Chapter 8
Eye Tracking Methodology

S. B. Hutton

Contents

Abstract This chapter has two main aims. The first is to introduce readers to the range of eye tracking technologies currently available, and describe the basic principles on which they operate. The second is to provide readers with an understanding of the main determinants of eye tracking data quality and the ways in which this can be quantified. A greater understanding of how eye tracking technology works, and the key determinants of data quality has two important benefits. Firstly, it will improve the likelihood of researchers being able to maximise the quality of the eye tracking data they generate themselves, using eye tracking technology that is appropriate for their research goals. Secondly it will increase their ability to critically evaluate eye tracking research produced by other researchers. Holmqvist et al. (2011) identify several distinct categories of eye tracker users, including usability and media consultants as well as those interested in human-computer interaction and gaze controlled interfaces. This chapter assumes that the majority of readers are academic researchers, probably working in the fields of psychology or cognitive neuroscience and related

S. B. Hutton (✉)
Sussex University, Brighton, UK
e-mail: s.hutton@sussex.ac.uk

C. Klein and U. Ettinger (eds.), *Eye Movement Research*,
Studies in Neuroscience, Psychology and Behavioral Economics,
https://doi.org/10.1007/978-3-030-20085-5_8

disciplines, and as such are most likely interested in using eye tracking technology to establish point of gaze and oculomotor dynamics—and also be concerned with issues of accuracy, precision, sampling rate and timing. Section 8.2 of this chapter concerns eye tracking technology, and starts with a brief historical overview of early eye tracking techniques, followed by a description of some of the less common technologies that can still be found in research published today—albeit often in relatively niche areas. The vast majority of commercial eye trackers that are currently available are video based—and as such this approach is covered in most detail. Video-based eye tracking methodologies can be divided (broadly) into two categories: stationary, screen-based systems and mobile head-mounted systems. Clearly these two types of equipment are generally used in very different research scenarios—and differences in the technology concerned make comparisons across these categories difficult if not meaningless. As such they are treated separately in the chapter. Section 8.2 may provide some useful information for readers who are considering purchasing an eye tracker—but it is important to note that this is not its primary purpose—nor is it meant as an exhaustive description of the pros and cons of all currently available eye tracking techniques and commercial models for all possible research scenarios. Not only would such an undertaking become rapidly outdated, it would also involve comparing apples with oranges. Indeed, care has been taken to avoid mentioning specific manufacturers or models where possible. Hopefully any readers interested in purchasing an eye tracker will, after reading this chapter, be equipped with sufficient knowledge to make informed decisions as to which type of eye tracker would be most appropriate given their research goals—and be able to critically evaluate performance claims made by manufacturers and ask the right questions of sales people. Those already in possession of an eye tracker may gain a better understanding of how it works, and its capabilities and limitations, and be more confident that they are using it to its full potential. Section 8.3 considers eye tracking software—not only is software a central component of most commercially available eye trackers, and an important determinant of data quality, it is also one of the main factors determining the ease with which the technology can be used, and the range of uses to which it can be put. Again, this entire section is intentionally generic, and is not intended as an exhaustive evaluation of all currently available software. Section 8.4 addresses the other key topic of the chapter—data quality. It starts with an attempt to define important key terms such as "accuracy" and "precision". The section then considers how such concepts might be quantified. The final part of this section is intended to offer practical advice for maximising data quality—including some basic information on the importance of setting up participants and getting a good calibration. The precise setup and calibration details will differ depending on the type of eye tracker used—so the advice contained in this section again intentionally generic and aims to outline basic principles and good practices that apply to all or most eye tracking scenarios.

8.1 Introduction and Learning Objectives

The number of papers published that use eye tracking technology is increasing every year. Researchers who are interested in using gaze information to test hypotheses no longer have to build their own equipment from scratch, and have a wide range of techniques and commercial systems available to them. Many eye trackers are far simpler to operate and less uncomfortable for participants than the systems available even one or two decades ago, and some are capable of tracking eye movements with exceptionally high levels of accuracy and precision. Finally, advances in software mean that the analysis of oculomotor data is a far more efficient and rapid process than before. Current video-based eye tracking systems vary greatly in cost, and differ enormously with respect to their capabilities. Researchers need to make sure that any system they use (and the software that accompanies it) is capable of delivering data that will allow them to draw meaningful conclusions given their research goals.

The chapter has two key learning objectives: (1) To understand how various eye tracking techniques work (in particular the principles underlying video-based eye tracking). (2) To appreciate the various determinants of data quality, and understand how data quality can be measured and improved.

8.2 Eye Tracking Techniques

8.2.1 Historical Approaches

Wade and Tatler (2005) provide a fascinating history of the origins of eye tracking methodology in their book "The Moving Tablet of the Eye". Anyone interested in finding out more about early eye tracking pioneers and the techniques they developed should read this book. Techniques for tracking eye movements were first developed in the late 19th century. Pioneers such as Huey (1898) and Delbarre (1898) attached a crude "contact lens" to the anaesthetised eye of their participants. In Huey's apparatus, a lightweight aluminium rod attached to the contact lens connected to a pivot wheel, to which a second rod was attached. The movements of this second rod marked out the eye's movements on a rotating drum containing "smoke paper". Ingeniously, the second rod was electrified, creating sparks from its tip to the drum, allowing precise measurement of the eyes' speed. In Huey's words:

> *"In order to measure the speed [of the eye] an electric current from an induction coil was passed through the pointer to the drum. This current was interrupted at very regular short intervals by the vibrations of an electrically driven tuning-fork, the snap of the spark from the pointer's tip thus displacing a dot of soot on the paper record at each interruption. As the pointer flitted over the drum during the reading, a tracing was thus produced like that shown in Fig. 8.1".*

Fixations appear as "blobs" on the resulting trace—where many sparks have been emitted at roughly the same location, and saccades as a sequence of more widely

Fig. 8.1 The equipment used by Huey (left) and resulting recording (right)

spaced dots. By counting the dots between the blobs it is possible to measure saccade duration exceptionally accurately (a spark, and hence dot on the paper, was made every 10 ms by the "electronic tuning fork").

In the 20th century several alternative eye tracking techniques were developed—some involving more sophisticated contact lens arrangements in which rods were replaced with tiny mirrors, such as were used by Yarbus (1967). The reflection of a light source on these mirrors could be recorded on film and used to recover point of gaze. These approaches would still no doubt have been very uncomfortable for the participant—and the comparatively heavy contact lenses will certainly have altered the eye's dynamics to some extent.

A less invasive technique was pioneered by Dodge (see Dodge & Cline, 1901; Diefendorf & Dodge, 1908). His "Photochronograph" used a slowly falling photographic plate to record the reflection of a vertical strip of light on the cornea. In many ways this device was the precursor of modern video based oculography—the concept of measuring a reflection on the eye is employed by most of the video based eye trackers which currently dominate the commercial market.

Several other technical approaches were developed, mainly during the latter half of the 20th century. As many of these techniques (such as EOG) are still in use today, albeit with updated technology, they are described briefly in the following section.

8.2.2 Current Techniques

Currently available commercial eye tracking systems make use of a wide range of technologies. As mentioned previously, by far the most common are video-based eye trackers and consequently the majority of this chapter will focus on this technique. There are, however, several alternative approaches still actively used in research. These are outlined in the sections below.

8.2.2.1 Electro-Oculography (EOG)

As illustrated in Fig. 8.2, the eye acts as a dipole—a rod magnet—being slightly more negatively charged at the retina compared to the cornea (Mowrer, Ruch, & Miller, 1936). The magnitude of this corneoretinal potential is in the range 0.4–1.0 mV, and is generally attributed to the greater metabolic activity in the retina. Electro-oculograms are recorded via pairs of electrodes which are typically placed above and below the eye or to its left and right. When the eye rotates (horizontally or vertically) a potential difference occurs between the electrode it rotates towards and the one it rotates away from. This potential difference is measured and can be plotted and interpreted as gaze position on the horizontal or vertical axis (Fig. 8.3).

EOG has the advantage of being relatively cheap, and the technology is well established and readily available. The temporal resolution is high (the analogue signal can be sampled as often as desired). With some minor adaptations such as special electrodes the technique can also be used in neuroimaging environments such as MEG/MRI. Finally, EOG is capable of measuring eye movements whilst the eye is closed—essential if studying the eye movements that occur during sleep.

The EOG signal is measured with respect to the head—meaning that the technique is appropriate in head-fixed scenarios or for researchers for whom eye rotation rather than point of gaze is of primary importance. It remains a reasonably common technique in hospital settings, and is sometimes used by EEG researchers, who place

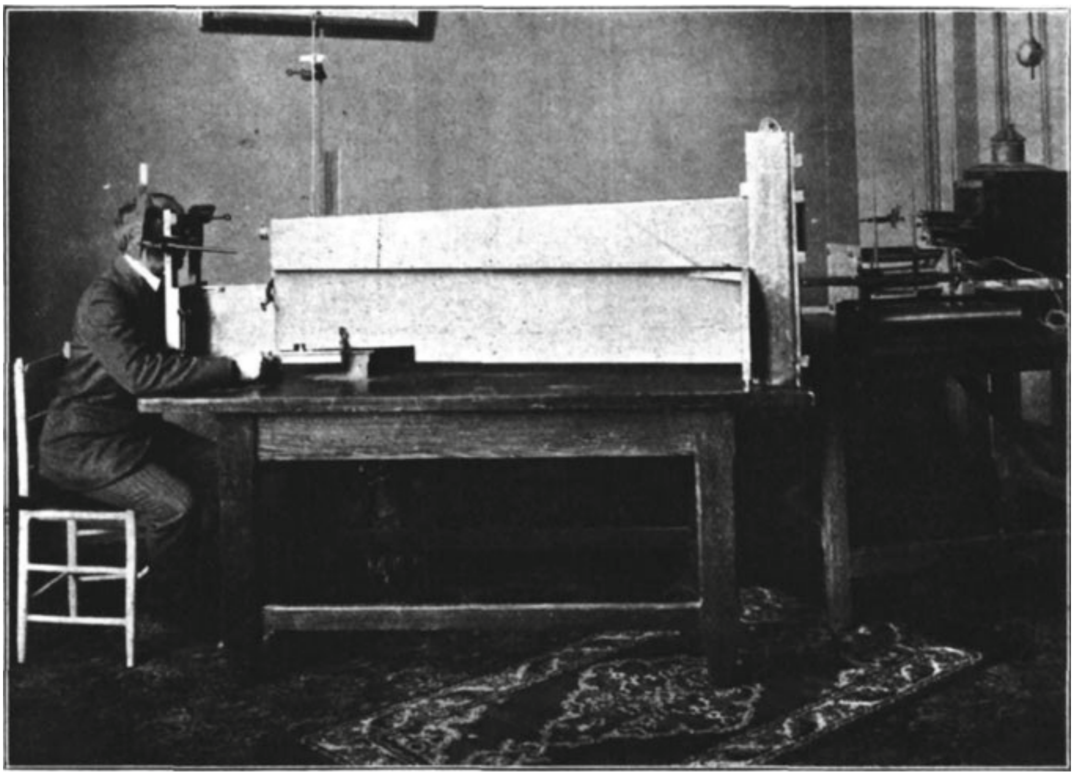


Fig. 8.2 Diefendorf and Dodge's Photochronograph

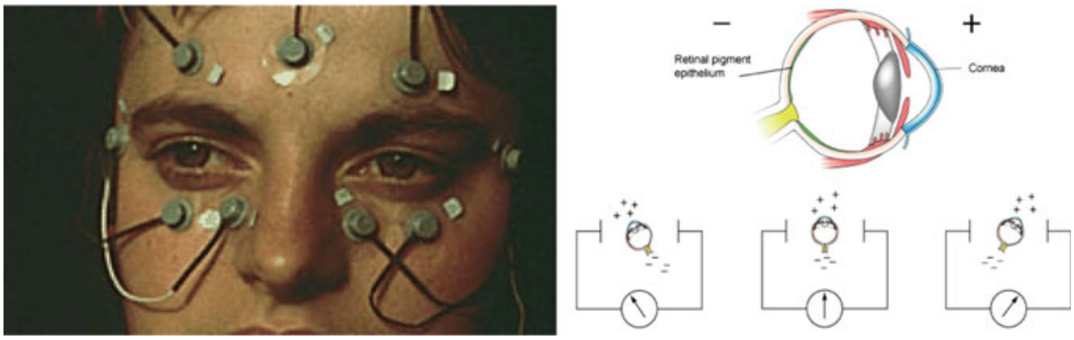


Fig. 8.3 Example of EOG electrode placement (left) and illustration of basic principles (right)

additional electrodes close to the eyes in order to pick up blinks or saccades which may confound their EEG data.

The main disadvantage of EOG is spatial accuracy because EOG is very prone to drift artefacts over time, typically due to impedance changes at one or more of the electrodes. These changes often result from sweating and mean that recalibration is frequently required. In addition, contraction of facial/neck muscles can also influence the signal, adding noise to the data. Spatial accuracy in EOG may also be compromised by the fact that the polarity of the eye is to some extent dependent on luminance changes, with increases in luminance increasing the level of signal. Finally, the signal is only truly linear for relatively small (less than about 10°) eye movements.

8.2.2.2 Limbus Reflection

Until video-based eye tracking techniques became dominant, infrared limbus trackers were fairly common. The basic principle is straightforward, and the components are relatively cheap. In essence a source of (typically infrared) light is shone at the eye, and a sensor (or array of sensors) measures the amount reflected back (see Fig. 8.4). The technique takes advantage of the limbus—the border of the coloured iris and predominantly white sclera of the eye. In essence the white sclera reflects more light than the coloured iris, and as the eye rotates more or less sclera is exposed to the IR emitters, and more or less IR is reflected onto the sensors.

Limbus trackers have a number of positive features. They are comparatively cheap and provide a direct measure of eye rotation within the head, which is important for researchers looking at basic oculomotor dynamics. As with EOG they are essentially analogue devices, and the analogue signal can be sampled at high frequency, allowing saccade dynamics to be observed in detail. However, as with DPI trackers, they require effective head restraint if point of gaze is required. If gaze shifts from a central target to a target 8° to the right, but the participants head simultaneously rotates 1° in the direction of the target, then Limbus trackers will only report a movement of 7°. One limbus based eye tracker has an ingenious solution to this issue. The system is head-mounted and the saccade targets are presented by what

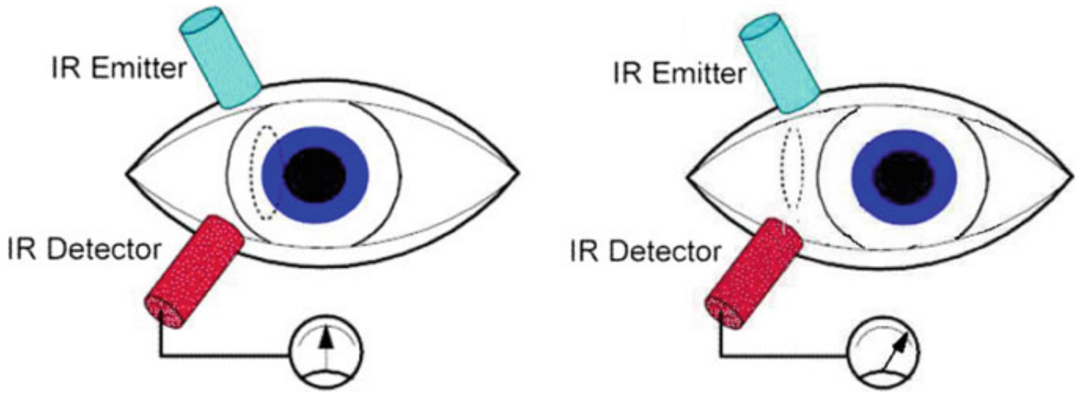


Fig. 8.4 The basic principle of limbus eye trackers. More IR is reflected in the image on the right because the limbus has moved away from the sensor as the eye rotates to its left

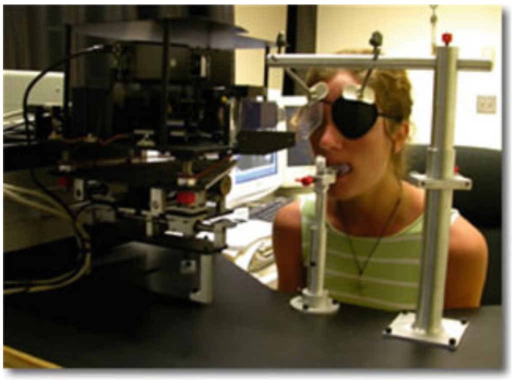
amounts to a laser pointer attached to the device. Participants can stand in front of any reasonably uniform surface such as an interior wall, and the targets are projected onto the wall. As such any head movement will result in the target also shifting, so the eye will always have to rotate the same amount as the target displacement in order for the target to be fixated.

Another limitation of limbus tracking devices is that due to the relationship between the emitters and sensors it is generally simplest to record either horizontal or vertical eye movements at any one time (although recording vertical eye movements in one eye and horizontal eye movements in the other is possible).

8.2.2.3 Dual Purkinje Trackers

Dual Purkinje Image (DPI) eye trackers provide high spatial accuracy. They are electro-mechanical devices and take advantage of the fact that a point of light shone at the eye results in four separate reflections—from the front and back of the cornea (the first and second Purkinje images) and from the front and back of the lens (the third and fourth Purkinje images). A detailed description of the mechanics of DPI systems is far beyond the scope of this chapter. The key principle to grasp is that rotational movements of the eye result in significant displacements of the locations of the first and fourth Purkinje images with respect to each other—whereas translational movements (such as those that arise from small head movements) result in no such displacement. The displacements between the first and fourth image occur during eye rotation because the two images are at different distances from the centre of rotation of the eye.

The two Purkinje images are separated by the optical components of the DPI eye tracker, and projected to photoreceptors via a complex series of lenses and mirrors, some of which are attached to servo-motors so that their angle can be rapidly adjusted. The signal from the photo-receptors is used to drive the motors so that the angle of the mirrors keeps the position of each Purkinje image "fixed" on its sensor. The



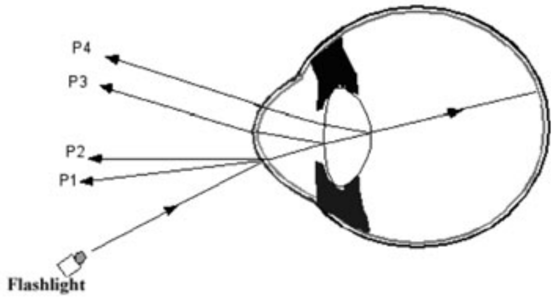


Fig. 8.5 A DPI eye tracker (left). Illustration of the 4 Purkinje Images (right)

"output" of the system is the difference in electrical signal sent to the various servo-motors in order to keep the two Purkinje images centred on their sensors. If both Purkinje images move the same amount (e.g. as would occur with a translational head movement) then the difference in the electrical signal sent to the two motors is zero—they both had to rotate their mirrors the same distance. When the eye rotates however, the two Purkinje images move a different distance—and hence the two motors have to move the two mirrors a different amount in order to stabilise their respective images. After a calibration procedure, the output of the system can be used to provide an exceptionally accurate estimate of eye position (Fig. 8.5).

DPI eye trackers are accurate up to 1 min of arc (a 60th of a degree of visual angle). As such the systems are perfectly capable of resolving very small micro-saccades. They were often used by psycholinguists, for whom spatial accuracy is of paramount importance (so that they can unambiguously determine which word, or even which letter within a word is being fixated).

On the downside, the nature of the optics means that whilst small head are tolerated, larger head movements (e.g. more than about 5 mm) can result in the eye becoming non-trackable. As a result, participants are often required to use bite-bars, which can quickly become uncomfortable.

Whilst DPI trackers provide a very accurate measure of eye position during fixations, they provide a slightly less accurate indication of eye position at the end of fast saccadic movements. This is because the fourth Purkinje image is reflected from the lens, which is suspended via muscles behind the cornea. At the end of a saccade, which is a very rapid movement that terminates very suddenly (see chapter by Pierce and colleagues in this volume), the lens "wobbles" slightly, causing small oscillatory changes in the position of the 4th Purkinje image, which are reported as a change in eye position by the tracker (see e.g. Deubel & Bridgeman, 1995). These "wobbles" are now referred to as "Post Saccadic Oscillations" (Hooge, Nyström, Cornelissen, & Holmqvist, 2015) and can cause ambiguity as to the precise timing of the onset of the subsequent fixation. As will be seen in Sect. 8.2.3, a related issue occurs with video based eye trackers which track the centre of the pupil.

Whilst exceptionally accurate (at least during fixations), DPI eye trackers are comparatively expensive—no doubt due to the complex optical and electromechanical components they involve. They also have a relatively limited operational range (in terms of how far the eye can rotate before one or other of the Purkinje images is lost), and this range can be very dependent on individual differences in pupil size as the fourth Purkinje image is seen through the pupil. Improvements in video based eye tracking systems—which offered similar levels of accuracy combined with a lower cost and greater ease of use and flexibility—resulted in many researchers moving away from DPI tracking, although there are still plenty of systems in active use.

8.2.2.4 Search Coils

Scleral search coils are essentially large contact lenses that contain a very thin filament wire wound round them (see Fig. 8.6). In humans the lenses are attached after the application of a topical anaesthetic. In non-human primates they may be surgically attached—removing any possibility of the lens slipping. The participant is placed within a magnetic field (created by larger wire coils, often housed within a "Necker Cube" type frame). When the coil in the contact lens moves, the magnetic field induces voltage change within the coil which can be directly measured. A figure of 8 winding in the lens allows torsional (rotational) eye movements to be measured, whilst a separate circular winding allows horizontal and vertical movements to be recorded.

The resulting voltage changes provide a very accurate measure of eye rotation within the head, and along with DPI trackers, scleral search coils systems are considered another "gold standard" in terms of eye tracking accuracy. There are, however, some disadvantages, not least the relatively invasive nature of the technique. Corneal scarring (albeit mild) can occur during lens insertion. The equipment is also relatively expensive. In addition, the lenses have been shown to slightly alter the kinematics of the eye (Frens & van der Geest, 2002), and even the smallest amount of slip can lead to inaccurate data.

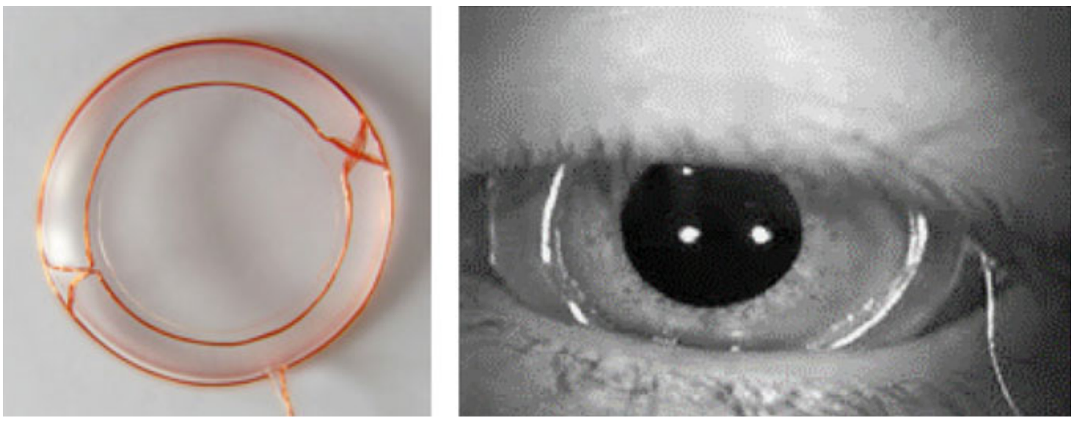


Fig. 8.6 Figure of 8 winding suitable for measuring torsional eye movements (left). Search coil in use (right)—note the trailing wires

An interesting variant of the scleral search coil technique is the dual induction system developed by Bour, van Gisbergen, Bruijns, and Ottes (1984). Rather than silicon lenses containing coiled wire, a gold annulus is placed on the eye. The distortions in the magnetic field induced by rotations of the eye are picked up by a secondary sensor placed in front of the eye. This technique avoids the wires trailing out from the lens—but is still comparatively invasive, and is not able to track torsional eye movements.

As with DPI trackers, the comparative ease of use and non-invasive nature of video based eye tracking has meant that scleral search coil techniques are used less often in human populations than they used to be—although it remains a relatively common technique for tracking the eyes of non-human primates. A recent publication by Kimmel, Mammo, and Newsome (2012) described a simultaneous comparison of search coil and video based eye tracking. They found very close agreement between the positional data from both systems and concluded that the video based system was appropriate for many if not most applications for which search coils have been used.

8.2.3 Video Based Eye Trackers

8.2.3.1 Basic Principles

Video based eye trackers are in some respects the modern descendants of Dodge's photochronograph in that they rely on (video) photography. They may also take advantage of a principle employed by DPI eye trackers to compensate for small head movements, that of two landmarks on the eye moving differently with respect to each other during rotation, but similarly during translation.

Rather than track two Purkinje images, most video based eye trackers track the first Purkinje image and/or the centre of the pupil (see Fig. 8.7).

The sampling rate of video based eye trackers is a function of the camera speed, that is the number of images it can capture per second. Commercial systems operate at anything from 30 to 2000 Hz. As is discussed in Sect. 8.4.1.3, sampling rate is a potentially important determinant of data quality. Only high speed systems are capable of delivering the high levels of spatial and temporal resolution previously associated with DPI and search coil techniques.

Video based eye tracking relies on image processing software that is able to rapidly identify, in each image, the precise location of the necessary landmarks (e.g. pupil and corneal reflection). When identifying the centre of the pupil two approaches can be taken. The most common is known as "dark pupil" tracking, and takes advantage of the fact that under infrared illumination, the pupil appears as a comparatively dark circle compared to the rest of the eye (e.g. the iris and sclera). The alternative approach is "bright pupil" tracking—which takes advantage of the well-known "red eye" effect—when light is shone directly through the pupil it bounces off the back of the retina. Most video based eye trackers rely on an infrared light source to illuminate

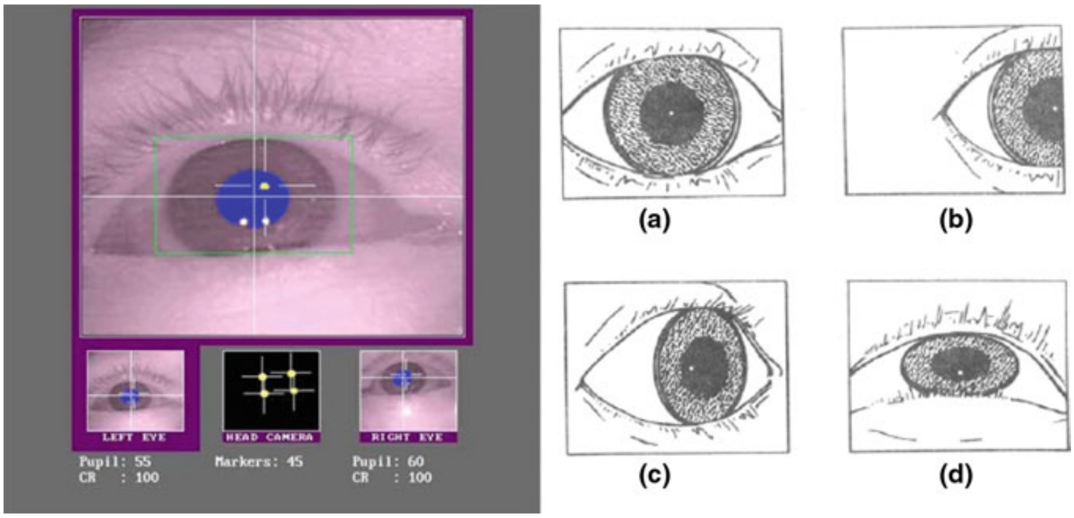


Fig. 8.7 Video based eye tracker using the centre of the pupil and first corneal reflection to track eye position (left). Illustration of the change in position of pupil and corneal reflection as seen by camera pointed straight at the eye (**a**) for rotational (**c**, **d**) and translational (**b**) movements. (Right, From Young & Sheena, 1975)

the eye and provide the corneal reflection. Systems that use infrared illumination and cameras are relatively untroubled by variations in natural light levels.

Complex image processing algorithms identify the centre of landmarks such as the pupil and corneal reflection, usually in real time, and after a calibration procedure has been performed the relative locations of the landmarks can be used to determine point of gaze. The precise details of these algorithms are generally proprietary, and a detailed discussion of the various approach is beyond the scope of this chapter.

As with DPI trackers, small head movements (which are translational) are differentiated from actual eye rotations by the different effect these movements have on the relative location of the pupil and CR centre on the camera's sensor—head movements result in both the CR and pupil shifting to the same extent, whereas eye rotations result in the pupil and CR moving a differing amount.

8.2.3.2 Monocular Versus Binocular Systems

Depending on the make and model, video based eye trackers may either track one eye or both eyes (or give the user the choice of tracking monocularly or binocularly). Most head mounted eye trackers are binocular. One advantage of binocular eye tracking is that vergence movements can be quantified and therefore some estimate of depth of focus can be established. This is particularly important for head mounted systems as participants are free to focus on objects at different distances, whereas for desktop systems the monitor is kept at a fixed distance. In general our eye movements are conjugate. In other words, both eyes move in the same direction at the same time, and the majority of eye tracking studies in the fields of Psychology and Cognitive Neuroscience report monocular data. However, when we change our focal depth eye movements can be disconjugate—looking at something straight ahead that is very close requires both eyes to rotate nasally compared to when looking at an object in the distance. Researchers interested in depth perception therefore require a binocular eye tracker. If presenting stimuli using a 3D monitor, or stereoscope, it is important to check that the eye tracker is still able to operate. For example, some 3D shutter glass systems require a strobing infrared light to synchronise the glasses with the monitor—and this could interfere with some video based eye trackers.

Another situation in which binocular data is useful is for detecting micro-saccades (see chapter by Alexander and Martinez-Conde in this volume). These saccades can be very close to the minimum spatial resolution of eye trackers, and therefore ambiguity as to whether a detected event is or is not a micro-saccade is reduced if the same movement occurred, in the same direction, in both eyes simultaneously. One of the most widely used algorithms for detecting micro-saccades (Engbert & Kliegl, 2003) can make use of binocular data.

8.2.3.3 Stationary Versus Mobile Systems

As mentioned in the introduction, video based eye trackers can be broadly categorised into stationary vs mobile systems. Stationary systems generally track eye movements whilst participants view stimuli presented (typically on a computer monitor) in front of them. These systems vary in the extent that head movements can be tolerated, requiring either some method to fix the head's position with a chin/forehead rest for instance, or allowing head movements within a limited "head box". Mobile systems allow participants to move around freely and look at objects other than a computer monitor, with the eye tracking components worn as glasses or a lightweight head-set. Stationary systems are typically used in research scenarios in which stimuli such as saccade targets, images or videos are presented to participants on a computer monitor. Mobile systems are generally more suitable for more "real world" type tasks in which participants are free to interact with people and objects around them as they go about a specific task (e.g. playing a sport/shopping/making a cup of tea).

Due to their need for small and lightweight components, most mobile eye trackers tend to have lower sampling rates and poorer spatial resolution than stationary systems (see next two sections).

8.2.3.4 Stationary Video Eye Trackers

There are a number of commercially available stationary systems, with large variations in cost, functionality and performance, but all operate on the same basic principles outlined above. Most are used in standard "laboratory" settings, in which a participant sits at a table in front of a monitor, but some stationary systems can also be used in neuroimaging environments. Their distance from the head means that they do not interfere with EEG electrodes for example. Some systems can operate in fMRI settings, with an MR compatible camera and illuminator being placed in a position where they can see the eyes either directly, or via a mirror above the head-coil.

Stationary systems can be divided further into two subtypes—remote eye trackers (that typically sit on a desktop) and tower mounted systems (which may be free standing or clamped to the desktop). Tower mounted systems combine the eye tracker with a head-restraint solution, and allow the camera to be placed closer to the participant's eye. In many cases the camera is mounted above the eye, looking down to a reflection of the eyes in a "hot mirror"—a mirror which is transparent to visible light (so that the participant can look through it) but which reflects infrared light. By placing the camera above the eye, out of the field of view, tower mounted systems have the advantage that the participant can interact with the stimuli being presented (e.g. by pressing on a touch screen) without interfering with the view from the camera or the IR light source. Tower mounted systems also often have a much larger trackable range.

Some systems require the participants head to remain relatively stationary, with some kind of chin rest or other head-restraint. Other "head free" systems allow greater freedom of movement, and some systems employ distinct algorithms for head fixed versus head free eye tracking. In general, greater accuracy can be obtained during head-fixed recordings. Research goals should determine which approach is most appropriate for any given project. Some populations (for example young children, infants or some patient groups) may not be able to use head-supports. Similarly, if the research goal is to explore eye movements during a relatively natural interaction with a computer (as in certain usability or HCI settings) then head-restraint is again undesirable. If head free eye tracking is required, it is important to consider the size of the "head box", that is the area in which the head is free to move whilst eye tracking is maintained.

One potential limitation of high speed video-based eye trackers that rely on the pupil centre to compute the point of gaze, is that the pupil is continually changing size. There are two ways in which this can be problematic. The first concerns dilations and constrictions—these are typically driven by luminance changes (including local luminance changes in the image being viewed), but can also be elicited by factors such as cognitive work load, arousal or fatigue (see chapter on pupillometry by Alnaes and Lang in this volume). It might be tempting to assume that as the centre of a large circle is in the same location as the centre of a small circle, pupil size should not influence the reported location of gaze. Unfortunately, pupils do not dilate or constrict symmetrically around their centre (see, e.g., Wildenman & Schaeffel, 2013). This means that large luminance changes both within and between trials can potentially cause slight spatial inaccuracy in the reported gaze location. The second potential problem that tracking the centre of the pupil can cause is post-saccadic oscillations. As mentioned previously, these are an issue in DPI trackers because the fourth Purkinje image is reflected from the back of the lens, which can wobble at the end of saccades. The wobbling of the lens may possibly influence the shape of the pupil, and thus its computed center. Video eye trackers with low sampling rates do not suffer from this issue as they lack the temporal (and often spatial) resolution necessary to resolve these tiny changes in the location of the pupil's centre. Nyström, Andersson, Magnusson, Pansell, and Hooge (2015a), Nyström, Hooge, and Holmqvist (2015b) provide a comprehensive discussion of the issue.

8.2.3.5 Mobile Video Eye Trackers

Some mobile eye trackers are worn as glasses—with the critical eye tracking components housed within the frame. Other mobile systems have the components attached, one way or another, to some kind of head-set—this may be as simple as a modified spectacle frame on which is mounted one or more miniature cameras and possibly a mirror, or a more complex headset carrying more or larger components. As with stationary video based eye trackers, the key components of mobile systems are cameras, capable of viewing the eyes, and one or more infrared light sources capable of illuminating the eyes and providing a corneal reflection. Due to the obvious need for key components (e.g., cameras) to be lightweight, head mounted eye trackers tend to have lower specifications than desktop systems.

A key feature of most mobile eye trackers is that in addition to the camera or cameras that are recording the eye, they contain a "scene camera", that is a camera that is capable of recording the scene the participant is looking at, capturing the participants view of the world, and over which a gaze cursor can be plotted both in real time and any subsequent analysis stage. This feature is necessary for any use in which the person being eye tracked is interacting with the real world, rather than simply observing stimuli presented on a computer screen.

Data quality is more likely to be an issue for mobile eye trackers than stationary eye trackers. This may be due in part to the relatively low specifications of the cameras and other components used (in order that the systems remain wearable), but perhaps more important are the challenging scenarios in which mobile trackers tend to be used, for example driving or sport. Such scenarios clearly limit the control the experimenter has over potentially important variables. For example, bright sunlight can be particularly problematic for mobile eye trackers. Not only can it "bleach out" the infrared illumination systems, it can also induce a dramatic reduction in pupil size, and even squinting behaviour—making the eye difficult or impossible to track.

8.3 Software Used in Eye Tracking Research

A critical component of modern eye tracking methodology is software, which can be divided into three broad categories: Firstly, there is the software used by the eye tracking system itself. At a minimum this software provides a user interface and controls the actual recording of the eye movement data. Secondly, there is software for stimulus presentation. This is a requirement for most, but by no means all eye tracking research. Finally, there is software that facilitates the analysis of eye tracking data.

The aim of this brief section is not to provide an exhaustive list of the capabilities of what are generally rapidly evolving pieces of software. Nor is it meant to be a comprehensive comparison between various approaches and solutions. Its central purpose is to make readers aware that when evaluating published eye tracking research, or when considering which type or model of eye tracker would be most appropriate for their needs, they should not limit their evaluation to hardware considerations. People looking to buy an eye tracking system should look in detail at the capabilities of the software itself: Does it provide the functionality you need? How it is licensed? Will future upgrades incur a cost? People evaluating research should consider the extent to which critical factors (such as how fixations and saccades are defined) are controlled by the eye tracking software.

An informed understanding the eye tracking software—what it is doing and how it is doing it—can also lead to improvements in participant set up and ultimately data quality. By understanding what your equipment's software is trying to do, you are better placed to know when (and why) it might be struggling—and what you can do to help.

8.3.1 Software Controlling the Eye Tracker

One of the key functions of eye tracking software is to provide the user with an interface through which participant set up and calibration can be controlled. Important points to consider are its ease of use (particularly if working with populations such as infants or patients) and its flexibility (can different calibration targets be used for example?). The software will also save the eye movement data to a file.

At the heart of all video based eye tracking software are image processing algorithms. Essentially the software receives a series of images of the eye(s) from the eye tracking camera(s), and needs to identify (at a minimum) the location within that image of the centre of the pupil and the corneal reflection(s). Manufacturers differ in the extent to which the various algorithms (filters, parsing rules etc.) used by the eye tracking software are revealed to customers, and in the extent to which these can be modified, but these low level algorithms are often proprietary, and may not made available to researchers. Their robustness is one of the key determinants of eye tracking data quality, and as is argued in Sect. 8.4, setting up participants optimally involves ensuring that the algorithms are able to do their job.

Eye tracking software also often handles communication with the presentation software. Such communication is important as it allows stimulus events (such as the onset of a target or image, or the offset of a fixation cross) that are controlled by the presentation software to be logged in the eye tracking data. Without such communication calculating saccade latencies, and when fixations occur with respect to the stimulus onset can be difficult. Another important aspect of this communication is whether it is bidirectional. In other words, can the eye tracking software make the results of its gaze calculations available to the presentation software? Such communication is critical for gaze contingent tasks (in which for example the location of a saccade target is shifted during the saccade itself). If such bi-directional communication is required, then it is important to establish the end to end delay (that is, the delay between the eye changing position and that change being calculated by the eye tracking software and sent back to the stimulus presentation software; see Sect. 8.4).

8.3.2 Stimulus Presentation Software

There are two broad categories of stimulus presentation software—software supplied by the eye tracking manufacturer, and generic stimulus presentation software (which can be further divided into commercial software such as E-Prime, Matlab, NBS-Presentation, and free solutions such as OpenSesame or PsychoPy). When considering stimulus presentation software, it is important to evaluate their timing accuracy (their ability to present stimuli at a known point in time) and synchronisation abilities (the extent to which the time that a stimulus event occurs is accurately flagged in the eye tracking data stream).

Software provided by manufacturers ranges from basic products that essentially implement some kind of "slideshow" of text, images or possibly videos, to powerful and sophisticated software capable of implementing a very wide range of experimental scenarios. The main advantage of software provided by the eye tracking manufacturer is that integration between the stimulus presentation and recording software is generally handled very straightforwardly, for example if calibration routines and gaze contingent triggers are easily set up. Manufacturer supplied display software may also offer straightforward integration with any analysis software that they provide, for example so that interest areas defined in the experiment are available automatically at the analysis stage. Another advantage is that the manufacturers may be able to offer more support to customers using their software for stimulus presentation.

Commercial stimulus presentation software is often highly featured and well supported. Most modern eye trackers will provide libraries or some means by which commercial stimulus presentation packages can communicate and interface with the eye tracker—for example to start or stop recording and signal key stimulus events such as fixation onset or target onset etc. Commercial programs differ with respect to their timing capabilities, as well as their ability to accurately display stimuli of different types (e.g. audio or video). Users should make sure that they choose presentation software that is capable of implementing the eye tracking tasks they plan to run.

A final distinction that can be made is between presentation software that is essentially a programming environment (Matlab with Psychtoolbox or Presentation) and software that is uses a graphical interface (e.g. E-Prime or PsychoPy). Whilst programming environments are often very versatile, they can take a relatively long time to learn. Conversely, while software that uses a graphical approach can be very easy to work with, and allow functional experiments to be produced quickly, they may lack flexibility required for more complex experimental designs.

8.3.3 Data Analysis Software

Much like EEG and fMRI, eye tracking can very quickly result in very large amounts of raw data. A relatively modest experiment containing 24 trials in which eye movements were recorded for 8 s per trial, would result in 192,000 samples from a high speed eye tracker recording at 1000 Hz. Each sample would contain, at a minimum, a timestamp and the X and Y location of gaze (and often pupil area) at that point in time. Multiply the number of data points by the number of participants and the disadvantages of analysing the data without some kind of software solution are clear.

For researchers happy to deal in sample level data, programs capable of dealing with large quantities of data such as Matlab and R can be used. Sample level position data can readily be converted to velocity, allowing the data to be parsed into saccades and fixations using whatever criteria the researcher feels are appropriate. Such software also allows for flexibility in terms of applying filters should these be desired or be felt necessary due to data quality issues.

There are some commercial software solutions designed specifically for eye movement data analysis, that can work with data from a number of different eye trackers (either in its raw form, or converted to ASCII format). There are some free eye movement data analysis packages—but they are typically aimed at relatively discreet types of analysis. For example iMAP (Caldara & Meillet, 2011) is a powerful package of Matlab functions that allow sophisticated fixation maps (heatmaps) to be created and compared statistically. Most analysis software, however, is supplied by manufacturers, and will work only with data collected on their eye tracking systems.[1]

As with stimulus presentation software, analysis solutions provided by manufacturers may vary enormously in their sophistication. At the most basic level are programs that allow a single data file to be loaded, perform some basic parsing to allow fixations to be identified, and plot these over the stimuli that were presented. Such software may perform some other basic visualisations—scan paths for example (see chapter by Foulsham in this volume)—and calculate basic metrics such as fixation count and average fixation duration. More sophisticated software will allow data to be loaded from more than one participant, and will likely involve some integration with the stimulus presentation software. Often such software will also allow the creation of interest areas (also known as regions of interest) and be capable of outputting reports that provide a range of summary data at the interest area level (such as dwell time). The most sophisticated analysis software will be tightly integrated with both the recording and stimulus presentation software, and be able to present the data with a range of visualisations, including, for instance, the creation of heat maps etc. Other functions may include the ability to group data by trial variables, "clean" data, limit analyses to specific interest periods, create dynamic interest areas, and will be capable of outputting hundreds of useful dependent variables for analysis.

[1] A reasonably up to date list of software for analysis can be found here: http://www.eyemovementresearch.com/software/.

8.4 Data Quality

No matter what eye tracking technology is used, or what population is being tested, all researchers should be aiming to record the best quality data that they can. Good quality data makes subsequent analysis far simpler and any effects easier to describe and prove statistically. It is, however, important to bear in mind that different aspects of data quality may be more or less important to researchers depending on their research goals. For example, psycholinguists interested in which part of which word was fixated during reading, and exactly when and for how long it was fixated, will require exceptional spatial and temporal accuracy and precision. Researchers using a dot-probe like task, or a preferential looking task with infants, may only need to know whether the first eye movement after target presentation was to the left or to the right, in which case spatial accuracy and precision will be less important. Due to space limitations this following discussion should be considered introductory. Readers interested in a more detailed discussion on this topic can find much useful information in Reingold (2014) and Holquvist and Nystrom (2010).

As will become clear, some determinants of data quality are a function of the eye tracker itself such as its sampling rate, noise level, the filtering algorithms it employs. However, other determinants are, to some extent at least, within the control of the experimenter and the participant. The aim of this section is to ensure that researchers working with eye trackers have sufficient understanding of those aspects of data quality that they have some control over to be able to take positive steps to ensure they are getting the best data they can, given their system's potential.

8.4.1 Key Terms

Any attempt to discuss eye tracking data quality involves stepping into a terminological minefield. Terms such as "accuracy" and "precision" are often used by different researchers to mean rather different things. This section starts with an attempt to define various key terms (highlighted in bold), as they are used in this chapter—which, it should be born in mind, is not necessarily how they are always used in published research or manufacturer's literature.

As a starting point, I follow Reingold (2014) in defining **data quality** as "the fidelity with which the continuous variation in the eye movement signal is reflected in the values measured and reported by the eye tracker". In other words, we have something real that changes over time (the position of the eye) and we have our measurement of that position. In crude terms data quality can be considered as the extent to which the measurement agrees with reality, with high levels of agreement equating to high data quality. Any attempt to measure data quality, defined as such, obviously faces an immediate problem: When recording from an actual eye we have no independent estimate of its "real" eye position other than our measurement of it via the eye tracker. This point (and possible solutions) is returned to in Sect. 8.4.2

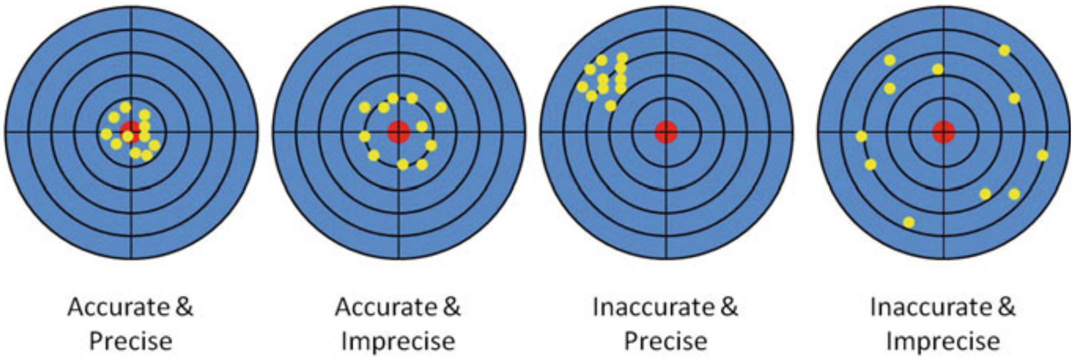


Fig. 8.8 illustration of the difference between, and independence of, accuracy and precision

on measuring data quality. The remainder of this section will attempt to provide definitions of several other important concepts.

8.4.1.1 Accuracy and Precision

Perhaps the two most important (and widely used) concepts in any discussion of eye tracking data quality are "accuracy" and "precision". These two concepts are, as Reingold (2014) points out, often treated (erroneously) by researchers as equivalent. If a single actual (true) value is repeatedly measured, accuracy can be defined as the mean difference between the measured and the true value. In contrast, precision (or reproducibility or repeatability), can be considered as the extent to which repeated measurements of a set of true values produces the same or similar set of measured values—regardless of the accuracy of these values. In other words the measured values may not be accurate, but if they are consistent then precision is high. The concept is illustrated in Fig. 8.8:

8.4.1.2 Noise

Another concept, which can be related to both accuracy and precision, is often referred to as "noise". This can be considered as variations (unwanted or of unknown source) in the gaze position signal, a signal which, at least during fixations, is often implicitly assumed by researchers to be stable. It is important to understand that there are several potential sources of noise in eye tracking data. One major source of noise is the human oculomotor system itself. The eye is never truly still; even during fixations movements such as drift, tremor and micro-saccades are readily observed. Such oculomotor "noise" is real and can be accurately measured, but it can be considered as noise in some research contexts because it adds variation to what many researchers assume is a steady signal (for instance, the eye during fixation). The second source of noise can be termed environmental, caused for example by vibrations disturbing the eye tracking camera, or electromagnetic interference. The final source of noise is the eye tracking equipment itself. When an artificial eye is tracked, and environmental factors are adequately controlled, the contribution to overall noise levels made by the eye tracker itself can be isolated and the "spatial resolution" (the smallest change in position that can be detected) of the system can be calculated. High end video based eye trackers have spatial resolutions of around 0.01–0.05° of visual angle.

8.4.1.3 Temporal Resolution, Latency and Delays

Our discussion of data quality has thus far focussed on the spatial dimension, but timing issues are often equally important. Perhaps the simplest concept to grasp is that of "temporal resolution", being essentially the sampling rate of the eye tracker, as typically expressed in Hz and corresponding to the number of times the eye tracker is capable of sampling the eye's position each second. As mentioned previously, video based eye trackers have sampling rates ranging from around 30 to 2000 Hz. Sampling rates are important for a variety of reasons, and high sampling rates in general confer benefits rather than disadvantages.

If fixation duration, or saccade onset latency are critical variables, then having an eye tracker that can establish these durations or timings to the nearest 1 or 2 ms (e.g. a 1000 or 500 Hz system), as opposed to the nearest 16.67 ms (as is the case for a 60 Hz eye tracker) is preferable. Holmqvist et al. (2011) point out that the effects of low sampling frequency can be compensated for by collecting sufficient data. But collecting sufficient data to overcome the limitation of a slow eye tracker can be a high price to pay if the expected difference in values (e.g. fixation durations) between conditions is small, and in many experimental settings is simply not an option.

High sampling frequencies are also needed if you are interested in detecting microsaccades, or accurately measuring saccade velocity or amplitude. Another advantage of relatively fast sampling rates is that it can reduce "recovery time". During eye tracking recordings it is not unusual for the eye tracker to periodically be unable to track the eye, with blinks being by far the most common reason. When the eyelid descends it obscures the pupil and corneal reflection and eye tracking is not possible. Head rotations that obscure the camera's view of the eye, or head movements that take the eye out of the camera's field of view can also result in "tracking loss". Eye trackers with high sampling frequencies are much quicker to "reacquire" the eye than trackers with low sampling frequencies.

Another important term is "end to end delay" which can be defined as the time between an actual movement of the eye taking place, and the eye tracking system signalling that the movement has occurred. Having a low latency eye tracker is particularly important for users who want or need to run gaze-contingent tasks in which some aspect of the stimulus display is dependent on the participants gaze position. Examples of gaze contingent tasks include the "moving window" type tasks popular in psycholinguistics, and saccade adaptation tasks, in which the target location is shifted during the saccade itself.

A related concept is "temporal precision", which can be defined as the standard deviation of eye tracking latencies. Poor temporal precision can occur if the computer running the eye tracking software allows the operating system to "hog" the processor, taking priority over the gaze calculations. Poor temporal precision can create huge difficulties for researchers, particularly when trying to synchronise their eye movement data with stimulus events. It can also contribute ambiguity to the onset and offset of fixations and saccades.

8.4.1.4 The Interaction Between Noise and Sampling Rate: Velocity Noise

The velocity of an object can be defined as its change in position per unit time. Eye velocity is typically measured in degrees of visual angle per second. Velocity can be calculated from position data, using a variety of models. A detailed treatment of these models is beyond the scope of this chapter, but in essence velocity can be thought of as the speed of the eye, and is therefore related to the difference in eye position between samples.

Velocity is often calculated, either during recording, or after data collection, and can be used for defining the onset and offset of saccades (and hence the onset and offset of fixations). Because velocity is based on the differences between positions, any positional noise will contribute to what is known as velocity noise. Reingold (2014) illustrates the critical relationship between sampling speed and velocity noise with the following diagram. It is clear that a moderate amount of spatial "noise" can have a dramatic impact on the number, duration and timings of fixation and saccade events. Holmqvist et al. (2012) provide another important demonstration and discussion of this issue.

In the top panel of Fig. 8.9, moderate noise has been added to two samples of positional data recorded at 200 Hz (indicated by asterisks). When the positional data is converted to velocity, the velocity noise causes both the false detection of a saccade, and errors in the duration and onset of two fixations.

8.4.1.5 Filtering

Modern eye trackers may employ some kind of filtering to attenuate variations in the eye tracker signal that are not related to the eye movement itself (e.g., noise). Filtering can be applied by the tracker software during the recording itself, and/or by the analysis software after the data has been recorded. Different eye trackers use different filtering techniques, typically either "heuristic" or rule of thumb filters, or some mathematical filtering process (e.g., a moving average filter). It is important to note, however, that certain filters may also distort the signal, for instance, by changing the latencies of certain events (e.g., saccades) or filtering out others (e.g., micro-saccades). It is therefore important to understand the impact of any filtering on your data.

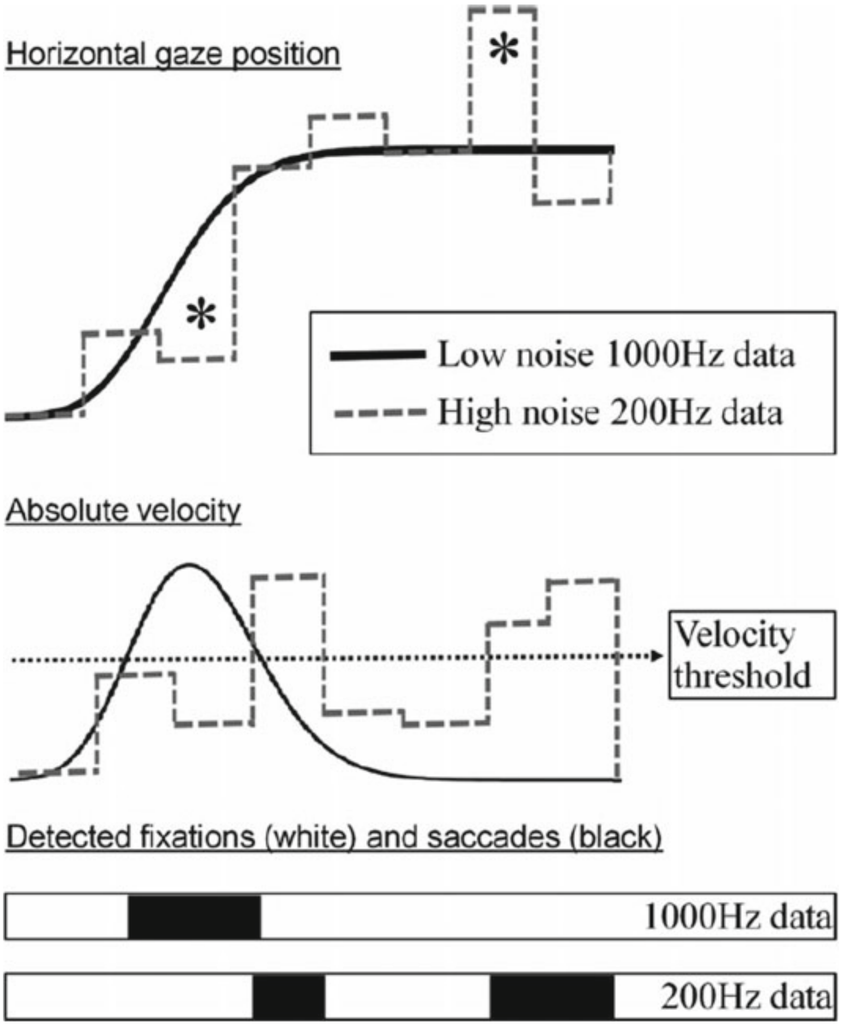


Fig. 8.9 Illustration of the effect of velocity noise on parsing. From Reingold (2014)

8.4.1.6 Parsing

Another issue when considering data quality concerns what exactly we are referring to by "data". Eye trackers output "sample level" information, in other words a single estimate of the eye's location (or point of gaze) for each measurement taken (e.g., 60 times per second for an eye tracker with a 60 Hz sampling rate or 1000 times per second for a 1000 Hz system). Whilst low level sample data is of enormous importance to many researchers, particularly those concerned with the basics of oculomotor control, saccade dynamics and curvature, smooth pursuit etc., it is of less importance to many others, at least in the sense that they are unlikely to want to deal with their data at the sample level. This is particularly the case for those researchers whose primary measures are based on fixations, including their location and duration. For example, an infant researcher may want to know which of two simultaneously presented faces was fixated first, and for longest. A psycholinguist may want to know which words were fixated, in which sequence and for how long.

Such analyses generally do not require looking at data at the sample level. They do, however, require that the sample data be parsed, which means that a set of rules is required to allocate any given sample to some event-state, typically fixation, saccade or blink. The properties of these events (e.g., the location of a fixation, its onset time, its duration, the amplitude of a saccade) are the key metrics of interest. If dealing with parsed data it is important to understand the parsing process, in order to maximise the potential of it delivering meaningful decisions, as well as understand its limitations.

Some video based eye trackers parse the data in real time. Others leave the parsing until after data collection—at which point the parsing is typically performed by the analysis software. There are two broad approaches to parsing. One is a fixation detection approach, with fixations typically being defined by properties such as duration and dispersion. The other is known as a saccade picking approach, with the onset/offset of saccades being defined by velocity/acceleration thresholds. High speed systems tend to use the saccade picking approach as, with sufficiently high temporal resolution and low velocity noise, the onset of a saccade is unambiguous. Typically, saccade onset will be determined by a relatively simple set of rules (e.g. eye velocity increases above X degrees per second). When a saccade picking approach is used, fixations are defined simply as being those samples that are not in a saccade or blink. In other words, the onset of a fixation is defined by the end of the previous saccade, and the end of a fixation is defined by the onset of the subsequent saccade.

8.4.2 Measuring Data Quality

As discussed in Sect. 8.4.1, if data quality is defined as the correspondence between the measurement and reality then we are faced with a problem, namely finding some "true" measure of eye position. In the relatively sparse literature that has directly addressed eye tracking data quality two different approaches have been employed. The first is to use some kind of exceptionally accurate and precise eye tracking technique to act as a proxy for the "real" eye position, and measure eye movements simultaneously with both the "gold standard" technique and the eye tracker whose data quality is of interest. As discussed in Sect. 2, whilst both DPI and scleral search coil techniques have been claimed as "gold standards", there is no such thing as a perfect eye tracker. As mentioned in Sect. 8.2 of this chapter, search coils are known to influence the kinematics of eye movements and lens slippage can also be problematic. DPI trackers report tiny movements of the lens at the end of saccades as changes in eye position. There are also physical limitations with respect to which eye tracking techniques can be performed simultaneously. For example, the infrared emitters in limbus tracking systems could potentially interfere with the infrared illuminators often used in video-based systems. Similarly, it would be very difficult to simultaneously record using a DPI tracker and a video based system given the need for both systems to have an uninterrupted view of the eye being tracked.

The alternative approach is to use some kind of artificial eye whose absolute position at any point in time can be known (typically because it is invariant). This approach (see Reingold, 2014) also neatly avoids another problem inherent in assessing data quality, namely the fact that any spatial inaccuracy in eye movement data is a product of both the eye tracker related error, and participant related error. Eye tracker related error can only be truly estimated if the precise position of the eye being tracked is known—as with an artificial eye. Some manufacturers will report values for spatial resolution that are based on tracking an artificial eye. It should be pointed out that artificial eyes may be easier or harder to track than actual eyes, and as such, and evaluation of eye tracking accuracy and precision should attempt to utilise data from both.

One interesting variant of the artificial eye is one that uses two LEDs to mimic the corneal reflection. If these LEDs are in different locations, rapidly switching from one to the other simulates a saccade. Such devices are particularly useful for measuring latency and end-to-end delay. Whilst a variety of static artificial eyes are commonly used, recent advances now allow "robotic" artificial eyes to be used instead of human eyes (Reingold, 2014).

8.4.2.1 Measuring Accuracy

Accuracy can be calculated as the average difference in the reported location of gaze to the actual location of the target. In practical terms it is often measured at the fixation level. In other words, it is the reported location of the fixation that is compared to the actual location of the target. As is explained in Alexander and Martinez-Conde's chapter in this volume, the eye is never truly still, and there are a large number of different approaches to defining what exactly constitutes a fixation. As such, the algorithm involved in computing the fixation onset or offset and location is a critical determinant of accuracy. Assuming that fixation duration has been established, one relatively straightforward solution to defining a fixation's location is to simply take the average X and Y location of all samples that are included within its duration.

In practical terms, accuracy can be measured by asking participants to repeatedly fixate a target, or targets with known locations. Accuracy can be expressed as the difference in position of the fixation and the target, in either the X or Y dimensions (or both if Pythagoras' theorem is used to calculate the angular distance between the fixation and the target locations).

8.4.2.2 Measuring Precision

Precision is generally reported at the sample level and can be expressed either as the standard deviation of the X, Y locations of samples, or in terms of the distances between samples—in which case it is often expressed as Root Mean Square (Sample to Sample), or RMS-S2S.

8.4.3 *Maximising Data Quality—Participant Set up and Calibration*

Without doubt, one of the most important determinants of data quality (particularly spatial accuracy) is the initial set up and calibration procedure, but it is one that often gets rushed or is done sub-optimally for a variety of reasons.

Obviously, the precise details for optimal participant set up will differ as a function of the eye tracker being used—for example setting up a head-mounted eye tracker is very different to setting up a desktop based system. However, various basic principles are common to all video-based systems, desktop or head mounted. The camera needs a good view of the eye and the eye itself needs to be well illuminated with both pupil and corneal reflection(s) visible. Similarly, a number of factors such as glasses, eye makeup and droopy eyelids can be potentially problematic for all video based eye trackers.

Ideally set-up and calibration would involve a knowledgeable and practiced experimenter, and an alert and motivated participant. Techniques for setting up and calibrating infants and other potentially non-motivated populations are provided in Sect. 8.4.3.2.

8.4.3.1 Set up

Many data quality problems can be avoided if the participant and eye tracker are set up optimally. There are several factors to consider, the relative importance of each depending to some extent on the make and model of eye tracker being used. The following advice is meant to be as generic as possible.

One key factor is the height of participant with respect to the monitor on which the stimuli are to be presented. Many researchers assume that the eye should be aligned with the centre of the screen. In fact, high-end desktop based eye trackers may actually function best if the level of the eye is closer to the top of the screen. Having a height adjustable chair or adjustable table is critical.

Another important factor is distance of the eye from the camera. This is important for two reasons. Firstly, if the eye is too far from the camera, the pupil and corneal reflection will appear smaller, and occlude fewer sensor pixels. The image processing algorithms will then have less data to work with and will produce less stable estimates of their centres. In other words, noise in the data will increase and data quality will decrease. Secondly, if the eye is too close to the camera it is possible that eye movements to targets placed towards the edge of the screen will exceed the trackable range of the system. All eye trackers have a trackable range, often described in degrees of visual angle, which correspond to the maximum rotation the eye can make before tracking becomes inaccurate or impossible (as the corneal reflection will “fall off” the cornea and onto the sclera). Whilst head mounted systems will also have a trackable range, it often less of an issue as the participants are free to turn their head towards objects of interest, thus minimising the rotation of the eye itself. For desktop systems, however, if the monitor is placed too close to the participant, it can be very easy to engineer a situation in which the corners or edges of the monitor fall outside of the trackable range of the system. A good rule of thumb is that at about 60 cm distance 1 cm on the screen equates to 1° of visual angle. Hence a 40 cm wide monitor 60 cm from the participant's eyes would subtend ±20° of visual angle horizontally.

Before starting the calibration, it is always a good idea to check the stability of the set-up by asking the participants to look at the four corners of the calibration space. If the eye tracking software allows it, make sure that the pupil and corneal reflections are visible at each corner, and their shapes are not in any way distorted. One very common issue is for the corneal reflection to become distorted or disappear altogether in one of the corners. This is often a sign that something is sub-optimal with respect to the geometrical arrangement between the participant, the eye tracking camera and the monitor. The simplest solution is often to simply move the monitor slightly further away, or alternatively adjust the height of the monitor or participant such that they are not having to rotate their eyes up or down so much in order to view the problematic area. So for example of the problem is with the top corners, lowering the monitor or raising the participant would probably provide a solution. If the bottom corners then become problematic, the monitor should be moved further away.

Factors that can complicate participant set up or increase "noise" and some possible solutions:

(1) Make up: Eye make-up, particularly mascara, can occasionally cause problems as it tends to be the same colour as the pupil (e.g. black) and can therefore potentially interfere with the image processing algorithms that are attempting to locate the pupil's edge. The extent to which mascara can be an issue depends very much on the eye tracker. Particularly large or even false eye lashes, which can curve such that they overlap with the pupil from the camera's view point can be problematic for even the most robust video based system.

(2) Glasses: Glasses with particularly thick or dirty lenses can obscure the view of the eye from the camera. In addition, glasses may diffuse some of the infrared light typically used to illuminate the eye and provide the corneal reflection. This can be particularly problematic if the lenses have been coated to reflect infrared light. The reflection of the infrared light source from the glasses can sometimes obscure the eye. In many cases simple repositioning of the participant's head by asking him or her to thrust their chin forward towards the camera will move the reflection out of the way. Other tips include bringing the camera closer and tilting it up at a steeper angle. If tracking monocularly, often the simplest thing to do is try the other eye. Often the reflection will turn out not to be problematic for that eye.

(3) Contact lenses: Soft contact lenses tend to remain fixed to the cornea, and are generally not a problem for most video based eye trackers. Hard lenses, on the other hand, are generally smaller, and tend to "float" over the eye—particularly during saccades. As a result, the edge of a lens can get close or even cut across

the pupil, distorting its shape as it appears to the camera, and confusing the algorithms that are attempting to calculate the pupils centre. The position of the lens can also impact on the position of the corneal reflection. In general, the best solution is to ask the participant if they have any glasses they can wear as an alternative.

(4) Monitor distance: It is surprising how often users set equipment up in ways that mean it they eye is forced to rotate beyond the trackable range of the camera in order to view the corners of the monitor. The camera should be placed at the distance from the eye recommended by the manufacturer, and the monitor positioned at a distance that ensures it falls within the trackable range.

(5) Eye (iris) colour: Video based eye trackers rely on being able to distinguish the boundary of the pupil and iris. Certain eye colours (particularly brighter or paler irises which provide less contrast with the pupil under infrared illumination) may be problematic for some models of eye tracker. In very young infants (3 to 6 month olds) the iris lacks pigment, again making it difficult for some eye trackers to discriminate between the pupil and iris under infra-red illumination (a similar issue occurs in people with albinism). Some eye trackers can operate with different frequencies of infrared illumination which improve the pupil/iris contrast in very young infants.

(6) Pupil size: As discussed in Sect. 8.2, a somewhat unfortunate property of pupils is that they do not dilate symmetrically around their centre. As a result, it is important to try to avoid drastic changes in luminance across trials or during the course of a single trial (e.g. black fixation cross on white background followed by face on a black background).

(7) Ocular dominance: Approximately 80–90% of participants are right eye dominant. Perhaps the simplest test of ocular dominance, which works well with both adults and children (who can find winking difficult) is to ask them to extent their arms and form a small opening in their hands through which they look at a distant object (such as the experimenter's nose). Then ask the participant to bring their hands up to their face, whilst continuing to keep the experimenter's nose in view through the opening in their hands. They will bring their hands to their dominant eye. The "extent" of ocular dominance varies greatly between individuals—but tracking from a non-dominant eye can be a cause of accuracy problems. Symptoms of tracking from a non-dominant eye include significant (readily observable) drift during fixations, and poor/inaccurate calibration models/difficulty with validation (because the non-dominant eye tends to go to a slightly different place each time).

A useful approach if recording monocularly is to start calibrating the right eye. If calibration models are poor/attempts at validation give bad results, then one possibility is that you are in fact tracking from the non-dominant eye, so try calibrating the left eye and see if things improve.

(8) Ptosis (droopy eye lids): This can be a particular problem in elderly participants (although it can also occur with tired or drowsy younger participants). The top eyelid can droop sufficiently that it obscures the top of the pupil – potentially causing problems for the image processing algorithms that are try-

ing to determine its centre. A related problem can occur in low light levels if the pupil dilates sufficiently that its edge is partly obscured, even in the absence of a drooping eyelid. Some video based eye trackers allow the user to switch between "centroid-based" and model based estimates of the pupil centre. Model based solutions are often better when the pupil is partially obscured as the model "fills in" the missing edge. Genuine ptosis (as opposed to drowsiness) cannot be corrected other than by taping up the eyelids, a procedure that can be both embarrassing and uncomfortable for elderly participants, and which should be avoided if at all possible. Occasionally repositioning the camera, so that it is looking up into the eye at a steeper angle can help.

8.4.3.2 Calibration

The calibration procedure is the most important determinant of subsequent spatial accuracy—but as with participant set up, it is often rushed, or unwittingly done sub-optimally—with seemingly little or no awareness of the consequences this can have. The aim of the calibration procedure is to provide a mapping function that allows raw eye tracking data (the location of the pupil and CR centres on the camera sensor for example) to be converted to screen (typically pixel) co-ordinates. It usually involves presenting targets at a sequence of known locations (anywhere between 3 and 25 depending on the system and needs of the researcher—although some head mounted eye trackers allow single point calibrations). Calibrations can be one-dimensional (e.g. horizontal only) or two-dimensional.

Once the values of the raw eye tracking data at each location are known various mathematical/modelling approaches can be applied to allow the point of regard to be established for any intermediate raw eye movement values. The simplest approach would be to use a linear regression. Limiting our discussion to the horizontal dimension for simplicity's sake, assume that the participant fixated the following three target positions on the X axis: −200, 0, 200 (assume that the units are pixels and 0 represents the centre of the screen). If the raw X signals from the eye tracker when the eye fixated those 3 locations (averaged over say 100 ms of a fixation) were 100,500 and 900 respectively, the following regression equation ($y = mx + c$) describes the relationship:

$$\text{ScreenPixelX} = 0.5 * \text{ EyeTrackerX} - 250$$

Thus a raw eye tracker value of 700 would equate to a screen pixel position of $(0.5 * 700) - 250 = 100$.

In reality the calibration procedure may involve considerably more complicated non-linear regression or modelling techniques, but the above example illustrates the central concept: the calibration serves to provide some mapping function that allows raw tracker values to be converted to more meaningful units related to the calibrated space (typically a monitor, and thus often expressed in screen pixels). Therefore, it is one of the most important determinants of the spatial accuracy of any eye movement recording.

One important consequence of any calibration that involves regression is that any inaccuracy within the model may become amplified towards the extreme edges of the calibrated space. Taking the simplified example above, assume that when the participant was supposed to be looking at the target 200 pixels to the right of the screen, the experimenter pressed the "Accept" button before they eye made a secondary saccade to get on target. As a result, the participant was actually fixating pixel 195, and the raw tracker value for that calibration target was 880.

The (inaccurate) mapping function is now

$$\text{ScreenPixelX} = 0.51 * \text{EyeTrackerX} - 253.$$

When the participant looks at the centre of the screen (500 in EyeTrackerX; 0 in ScreenPixelX) the equation will now return a value for ScreenPixelX of 3.84, approximately 4 pixels out from the "true" gaze position. However, when the participant looks at the target on the right (900 in EyeTrackerX, 200 in ScreenPixelX) the equation returns a value of ~208, in other words an 8 pixel discrepancy between the true eye position and the reported eye position, that is twice as much error as when the participant looked at the centre.

The purpose of this somewhat laboured explanation is to try to highlight the importance of having an accurate calibration model for getting spatially accurate data. Note that a faulty calibration model will not have any impact on precision, in the case above every time the participant looks at the target at +200 the eye tracker will say they are looking at +208.

Reingold (2014) points out that at the heart of the calibration procedure lies an assumption that is almost certainly false. The assumption is that at the point at which the raw eye tracking data is sampled and assigned to the target's location, the eye was looking exactly at the centre of the target. This assumption is clearly false as we know that the eye is never truly still (see chapter by Alexander and Martinez-Conde in this volume). In addition, there is the simple fact that the participant's actual point of gaze may not be on the target's centre. Foveal vision is capable resolving in high acuity an area of the visual scene 1–2° in diameter—typically larger than the size of a standard adult calibration target. In other words, in order to "see" the calibration target, the participant does not necessarily have to fixate its exact centre. As Reingold (2014) points out, we have limited control in accurately directing our gaze and limited awareness of our actual gaze position. So whilst the calibration procedure will necessarily assume that the participant is looking at the centre of the target, this may not necessarily be the case. Another potential source of error during calibration, as in the worked example above, is that the experimenter presses the accept button (or the tracking software automatically accepts the fixation) before the participant had an opportunity to make a secondary saccade that would have taken their eye closer to the target. In other words, even if the participant was actually capable of looking at the exact centre of the target and keeping gaze fixed in exactly that location, the calibration still requires the operator (or eye tracking software) to make some decision as to whether this is in fact the case (in the absence of a calibration model that provides any information on accuracy), and this decision itself can be erroneous.

So, armed with this knowledge, what can be done to improve the chances of getting an accurate calibration model and maximising the accuracy of your data? Perhaps the most important, and often neglected factor under the experimenter's control is participant motivation. It is necessary to ensure that the participant is alert and trying their best to fixate the centre of each target. The experimenter can assist them in their endeavours by providing them with calibration points with clearly defined—but most importantly, it will be required to tell them what to do. In other words, it is important to explain to participants the importance of the calibration procedure, and that they are supposed to look closely at the centre of each calibration point in turn. Obviously such motivation cannot be expected of children or infants—so other strategies are required. These are covered in the following section. Some eye tracking systems allow the experimenter to choose between "manual calibration" (in which the experimenter accepts each point in turn) and "automatic calibration" (in which the eye tracking software determines when to accept the fixation). It is usually preferable to use the manual approach, particularly with un-practiced participants.

Most modern eye tracking systems provide the user with some kind of feedback as to the robustness of the calibration model. Users need to understand what this feedback is telling them, and use it to ensure that their calibration model is as good as it can be—given the limitations of both the human oculomotor system and their tracker. Corners are often weak points in calibrations—often because of poor set-up. If one or more corners are consistently "out" then consider moving the monitor further back, or adjusting the monitor or participant height.

Calibrating gaze positions of infants and young children

The key to successful infant eye tracking is to ensure that the screen you are using to present your stimuli is by far the most interesting thing in the infant or child's immediate environment. Visual stimuli should be bright/high contrast, and if possible accompanied by a sound. Children and particularly infants will rapidly tire of the same stimulus, so a good strategy is to ensure that you have a sufficient supply of attractive and noisy "attention grabbers" which can be employed whenever their attention wanes. It can help to have a cartoon running on the screen during camera set up if your system allows this.

Calibration targets should be chosen with particular care. Whilst large and ideally animated targets are best at attracting infant's attention, if they are too large, or contain no obvious focal point, the calibration process can be compromised as the infant may look at the bottom left of the target at one position and the top right of the target at another position. One useful approach is to use "looming/shrinking" targets—these are animations which start off small and then expand to attract the infants' attention, and then rapidly shrink back to the calibration location—infants' eyes will tend to follow the object as its shrinks, and the trick is to accept the calibration at the point when the target is at its smallest.

When eye tracking infants it may be preferable to limit the number of calibration points—a 3 or 5 point calibration may suffice. This may have a slight cost in terms of spatial accuracy over some parts of the screen—but many researchers feel that some data is better than no data.

In summary, ascertaining the best possible data quality is an obvious requirement in the recording of eye movements if researchers are to fully exploit the great potential of eye tracking technology in the study of psychological processes. Section 8.4 was meant to provide some practical guidelines in achieving this important goal.

8.5 Suggested Readings

Wade, N. J., & Tatler, B. W. (2005). The moving tablet of the eye: the origins of modern eye movement research. Oxford: Oxford University Press.

– *A fascinating history of eye tracking research.*

Reingold EM. (2014) Eye Tracking Research and Technology: Towards Objective Measurement of Data Quality. Vis Cogn. 22(3):635-652.

– *A very interesting and thorough discussion of the issues surrounding data quality.*

8.6 Questions Students Should Be Able to Answer

What techniques are currently used to track eye movements?
What are the main advantages of video-based eye tracking?
Why do video-based eye trackers often track the corneal reflection as well as the pupil?
What is the "trackable range" of an eye tracker?
What does "end to end delay" refer to?
How do you decide which of the available eye tracking systems is suitable for your planned study?
How can data quality be measured?
How can an eye tracker be precise but not accurate?
What steps can the experimenter take to ensure the best data quality?

References

Bour, L. J., van Gisbergen, J. A., Bruijns, J., & Ottes, F. P. (1984). The double magnetic induction method for measuring eye movement–results in monkey and man. *IEEE Transactions on Biomedical Engineering, 31*(5), 419–427.

Caldara, R., & Miellet, S. (2011). iMap: a novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods, 43*(3), 864–878.

Delabarre, E. B. (1898). A method of recording eye-movements. *American Journal of Psychology, 9*(4), 572–574.

Deubel, H., & Bridgeman, B. (1995). Fourth Purkinje image signals reveal eye-lens deviations and retinal image distortions during saccades. *Vision Research, 35*(4), 529–538.

Diefendorf, A. R., & Dodge, R. (1908). An experimental study of the ocular reactions of the insane from photographic records. *Brain, 31,* 451–489.

Dodge and Cline. (1901). The angle velocity of eye movements. *Psychological Review, 8,* 145–157.

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research, 43*(9), 1035–1045.

Frens, M. A., & van der Geest, J. N. (2002). Scleral search coils influence saccade dynamics. Journal of Neurophysiol. *2*, 692–698.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: University Press.

Hooge, I., Nyström, M., Cornelissen, T., & Holmqvist, K. (2015). The art of braking: Post saccadic oscillations in the eye tracker signal decrease with increasing saccade size. *Vision Research, 112,* 55–67.

Huey, E. B. (1898). Preliminary experiments in the physiology and psychology of reading. *American Journal of Psychology, 9*(4), 575–886.

Kimmel, D.L., Mammo, D., Newsome, W.T. (2012). Tracking the eye non-invasively: Simultaneous comparison of the scleral search coil and optical tracking techniques in the macaque monkey. Frontiers in behavioral neuroscience *14;6*, 49.

Mowrer, O. R., Ruch, R. C., & Miller, N. E. (1936). The corneoretinal potential difference as the basis of thegalvanometric method of recording eye movements. *American Journal of Physiology, 114,* 423.

Nyström, M., Andersson, R., Magnusson, M., Pansell, T., & Hooge, I. (2015a). The influence of crystalline lens accommodation on post-saccadic oscillations in pupil-based eye trackers. *Vision Research, 107,* 1–14.

Nyström, M., Hooge, I., & Holmqvist, K. (2015b). Post-saccadic oscillations in eye movement data recorded with pupil-based eye trackers reflect motion of the pupil inside the iris. *Vision Research, 92,* 59–66.

Reingold, E. M. (2014). Eye tracking research and technology: Towards objective measurement of data quality. *Visual cognition, 22*(3), 635–652.

Wade, N. J., & Tatler, B. W. (2005). *The moving tablet of the eye: The origins of modern eye movement research*. Oxford: Oxford University Press.

Wildenmann, U., & Schaeffel, F. (2013). Variations of pupil centration and their effects on video eye tracking. *Ophthalmic and Physiological Optics, 33*(6), 634–641.

Yarbus, A. L. (1967). *Eye movements and vision*. Springer.