstrating that the distributions from both cameras are consistent with essentially the same mean and standard deviation. A two sample permutation test (significance level of 0.05) was used to test the null hypothesis that the difference between the experimental means in each pair of saccade-latency distributions is zero. None of the means were significantly different across all four subjects ($p > 0.05$).
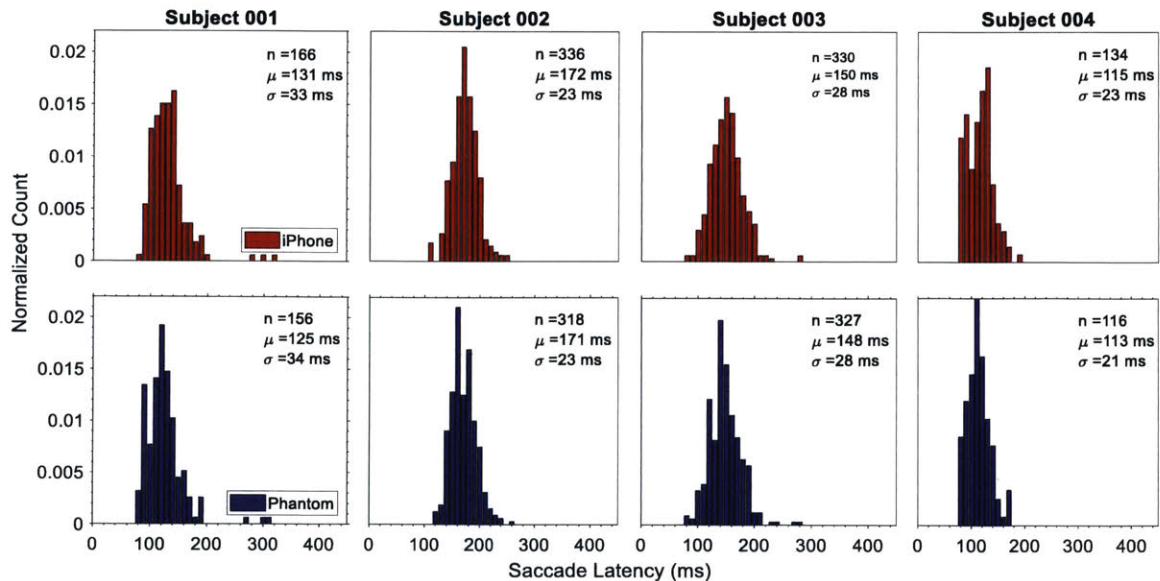

Figure 6-1: Example saccade latency distributions obtained from *(top)* a consumer-grade camera (iPhone) and *(bottom)* a high-speed camera (Phantom [41]) on four subjects.
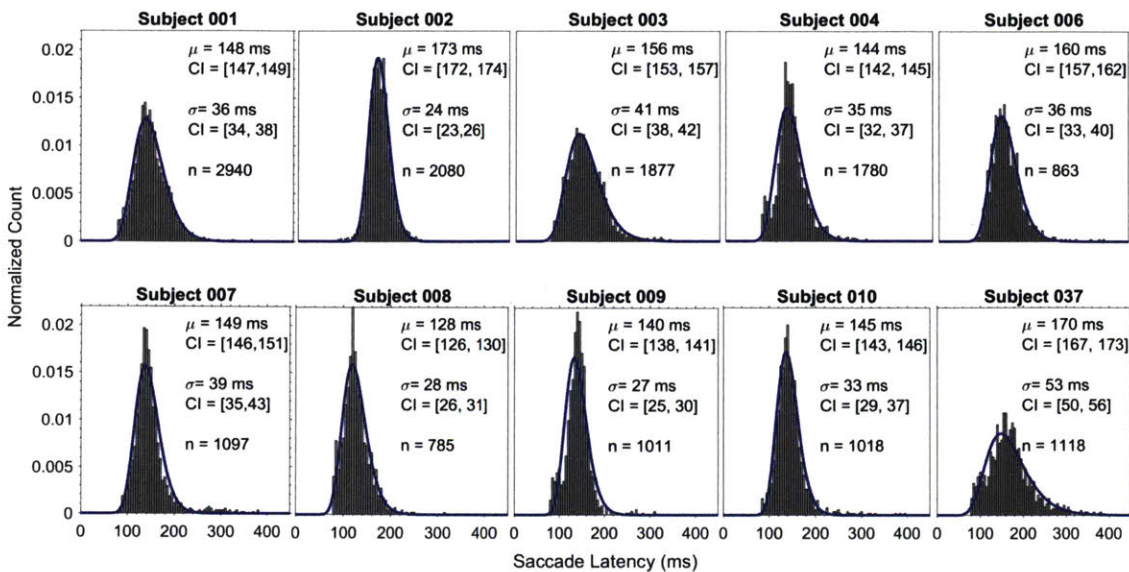
6.2 Saccade latency in healthy individuals

Figure 6-2: Saccade latency histograms for ten healthy individuals. μ is the sample mean, σ is the associated sample standard deviation, and n is the total number of observations. The 95% confidence interval (CI) for the sample mean and standard deviation are shown below each test statistic. Saccade latencies below 80 ms were censored. The estimated log-normal probability density functions are shown in blue.

In this work, a total of 20,000 saccades were recorded in 30 healthy subjects. Among the 30 subjects, we recorded 10 on at least 10 separate occasions (sessions). The aggregate saccade-latency distributions of these ten subjects are shown in Fig. 6-2. The sample mean saccade latency across subjects ranges from 128 to 173 ms, while the sample standard deviation varies from 24 to 53 ms. All saccade fits with a NRMSE< 0.1 were included in the histograms. Rejected saccades include those contaminated by blinks or initially directed toward the opposite direction, as well as noisy eye traces. Saccade latencies smaller than 80 ms were censored, as they were considered anticipatory movements [54]. Taking these rejection criteria into account, the average fraction of good saccades per recording session across subjects is 81 $\pm$ 18% (about 97 $\pm$ 22/120 saccades). Some subjects had a tendency to initiate saccades

toward the opposite direction or make anticipatory movements, which reduced the fraction of good saccades per session.

The sample mean and standard deviation of the individual aggregated distributions (Fig. 6-2) were bootstrapped to produce a 95% confidence interval (CI) for these parameters. Bootstrapping helps quantify the variability around an estimated parameter by resampling the sample data with replacement. Once we have an estimate of the variability around a specific parameter, we can compute the desired CI. In this particular experiment, the process of resampling and recalculating the test statistics (μ and σ) was repeated for 5,000 iterations. The 95% CI for the sample mean and standard deviation of each saccade latency distribution is illustrated in Fig. 6-2.

The saccade-latency distributions from individual subjects show variable degrees of positive skewness, with saccade latencies larger than 200 or 300 ms not uncommon (Fig. 6-2). The hypothesis that reaction times follow a log-normal distribution was tested by fitting censored (at 80 ms) log-normal distributions to the saccade-latency data (Fig. 6-2). To robustly fit a log-normal model, the distribution of the log-saccade latencies is inspected. If the saccade latency data is truly log-normally distributed, a log transformation of the distributions should yield normal distributions truncated at $log(80(\text{ms}))$. Normally distributed data have a cumulative distribution function (CDF) that when plotted on a probit scale [8] appears as a straight line. The slope and x-intercept of this line are the mean and standard deviation of the normal distribution. Hence, by fitting a line to the probit-scale CDF, the parameters of the log-normal distribution are estimated.

The Kolmogorov-Smirnov test (significance level of 0.05) was used to test the null hypothesis that the experimental saccade-latency distributions can be described by a truncated log-normal distribution. Among the individual saccade-latency distributions (one for each recording session) across all subjects, 77 out of 82 (94%) distributions were not significantly different from a log-normal distribution ($p < 0.05$). When the individual data for each subject were aggregated into a single distribution, like in Fig. 6-2, seven out of ten distributions were not significantly different from a log-normal distribution ($p < 0.05$).

In addition, the distribution of the mean saccade latencies across 30 subjects is reported. The histogram in Fig. 6-3 shows the distribution of the mean saccade latencies from the first recording session across all subjects, and indicates that there is significant variability in mean saccade latencies even in the healthy population.

6.3 Longitudinal analysis of saccade latency

Fig. 6-4 shows the mean saccade latency across recording sessions for ten subjects. Four of the ten subjects were trained subjects, i.e., they had prior familiarity with the visual tracking task. Although the recording sessions are sequential, they were not taken on consecutive days. The error bars in the figure represent one sample standard deviation above and below the mean of each session. The dashed lines connecting each subject's mean saccade latency are the best-fit regression lines.

The coefficient of determination, R^2, and confidence limits on the slope parameters were calculated. The R^2 values for the regression lines were between 0.01 and 0.84. The slopes of the linear regression models were statistically different from zero in four of the ten subjects. Of the four, three had a negative slope and one had a positive slope.
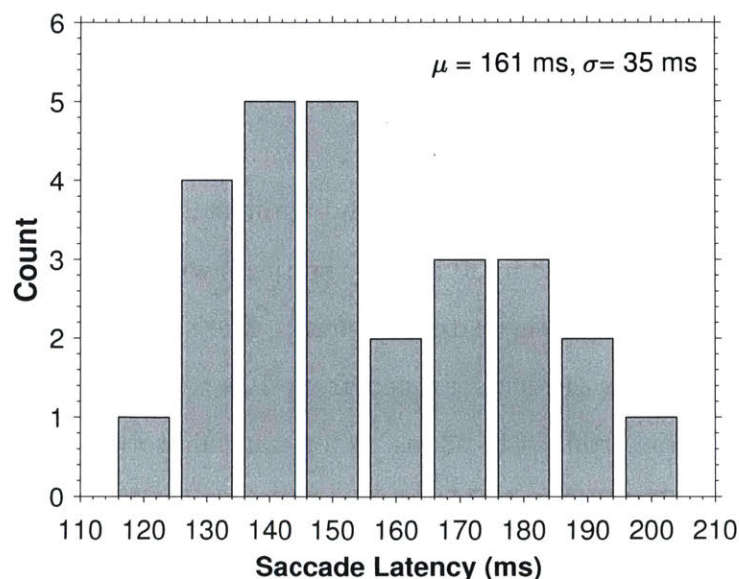

Figure 6-3: Distribution of mean saccade latency values from 30 healthy individuals.
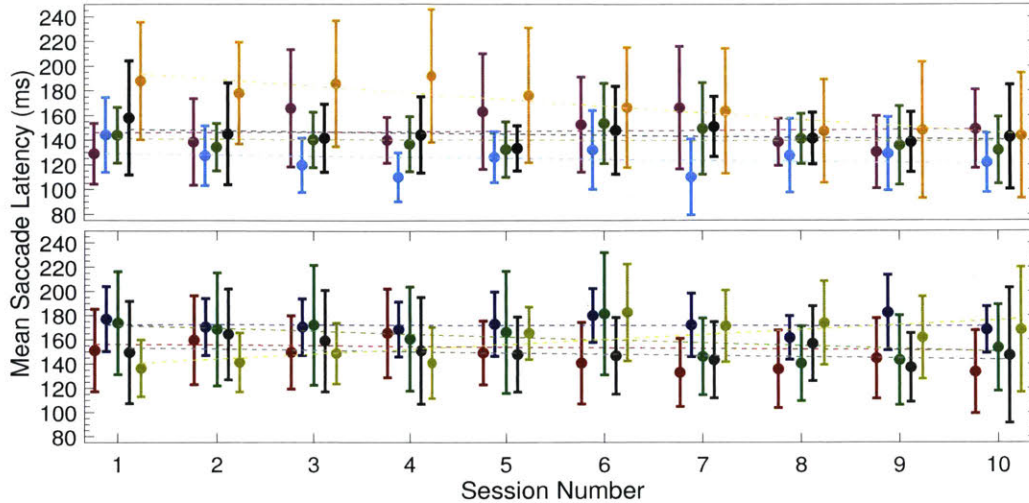
Figure 6-4: Changes in the mean saccade latency across recording sessions. An individual subject is represented by a distinct color. The error bars specify one sample standard deviation above and below the mean for each recording session. The dashed lines indicate linear regression on the mean values across recording sessions.

6.4 Summary

In this chapter, we presented the results of an analysis performed on the saccade latencies measured with our system. Before analyzing the saccade latencies recorded with the iPhone, we compared them with values recorded using a high-speed, research grade camera. To compare, we took simultaneous recordings on four subjects using both cameras and compared the mean and standard deviation of their respective saccade latency distributions. The resulting distributions from both cameras are consistent, with essentially the same mean and standard deviation.

We recorded 20,000 saccades in 30 healthy subjects with our system, ten of which were recorded on at least ten separate occasions. The sample mean of the saccade latency distribution across subjects ranged from 128 to 173 ms, while the sample standard deviation varies from 24 to 53 ms. A log-normal probability density function was fit to the ten distributions and a Kolmogorov-Smirnov test (significance level 0.05) showed that seven distributions were not significantly different from a log-normal distribution.

A plot of the mean saccade latency across recording sessions for ten subjects is seen in Fig. 6-4. A linear regression on the mean saccade latencies across recording sessions showed that a linear model can only explain 1% to 84% of the variability in the mean saccade latency. The slopes of the lines were statistically different from zero in four of the ten subjects.

Chapter 7

Discussion and next steps

In this chapter, we discuss the results from Chapter 6, summarize the key contributions of the thesis, and examine possible opportunities to expand upon the limitations of this work.

7.1 Discussion

Saccade latency in healthy individuals

Our system allowed for the measurement of 20,000 saccades in 30 healthy subjects, in which 120 saccade measurements take less than eight minutes. This is a substantial amount compared to the thirty something saccades per subject recorded in some clinical, video-based, eye-tracking studies [9, 10, 23, 25]. Considering the criteria to reject certain saccade latencies, we are able to retain, on average, a sizable amount (81%) of data per session across subjects (about 97 saccades). This amount of data allows us to draw significant conclusions about the behavior of saccade latency across healthy individuals.

One of the goals of this work is elucidating whether repeat measurements of saccade latency can be used to assess neurocognitve decline. Therefore, it is important to understand the behavior and variability of saccade latency in healthy subjects. Often in case-control studies, saccade latencies from multiple subjects are pooled together into one of two distributions: the healthy subjects distribution or the patient group

distribution. Unfortunately, saccade latency intra-subject variability is neglected with this approach. In this work, between 780 and 2,900 saccade latencies were measured in each of ten subjects, where the individual mean saccade latencies ranged substantially from 128 ms to 173 ms and the standard deviation from 24 ms to 53 ms (Fig. 6-2). Furthermore, the 95% confidence intervals on the sample mean saccade latency and standard deviation suggest that these parameters are distinct across individuals (Fig. 6-2). The variation across subjects is also documented in Fig. 6-3, where significant variability in mean saccade latencies is observed. These results support the notion that saccade latency intra-subject variability can be quite substantial and suggest that the common approach of pooling data from different subjects obscures important intra-subject variation that might be important to track.

According to the results of the Kolmogorov-Smirnov test, seven out of the ten distributions in Fig. 6-2 were not significantly different from a log-normal distribution. This result is expected considering that three of the distributions (subjects 004, 009, and 037) appear to be bimodal, with the smaller mode centered around 100 ms. However, the saccade latency distribution of the majority of subjects may be modeled as a log-normal distribution. This observation is consistent with [57], 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.

Longitudinal analysis of saccade latency

In our longitudinal analysis, we evaluated how the mean (μ) and the standard deviation (σ) of the saccade-latency distributions changed across time. These statistics were chosen over the mode or the median for two major reasons: (1) the mode, mean, and median tracked one another closely; and (2) clinical studies usually report the mean and standard deviation. In [52], we observed that multiple measurements of saccade latency per individual across days had a substantial amount of intra-subject variability. To reach this observation, the mean and standard deviation of the saccade-

latency distributions were analyzed as a function of the individual recording sessions. The recording sessions are sequential but were not necessarily taken on consecutive days.

With only five recording sessions from ten subjects, it was difficult to quantify the variability in these measurements across time [52]. In this work, we augmented our dataset by recording five additional sessions (ten sessions total) in ten subjects. A linear model on the mean saccade latency across sessions per individual showed that the slope of this line is not significantly different from zero in six out of ten subjects (Fig. 6-4). That is, although the variation in the mean and the standard deviation remain significant, both parameters seem to be stationary across time for the majority of subjects. Furthermore, the values of the coefficient of determination, R^2, suggest that the variability in these measurements cannot be exclusively explained by the number of recording sessions. For example, there are other factors (such as tiredness or test-taking strategies) that may introduce variability. More work is needed to elucidate and isolate these other effects.

7.2 Conclusions

This thesis paves the way for non-obtrusive monitoring of physiological variables with embedded platforms. Specifically, it proposes a system that measures saccade latency outside of a clinical environment, overcoming many of the challenges that make traditional clinical set-ups expensive and inaccessible. The key contributions of this thesis include:

- An automated pipeline to measure saccade latency outside of the clinical environment without the need of specialized equipment, such as IR illumination and research-grade cameras.

- A modified algorithm for determination of eye movements in video-recordings from consumer-grade devices.

77

- A complete assessment of the robustness of said algorithm to challenging recording environments.

- A model-based approach for determining the onset of a saccade.

- A goodness-of-fit metric that enables automated outlier detection and widespread data collection without the need for visual inspection of saccade traces.

- A thorough analysis and quantification of the saccade latency distributions of normal, healthy subjects and the intra-subject variability among them.

- A preliminary framework for the longitudinal analysis of saccade latency in normal, healthy subjects.

These contributions represent substantial advances towards recording saccade latency from a broad population, paving the way for understanding the behavior of this measurement in the general public. A deeper understanding of saccade latency is necessary to put into perspective the saccade latency changes seen in patients with neurocognitive decline.

7.3   Future work

Possible future work may be divided into two aspects: data acquisition and processing, and analysis of the saccade latency distributions. On the former, one venue of exploration could be gathering a thorough understanding of the effects of the visual task on saccade latency. In Section 2.5, we briefly discussed a set of visual tasks that are present in the clinical literature. Although not mentioned in this thesis, there are further variations that are known to engage brain areas that might be damaged by neurodegenerative disease. Hence, modifying the complexity of the visual task and understanding its effect on saccade latency could possibly yield more informative results than the current visual task (see Section 4.1.1).

Another area of improvement is the portability of our saccade latency measurement set-up. In spite of the fact that our system is already more accessible than

current clinical set-ups, a data collection system that can run entirely on a phone (without relying on an additional laptop or duplicate monitor) will further enable widespread data collection. This type of set-up would require using the front-facing iPhone camera, which can only record video at a speed of 30 fps: an eighth of the speed at which we record video currently (240 fps). Hence, for this to be feasible, it is necessary to push the limits of the current system and measure saccade latency at a much lower frame rate.

One innovation on the data processing front would be to develop a deep understanding of the mechanisms employed by the modified eye-tracking algorithm (iTracker-face). Even though we gather better eye movement traces with iTracker-face, we lack an intuition as to why this is the case. Knowledge in this aspect will help us tailor the algorithm further to our specific needs.

The analysis of the saccade latency distributions can improve substantially by implementing sophisticated models that quantify the subtle properties of a saccade latency distribution. In this work, a lognormal probability density function was used to model the distribution of saccade latencies. However, it was shown that the lognormal model is not suitable for some participants that exhibit two modes in their distribution. Models with higher complexity can quantify these subtle characteristics and improve the mathematical characterization of the distributions. Also, the longitudinal analysis of saccade latency can be enhanced by the exploration of other types of models and data representation schemes that shed additional light into the longitudinal behavior of this measurement.

Another enhancement to the data analysis procedure could be incorporating other saccadic metrics. Although the only eye movement measurement discussed in this thesis is saccade latency, there are other parameters - such as the error rate, saccade peak velocity, and saccade duration - that form a richer set of digital biomarkers and might be predictive of neurocognitive decline. This goes hand in hand with developing or comprehending mechanistic models that explain the effect of neurodegenerative disease progression on saccade latency. We believe that this level of physiological understanding will help us develop better models of saccade latency.

Appendix A

The Starburst algorithm

Early on, the Starburst algorithm [37] was considered a candidate eye-tracking algorithm for saccade onset detection. It was developed for images recorded with a consumer-grade camera and used eye-tracking techniques that were similar to those seen in clinical environments. This appendix contains a description of Starburst, its modifications, and the results of a performance evaluation on the modified version.

A.1   Algorithm description

Starburst is a *feature-* and *model-based* algorithm developed for a head-mounted eye-tracking system [37]. It relies on infrared (IR) illumination to provide a sharp boundary between the pupil and iris to extract the location of the pupil center and corneal reflection. Once the pupil center and corneal reflection are localized, a PCCR vector (see Section 3.1) is formed and related to coordinates in a scene.

The first step in the Starburst algorithm is to locate and remove the corneal reflection on the eye image. Localization of the corneal reflection is crucial to the PCCR vector and its removal reduces spurious detections of the pupil-iris boundary. The corneal reflection is detected using an adaptive intensity threshold on the image, in which the brightness and size of the corneal reflection are used as discriminant features.

Corneal reflection localization and removal is followed by a *feature-based* method

that detects the contour of the pupil. Individual feature points along the pupil-iris boundary are detected with multiple rays that extend from a central best guess of the pupil center [37]. The intensity gradient along each ray is computed until a fixed threshold is exceeded; a feature point is declared at that location in the ray. To improve the robustness of the algorithm to poor initialization, the process is repeated for each of the candidate feature points until the center of all candidate feature points converges to a stable location. That is, multiple rays extend from each of the candidate feature points to detect more features along the pupil contour.

The feature detection process occasionally selects erroneous features that do not belong to the pupil contour. The Random Sample Consensus (RANSAC) paradigm is used to eliminate these outliers, where an outlier is a sample generated through error that does not belong to the pupil contour. This framework achieves its goal by choosing a subset of candidate feature points, fitting an ellipse model to them, and selecting the model with the highest agreement between all candidate feature points. Subset selection and model evaluation is repeated for a number of iterations and by the end of the process, an ellipse is fit to the largest set of inliers. The ellipse is a model of the detected pupil boundary, where the center of the ellipse is the estimated pupil center. The complete Starburst algorithm is applied to each frame individually, producing an estimate of the pupil center and corneal reflection for each frame in the video.

A.2   Modifications

The original Starburst algorithm was optimized for images taken with a head-mounted device using IR illumination. This fixed camera pose relative to the eyes ensures that the eye is always in the same region relative to the camera, which makes algorithm initialization easy across trials. However, the benefits of IR illumination and head-mounting no longer hold when the eye movement is captured with an iPhone camera with a varying pose under natural light.

To address these limitations, we developed Starburst-phone: a modified version of

the original Starburst algorithm. One of the main modifications is that the iris contour is modeled instead of the pupil contour. This is motivated by the observation that in visible-spectrum imaging the iris boundary is often more distinct than the pupil-iris boundary [37] (Fig. A-1). Additionally, a circle model is fit to the iris contour rather than an ellipse. A circle has fewer parameters compared to an ellipse and gives a more stable estimate with fewer feature points. Contrary to a head-mounted device, the camera pose can vary with an iPhone, and thus the eye-crop position is manually determined. The same eye-crop position can be used for all frames by assuming minimal head movement during each test, which lasts under two minutes. The pupil center estimate used in the ray generation procedure is manually initialized in the first frame of each test. However, subsequent frames initialize the pupil center based on the previous frame, which allows for some minor head movement.
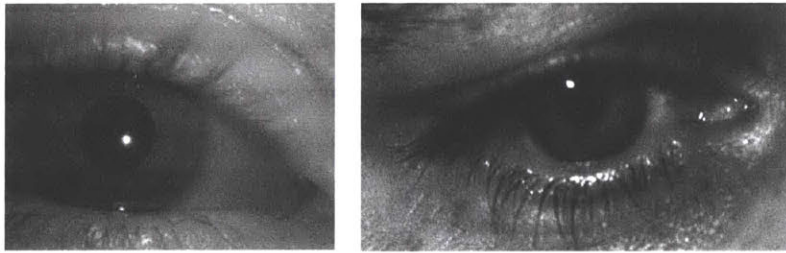

Figure A-1: Eye images with *(left)* infrared [37] versus *(right)* natural light. The iris contour is more visible under natural light than the pupil contour.

Fig. A-2 shows how the rays are generated from this initialization point and the gradient along each ray is calculated. The iris contour is detected by choosing the point with the maximum gradient along each ray rather than choosing the point that first exceeds a fixed threshold. Because the algorithm is modified to measure the iris boundary instead of the pupil boundary, it is important to consider that the upper eyelid partially occludes the former. Hence, the directions in which the rays are generated are restricted accordingly to prevent erroneous candidate feature points and reduce outliers. Finally, histogram equalization is selectively applied to the eye image for adaptation to the various lighting conditions.
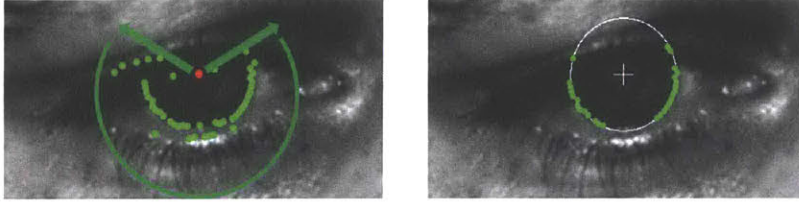
Figure A-2: The Starburst-phone algorithm operating under natural light. (*left*) Iris contour detection that avoids the upper eyelid. The red dot is the estimated center of the pupil from which the rays extend, while the green are the candidate feature points. (*right*) Iris model fitting after RANSAC succesfully eliminated outliers. Image from [33].

A.3 Algorithm performance evaluation

Similar to what is described in Section 4.1.2, the performance of Starburst-phone was evaluated on video sequences of subjects with and without glasses and under different levels of lighting. The illumination conditions in this evaluation consisted of three settings: 'room light', 'room light with LED panel lights on a low setting', and 'room light with LED panel lights on a high setting'. The subjects were recorded without glasses on the varying illumination scenarios. Only one expert annotator (compared to two) reviewed all saccade traces to determine whether each represented a typical saccade movement and had a sufficiently high signal-to-noise ratio to allow saccade-onset determination. Saccades that met these criteria were labeled 'good' - all other eye movement traces were labeled 'bad'. For the iTracker and iTracker-face evaluation described in Section 4.2.3, 9,600 saccades were reviewed by two expert annotators. Considerably less saccades were reviewed in this evaluation, with only 200 saccades in the glasses/no-glasses category and 40 saccades in the varying illumination scenarios.

The evaluation reults showed that Starburst-phone was the least robust algorithm compared to iTracker and iTracker-face. Only 11% of saccades were labeled 'good' by the expert annotator in the video sequences with glasses, while 43% were deemed reliable in the poorest illumination condition. In comparison, 91% of the iTracker-face saccade traces were labeled 'good' in the glasses condition and 90% in the worst illumination condition. The poor performance of Starburst-phone occurs because the algorithm mistakenly detects the frame of the glasses as part of the iris contour. Under insufficient lighting, it has difficulty detecting the iris-sclera boundary. With

fewer feature points detected on the iris contour, the iris model fitting frequently fails. In view of these results, Starburst-phone was no longer considered a candidate eye-tracking algorithm for saccade onset detection.

Bibliography

[1] Eyelink 1000 plus: The most precise and versatile eye tracker. `https://www.sr-research.com/products/eyelink-1000-plus/`. Accessed: 2018-08-13.

[2] Outlander anatomy: The eyes – part 2. `https://www.outlanderanatomy.com/tag/eyeball/`. Accessed: 2018-08-06.

[3] Psychophysics Toolbox 3. `http://psychtoolbox.org/`. Accessed: 2018-02-09.

[4] C.A. Antoniades and C. Kennard. Ocular motor abnormalities in neurodegenerative disorders. *Eye*, 29(2):200–207, 2015.

[5] C.A. Antoniades, Z. Xu, S.L. Mason, R.H.S. Carpenter, and R.A. Barker. Huntington's Disease: Changes in sacccades and hand-tapping over 3 years. *Journal of Neurology*, 257(11):1890–1898, 2010.

[6] J.A. Beintema, E.M. van Loon, and A.V. van den Berg. Manipulating saccadic decision-rate distributions in visual search. *Journal of Vision*, 5(3):150–164, 2005.

[7] P. Blignaut. Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, Perception, & Psychophysics*, 71(4):881–895, 2009.

[8] C.I. Bliss. The method of probits. *Science*, 79(2037):38–39, 1934.

[9] C. Bonnet, J. Hanuška, J. Rusz, S. Rivaud-Péchoux, T. Sieger, V. Majerová, T. Serranová, B. Gaymard, and E. Růžička. Horizontal and vertical eye movement metrics: What is important? *Clinical Neurophysiology*, 124(11):2216–2229, 2013.

[10] A.L. Boxer, S. Garbutt, W.W. Seeley, A. Jafari, H.W. Heuer, J. Mirsky, J. Hellmuth, J.Q. Trojanowski, E. Huang, S. DeArmond, J. Neuhaus, and B.L. Miller. Saccade abnormalities in autopsy-confirmed frontotemporal lobar degeneration and Alzheimer's Disease. *Archives of Neurology*, 69(4):509–517, 2012.

[11] D. Braun and B.G. Breitmeyer. Relationship between directed visual attention and saccadic reaction times. *Experimental Brain Research*, 73(3):546–552, 1988.

[12] D. Braun, H. Weber, T. Mergner, and J. Schulte-Mönting. Saccadic reaction times in patients with frontal and parietal lesions. *Brain*, 115(5):1359–1386, 1992.

[13] C.J. Bruce, M.E. Goldberg, M.C. Bushnell, and G.B. Stanton. Primate frontal eye fields. ii. physiological and anatomical correlates of electrically evoked eye movements. *Journal of Neurophysiology*, 54(3):714–734, 1985.

[14] J.D. Connolly, M.A. Goodale, R.S. Menon, and D.P. Munoz. Human fmri evidence for the neural correlates of preparatory set. *Nature Neuroscience*, 5(12):1345–1352, 2002.

[15] L.S. Costanzo. Neurophysiology. In *Physiology*, chapter 3, pages 65–110. Saunders, Elsevier, Pennsylvania, 2014.

[16] T.J. Crawford, S. Higham, T. Renvoize, J. Patel, M. Dale, A. Suriya, and S. Tetley. Inhibitory control of saccadic eye movements and cognitive impairment in Alzheimer's Disease. *Biological Psychiatry*, 57(9):1052–1060, 2005.

[17] W. Dai, I. Selesnick, J.-R. Rizzo, J. Rucker, and T. Hudson. A parametric model for saccadic eye movement. In *Proceedings of the IEEE Conference on Signal Processing in Medicine and Biology (SPMB)*, pages 1–6, 2016.

[18] J.F.X. DeSouza, R.S. Menon, and S. Everling. Preparatory set associated with pro-saccades and anti-saccades in humans investigated with event-related fmri. *Journal of Neurophysiology*, 89(2):1016–1023, 2003.

[19] 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. *International Psychogeriatrics*, 19(4):647–656, 2007.

[20] M.C. Dorris and D.P. Munoz. A neural correlate for the gap effect on saccadic reaction times in monkey. *Journal of Neurophysiology*, 73(6):2558–2562, 1995.

[21] E.R. Dorsey, S. Papapetropoulos, M. Xiong, and K. Kieburtz. The first frontier: Digital biomarkers for neurodegenerative diseases. *Digital Biomarkers*, 1(1):6–13, 2017.

[22] O. Ferhat and F. Vilariño. Low cost eye tracking: The current panorama. *Computational Intelligence and Neuroscience*, 3:1–14, 2016.

[23] S. Garbutt, A. Matlin, J. Hellmuth, A.K. Schenk, J.K. Johnson, H. Rosen, D. Dean, J. Kramer, J. Neuhaus, B.L. Miller, S.G. Lisberger, and A.L. Boxer. Oculomotor function in frontotemporal lobar degeneration, related disorders and Alzheimer's Disease. *Brain*, 131(5):1268–1281, 2008.

[24] D.W. Hansen and Qiang J. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):478–500, 2010.

[25] H.W. Heuer, J.B. Mirsky, E. L. Kong, B.C. Dickerson, B.L. Miller, J.H. Kramer, and A.L. Boxer. Antisaccade task reflects cortical involvement in mild cognitive impairment. *Neurology*, 81(14):1235–1243, 2013.

[26] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer. Estimating oculomotor events from raw data samples. In *Eye Tracking: A Comprehensive Guide to Methods and Measures*, chapter 5, pages 147–185. Oxford University Press, Oxford, 2011.

[27] 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*, chapter 2, pages 9–64. Oxford University Press, Oxford, 2011.

[28] S. Hoops, S. Nazem, A.D. Siderowf, J.E. Duda, S.X. Xie, M.B. Stern, and D. Weintraub. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson Disease. *Neurology*, 73(21):1738–1745, 2009.

[29] Q. Huang, A. Veeraraghavan, and A. Sabharwal. Tabletgaze: Unconstrained appearance-based gaze estimation in mobile tablets. *arXiv:1508.01244 [cs]*, 2015.

[30] P. Inchingolo and M. Spanio. On the identification and analysis of saccadic eye movements - a quantitative study of the processing procedures. *IEEE Transactions on Biomedical Engineering*, 32(9):683–695, 1985.

[31] T. Karantinos, E. Tsoukas, A. Mantas, E. Kattoulas, N.C. Stefanis, I. Evdokimidis, and N. Smyrnis. Increased intra-subject reaction time variability in the volitional control of movement in schizophrenia. *Psychiatry Research*, 215(1):26–32, 2014.

[32] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba. Eye tracking for everyone. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2176–2184, 2016.

[33] H.-Y. Lai, G. Saavedra-Peña, C.G. Sodini, T. Heldt, and V. Sze. Enabling saccade latency measurements with consumer-grade cameras. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2018.

[34] R.J. Leigh and C. Kennard. Using saccades as a research tool in the clinical neurosciences. *Brain*, 127(3):460–477, 2004.

[35] R.J. Leigh and D.S. Lee. The saccadic system. In *The Neurology of Eye Movements*, chapter 4, pages 170–251. Oxford University Press, Oxford, 2015.

[36] R.J. Leigh and D.S. Lee. A survey of eye movements. In *The Neurology of Eye Movements*, chapter 1, pages 1–16. Oxford University Press, Oxford, 2015.

[37] D. Li, D. Winfield, and D.J. Parkhurst. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 79–87, 2005.

[38] J.K.H. Lim, Q.-X. Li, Z. He, A.J. Vingrys, V.H.Y. Wong, N. Currier, J.B. Mullen, B.V. Bui, and C.T.O. Nguyen. The eye as a biomarker for Alzheimer's Disease. *Frontiers in Neuroscience*, 10(536):1–14, 2016.

[39] D.J. Mack, S. Belfanti, and U. Schwarz. The effect of sampling rate and low-pass filters on saccades - A modeling approach. *Behavior Research Methods*, 49(6):2146–2162, 2017.

[40] 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*, 128(6):1267–1276, 2005.

[41] T. Moyniham. A slo-mo camera with an insanely high frame rate (and price tag). https://www.wired.com/2014/07/phantom-v2511-camera/. Accessed: 2018-08-16.

[42] D.P. Munoz. Commentary: Saccadic eye movements: overview of neural circuitry. In *Progress in Brain Research*, volume 140, pages 89–96. Elsevier, 2002.

[43] D.P. Munoz, J.R. Broughton, J.E. Goldring, and I.T. Armstrong. Age-related performance of human subjects on saccadic eye movement tasks. *Experimental Brain Research*, 121(4):391–400, 1998.

[44] D.P. Munoz, M.C. Dorris, M. Par, and S. Everling. On your mark, get set: Brainstem circuitry underlying saccadic initiation. *Canadian Journal of Physiology and Pharmacology*, 78(11):934–944, 2000.

[45] I. Noorani and R.H.S. Carpenter. The later model of reaction time and decision. *Neuroscience & Biobehavioral Reviews*, 64:229–251, 2016.

[46] M. Nyström and K. Holmqvist. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, 42(1):188–204, 2010.

[47] 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. *Neurobiology of Disease*, 43(1):79–85, 2011.

[48] C. Pierrot-Deseilligny, I. Israël, A. Berthoz, S. Rivaud, and B. Gaymard. Role of the different frontal lobe areas in the control of the horizontal component of memory-guided saccades in man. *Experimental Brain Research*, 95(1):166–171, 1993.

[49] 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*, 126(6):1460–1473, 2003.

[50] C. Pierrot-Deseilligny, C.J. Ploner, R.M. Müri, B. Gaymard, and S. Rivaud-Pechoux. Effects of cortical lesions on saccadic eye movements in humans. *Annals of the New York Academy of Sciences*, 956(1):216–229, 2002.

[51] S. Rivaud, R.M. Müri, B. Gaymard, A.I. Vermersch, and C. Pierrot-Deseilligny. Eye movement disorders after frontal eye field lesions in humans. *Experimental Brain Research*, 102(1):110–120, 1994.

[52] G. Saavedra-Peña, H.-Y. Lai, V. Sze, and T. Heldt. Determination of saccade latency distributions using video recordings from consumer-grade devices. In *Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC)*, 2018.

[53] M.G. Saslow. Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Journal of the Optical Society of America*, 57(8):1024–1029, 1967.

[54] R. Shafiq-Antonacci, P. Maruff, C. Masters, and J. Currie. Spectrum of saccade system function in Alzheimer's Disease. *Archives of Neurology*, 60(9):1275–1278, 2003.

[55] J.B.J. Smeets and I.T.C. Hooge. Nature of variability in saccades. *Journal of Neurophysiology*, 90(1):12–20, 2003.

[56] C. Theleritis, I. Evdokimidis, and N. Smyrnis. Variability in the decision process leading to saccades: A specific marker for schizophrenia? *Psychophysiology*, 51(4):327–336, 2014.

[57] R. Ulrich and J. Miller. Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology*, 37(4):513 – 525, 1993.

[58] P. Viola and M. Jones. Robust real-time object detection. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid= E1A65DBACAD63547A793A006413752E8?doi=10.1.1.110.4868&rep=rep1& type=pdf. Accessed: 2018-08-29.

[59] D.E. Warren, M.J. Thurtell, J.N. Carroll, and M. Wall. Perimetric evaluation of saccadic latency, saccadic accuracy, and visual threshold for peripheral visual stimuli in young compared with older adults. *Investigative Ophthalmology & Visual Science*, 54(8):5778–5787, 2013.

[60] S.J. Wilson, P. Glue, D. Ball, and D.J. Nutt. Saccadic eye movement parameters in normal subjects. *Electroencephalography and Clinical Neurophysiology*, 86(1):69–74, 1993.

[61] A.M.F. Wong. The saccadic system. In *Eye Movement Disorders*, chapter 4, pages 60–71. Oxford University Press, Oxford, 2008.

[62] E. Wood and A. Bulling. Eyetab: Model-based gaze estimation on unmodified tablet computers. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA)*, pages 207–210, 2014.

[63] P. Xu, K.A. Ehinger, Y. Zhang, A. Finkelstein, S.R. Kulkarni, and J. Xiao. Turkergaze: Crowdsourcing saliency with webcam based eye tracking. *arXiv:1504.06755 [cs]*, 2015.

[64] Q. Yang, Z. Kapoula, E. Debay, O. Coubard, C. Orssaud, and M. Samson. Prolongation of latency of horizontal saccades in elderly is distance and task specific. *Vision Research*, 46(5):751–759, 2006.

[65] Q. Yang, T. Wang, N. Su, S. Xiao, and Z. Kapoula. Specific saccade deficits in patients with Alzheimer's Disease at mild to moderate stage and in patients with amnestic mild cognitive impairment. *Age*, 35(4):1287–1298, 2013.

[66] Z. Zhu and Q. Ji. Novel eye gaze tracking techniques under natural head movement. *IEEE Transactions on Biomedical Engineering*, 54(12):2246–2260, 2007.

[67] Z. Zhu and Q. Ji. Eye gaze tracking under natural head movements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 918–923, 2005.

[68] S.M. Zola, C.M. Manzanares, P. Clopton, J.J. Lah, and A.I. Levey. A behavioral task predicts conversion to mild cognitive impairment and Alzheimer's Disease. *American Journal of Alzheimer's Disease & Other Dementias*, 28(2):179–184, 2013.