

## 2 Eye-tracker Hardware and its Properties

---

This chapter introduces the machines that we call *eye-trackers*, their properties, and the people manufacturing and using them. It provides a minimal level of technological detail for researchers who record and use eye-movement data. Earlier drafts have been reviewed by staff from SR Research Ltd., SMI GmbH, and Tobii AB, and by independent researchers with expertise in developing eye-trackers. The chapter is structured like this:

- First, a brief historical look back on how eye-trackers were built and used.
- Manufacturer customer relations are the focus of Section 2.2 (p. 12).
- In Section 2.3 (p. 16), we present a list of issues important to consider before acquiring an eye-tracker, and before setting up an eye-tracking laboratory.
- Section 2.4 (p. 17) focuses on the properties of the environment where you make your records (the 'lab'), and the competences needed in it.
- Section 2.5 (p. 21) gives a very condensed overview of the eye movements measured by an eye-tracker, and then describes the dominating measuring principle: the pupil-and-corneal-reflection method.
- In Section 2.6 (p. 29), we review the quality of recorded data. Methods for measuring data quality are discussed here.
- Section 2.7 (p. 51) covers the set-up of eye cameras and infrared illumination. This set-up differs depending on your chosen hardware, and your particular research question. These issues will be addressed with various pictures taken from our lab to illustrate the types of problems you might encounter, and solutions to them.

If you plan to buy an eye-tracker, read the entire chapter before ordering, considering how the different technical aspects described will impact the specific type of research that you wish to carry out. If you are just now beginning with an eye-tracking project, read only the summary checklist given in Table 2.2, and then go to Chapter 3. If you already have an eye-tracking lab up and running, you may find this chapter useful for organizational and technical background. Or if you only require a quick introduction to the technical terms used to describe eye-tracker properties, read from Section 2.5 in this chapter.

### 2.1 A brief history of the competences around eye-trackers

The earliest eye-trackers were built in the late 1800s. They were technically difficult to build, mostly mechanical, and not very comfortable for the participants. Huey (1898) used a bite-bar with partially cooled sealing-wax attached to the mouth-piece; this ensured participants kept their heads still. Delabarre (1898) anaesthetized the eyeball by applying a solution of two to three per cent cocaine; a Paris ring was then attached to the eye which connected it to a mechanical level. Only at the beginning of the twentieth century did Dodge and Cline (1901) introduce the principle of photographing the reflection of an external light source from the fovea. This is much less invasive and in recent years has become the dominating technique for recording eye movements.

From around 1950, individual researchers developed a number of different techniques, the most common of which are the following:

- Lense systems with mirrors (some protruding, making blinking difficult) were used by Yarbus, Ditchburn and others in the 1950s to 1970s. Having a very high precision, these highly uncomfortable contact lenses made possible the recording of very detailed movements of the eye.
- Electromagnetic coil systems, which measure the electromagnetic induction in a silicon contact lense placed on the anaesthetized eye, were long considered the most precise method of measuring any eye movements (Collewijn, 1998), but are now known to alter the saccades of participants who wear them (Frens & Van Der Geest, 2002; Träisk, Bolzani, & Ygge, 2005). Contact lenses had to be modelled individually for each participant, and even then often remained uncomfortable.
- Electrooculography (EOG) systems measure the electromagnetic variation when the dipole of the eyeball musculature moves. Also, EOG systems typically only measured horizontal movements, and suffered from the electromagnetic noise of surrounding muscles. They still exist as a low-cost variety of eye tracking, having a high sampling frequency, although accuracy is poor due to drift.
- The Dual Purkinje systems from Fourward Technology were very expensive, difficult to maintain, had a very small visual field of recording, but were also extremely precise and accurate without having to place something directly onto the participants eye. Using video, they recorded data using both the first and the fourth Purkinje reflections (Crane & Steele, 1985). However, Deubel and Bridgeman (1995) presented data indicating that saccade endings are inadequately measured with this technique, due to the fourth reflection.

For a closer review of these and other early technologies, see Duchowski (2007, pp. 51–59), Rötting (2001, pp. 41–53), Ciuffreda and Tannen (1995, pp. 184–205), Young and Sheena (1975), or Ditchburn (1973, pp. 36–77). Additionally, Wade and Tatler (2005) offer an excellent historical overview of eye-movement research.

Throughout most of the twentieth century, eye-movement researchers were required to build their own systems before using them to do research. Ready-made, over-the-counter eye-trackers were simply not an alternative. Even in relatively recent times, they have had to build their own eye-trackers themselves, from electronics that had to be understood in detail, as the following methods section from Heywood and Churcher (1981) clearly exemplifies:

The experiments were carried out in a darkened room. Participants sat, their heads fixed by a dental bite and a forehead rest, facing a Tektronix 604 display CRT (31 phosphor) 51 cm away at eye level. The movements of their right eyes were recorded using an infrared photoelectric technique modified from those described by Wheelless et al. (1966) and Jones (1973). An image of the eye, lit with partially collimated infrared light from a GaAs LED (Texas Instruments, TIXL 16), was formed on a perspex screen, in which were mounted two infrared sensitive photodiodes (Texas Instruments, TILSI). The system was constructed so that with appropriate positioning of the two photodiodes, subtraction of their output signal yielded a signal linearly related to horizontal eye rotation over the central  $\pm 15^\circ$  of the visual field, and addition of their output gave a similar signal related to vertical eye rotations linear over the central  $\pm 10^\circ$ . In the present experiment we were concerned only with horizontal eye movements. The signals were amplified and were sampled by a computer (CAI Alpha LSI 2) that also controlled the participants' displays and recorded the movements of the target. The entire system had a half-power bandwidth of 330 Hz: the sample rate in the present experiment was 500 Hz, and the resolution of the system was better than 6 min arc.

Being required to build their own hardware had several disadvantages for scientists: Above all, it slowed down their research. It made eye tracking more exclusive and often completely impractical. However, running self-built eye-trackers had advantages, too: a researcher who has also built the system is more likely to know the properties of the data, and what filters and settings are necessary. Eye-movement behaviour can be more readily differentiated from the artefacts of the measurement system. Algorithms can be more closely attuned to actual eye-movement behaviour. Errors are easy to diagnose; and maintenance operations do not risk data quality as easily.

Beginning in the mid 1970s, this situation changed profoundly. Companies driven by engineers, such as ASL (Applied Science Laboratories), were beginning to build and sell eye-tracking systems to researchers. Ten years later, there were many companies offering eye-tracking hardware. Being able to buy eye-trackers made eye tracking more accessible and versatile. Suddenly researchers could focus on their research, leaving the technical issues to the manufacturer.

Of course, having one group of people *building* eye-trackers and another group *using* them gives rise to an unfortunate split of competencies. It is unfortunate for the researcher, because it is difficult to interpret (with absolute confidence) the data output from a system which they did not design. Likewise, researchers are often not trained in the technical skills required to maintain the system, and influence the technology and its development. Conversely, this situation is unfortunate for the manufacturer, because it is harder to build a system if you do not know exactly what it is going to be used for. It is uncommon for manufacturers to be trained in the principles of experimental design and statistical analysis, therefore the software and hardware requirements of researchers may not be met.

The number of researchers and others who use eye-trackers has grown enormously over the past 20 years. Current-day users of eye-trackers can choose between a large number of different systems from many competing manufacturers. Many of these strive to make it (seem) easy and effortless to do eye tracking. In fact, one line of current commercial development is going towards eye-tracking systems that require almost no system knowledge on the part of the user. In marketing research, some users are asking for eye-tracking systems that are so easy to use that it allows them to show a number of advertisements to any groups of participants, and then just press one button to get a diagnosis of the advertisement, without thinking about any of the technical properties of the eye-tracker, let alone experimental design.

In reality, eye-trackers are advanced physiological measuring systems, and they are produced in small series. Not enough people have tested and given feedback on them for you to be able to trust their functionality like you trust a DVD player, a microwave oven, or even a laptop computer. There will be difficulties in measurements, data quality issues and even bugs, and the diagnoses and workarounds will require system knowledge. We can use one DVD player for all disks, but not one eye-tracker for all studies. There are many technical aspects to eye-trackers that decide whether your particular study will be feasible with a particular eye-tracker. Researchers who understand their systems are much more likely to produce reliable results, and knowledgeable customers are much more likely to get a system they can actually use for the intended purpose.

Therefore, in this chapter, we will spend some time reviewing current types of eye-trackers and their properties, but also where eye-trackers should be located for optimal recordings and usability, and what sort of infrastructure is needed around them. First, however, we must discuss the current manufacturers and their complex relation to the researchers and others who buy equipment from them.

## 2.2 Manufacturers and customers

Before the 1980s, most researchers both built *and* did research with their eye-tracking system. Today, that single role is divided into two principal parties: the manufacturer and the researcher. Manufacturers have different origins. Some were founded by researchers, like SR Research with the EyeLink family of eye-trackers, co-founded by Dave Stampe and Eyal Reingold of the Department of Psychology at the University of Toronto, or Interactive Minds which grew from a group at the Department of Applied Psychology in Dresden. Others sprung out of engineering research work, for instance ASL (Applied Science Laboratory), which originates from MIT research in the 1960s and 1970s, and SMI (SensoMotoric Instruments), which spun off from an engineer's PhD thesis (Teiwes, 1991) on torsional eye tracking in neurological applications in the early 1990s.

In spring 2009, we could find 23 companies prepared to sell video-based eye-tracking systems to us. A handful of companies sell eye-trackers based on now less common principles, for instance coil systems, EOG systems, and diode-sensor systems. Of the 23 manufacturers of video-based eye-trackers, three were founded before 1985, and more than 50% after 2000. Most of them only sell one or two products, and several of those appear to have only a small customer group. The vast majority of companies have been founded by engineers and applied physicists, and very few of the companies had a psychologist in the group of founders.<sup>1</sup>

Over the past twenty years, several other manufacturers started but never grew large, and finally vanished. All the time, new people reinvent the wheel, possibly oblivious of the market situation, and this gives rise to media reports of "a new fantastic invention that can measure where people are looking", with predictions of the many applications such a tool could have. These inventions are seldom long-lived.

A few researchers continued building their own eye-trackers to give them the precise properties they required for their research. Mike Land's portable head-mounted eye-tracker from the mid-1990s is a good example, and Jeff Pelz and colleagues at Rochester also customize eye-trackers with similar goals in mind: to investigate eye movements in natural environments and when performing everyday tasks.

Sometimes a specific line of research develops its own eye-trackers, which is the case with the Visagraph, a series of family-based low-cost eye-trackers that have been used in diagnostic optometric reading tests in schools for several decades, but not in any other research. Similarly, the company Verify International, before going out of business in 2007, built their own eye-trackers for consumer research (known through publications such as Pieters & Wedel, 2004).

There is also an ongoing effort to develop smaller, less expensive, and more accessible eye-trackers, coordinated in the academic network COGAIN,<sup>2</sup> and it may not be long until it is technically feasible for each laptop to have a simple built-in eye-tracking function. Today, it is even possible to turn your webcam into a simple eye-tracker.<sup>3</sup>

In the following bulleted list we cover the main customer groups, focusing on the large manufacturers who supply them. We have first-hand knowledge of these companies, their eye-trackers, and the support which they offer.

- The academic researcher group is definitely the oldest, and probably the largest of the customer groups. It is fairly stable over time, but also very heterogenous in research themes. Dispersed over almost all disciplines of science, they are united by a desire to

<sup>1</sup>The exact number is not easy to determine, but it is safe to say that psychologists are part of the founders/owners in at least four of these 23 companies.

<sup>2</sup><http://www.cogain.org/>

<sup>3</sup>e.g., <http://www.gazegroup.org>

use proper experimental set-ups and statistics, often also emphasizing precise timing, accuracy, precision, and high sampling frequency in data. Researchers usually buy their eye-trackers as part of building or expanding a lab, or after having received funding for a project. In the authors' experience, for more than a decade, the leading manufacturers that provide for this demanding customer group are SR Research with the EyeLink system, SensoMotoric Instruments (SMI), and Applied Systems Laboratory (ASL), with Tobii Technology opening up and taking a leading role in some applied parts of academia.<sup>4</sup>

- Another large but much more recent group of eye-tracker customers is the media and advertisement consultants, who ask for eye-trackers that are simple to use. These consultants often want to rent the eye-tracker for a specific project, rather than buy it. They use eye-trackers as a method, among others, to decide whether to say "no" or "go" to an advertisement campaign, and they are often happy with heat map representations of data (see Chapter 7). Experimental designs and statistical significance tests typically do not give any added value to media or advertisement consultants. Also, some non-academic usability testers share methods and requirements with this user community. Many of the companies compete to sell to these users, but Tobii has dominated this customer category since the mid 2000s.
- Human factors researchers make up a small group that has existed for a long time. They have specific demands to be able to use eye-trackers in the field: in cars, nuclear plants, aeroplanes etc. Applied sports psychologists and consumer researchers can also have the same type of requirements, along with a few 'real-world' academic researchers. ASL, SMI, Smart-Eye, and Seeing Machines have traditionally had a focus on these users, but several others are also selling to this varied group.
- There is a group of clinical users of eye-trackers who are not interested in gaze positions per se, but are more concerned with movement patterns such as nystagmus, deviant saccadic forms, oculomotor dynamics, and torsional movements of the eye. Calibration must be possible even for participants who cannot fixate properly. The eye-trackers used for these purposes and the studies carried out, are often designed for diagnosing individual participants, or identifying core functional deficits of a visual disorder, rather than for testing large groups to find generalizable results.
- A previously small, but in later years very rapidly growing, group is the users of gaze-guided computer interfaces. They cannot operate a computer by other means, because of a disability or because the task requires their hands for other jobs. LC Technology was an early player in this group, now followed by Tobii and a few small companies. For this group, low price is often prioritized over high data quality. They want to interact with their computers using the eye-tracker on a one-to-one basis, and do not usually work with precise experimental designs and statistics over several users. Knowing the technical details of the system is often of little interest.
- Eye laser surgeons comprise a user group for which the primary interest is an accurate, precise, and quick eye-movement signal, that can be used to move a laser knife to compensate for the signal change when an eye movement occurs. Hardware and software requirements, as well as gaze estimation method, differ between this user group and all others.

<sup>4</sup>The strength of each company can be measured differently, for instance taking into account: 1) the number of peer-reviewed journal publications where their system was used, divided by research area and impact factors, 2) the number of systems sold, 3) the number of employees (divided into developers versus sales), 4) the presence of the manufacturer at conferences and meetings, and 5) the image they have in the community.

The users of eye-trackers, such as those listed above, also differ very much in their technical competence. Some labs, and some branches of research as a whole, have better skills than others to evaluate the technical properties of an eye-tracker before purchasing. Programming stimulus presentation, carrying out successful recordings, and developing algorithms and software for data analysis are all important considerations before buying an eye-tracker; proficiency in these abilities differs a lot between the user groups mentioned above, and this can effect the choice of which eye-tracker to buy and the validity of the studies carried out with it. Variance in the competencies required to evaluate and run eye-trackers, and program auxiliary software which compliments their use, exist both within academia and outside of it.

The *diversity* of customer groups has lead to manufacturers producing very different kinds of eye-trackers. Some of the manufacturers have a long history of providing eye-trackers of multiple types, for instance ASL and SMI, and to some extent SR Research (by offering optional extensions), while most of the others have concentrated on the specifics of their main target group.

Not only hardware and recording software, but also stimulus and presentation tools vary between manufacturers. The simple rule seems to be: the more a manufacturer provides for the academic community, the more versatile and powerful is their stimulus presentation tool, simply because the (predominantly academic) customers of those manufacturers have asked for solutions that support a large range of experimental designs. Manufacturers who mostly cater for the advertisement and usability customer group, typically offer slide shows with limited support for running sophisticated empirical studies; they rather emphasize web support in their stimulus tools.

The analysis software from manufacturers also reflects which customer groups they have. Many customers investigating applied domains (e.g. website usability) mainly care about *visualizations* of eye-tracking data (heat maps and scanpaths, see Chapters 7 and 8 for a thorough discussion), rather than graphs and appropriate statistical comparisons. The analysis software, and the way the salespeople present it, will then focus on visualizations that look good in demos, but will only have limited options for exporting data (other than raw data, fixation sequences, and area of interest hits). The researcher, however, is trained to trust a result *only* if it comes from a correctly performed experiment, with statistically significant effects. Therefore, the manufacturers with more academic researchers as customers have analysis software that allows for a variety of different experimental designs, and which can export a rich range of outcome measures (dependent variables).

The background of the manufacturers' *sales people* differs wildly. Before you take advice from any salesperson, find out what their *background and motivation* is: are they newly graduated engineers who know very little about eye-tracking research, or do they have comprehensive experience and know-how? If you press them on a technical issue, will they simply guess that their system handles it, or can they tell you about the technical properties of the system and the motivation behind it? How well do they understand the scientific aspects of eye-tracking research, and the role of their equipment and software in your workflow? What is their contact network in the scientific and applied fields of eye-tracking, and will they make it available to you? Can their claims about the role of their company and products among researchers and practitioners be supported by independent sources? Their prime motivation will always be to sell you a system, but does the salesperson have additional motivations that are beneficial to you: to maintain good relations to you as a customer, using you as a reference for future customers, to assist you in publishing papers in scientific journals; perhaps they are only interested in adding you as an additional node in their network clients to help them gain future sales.

Many customers over-estimate the *scientific competence of the manufacturers*, thinking that they not only produce the systems, but also know exactly how to use eye-trackers to

do research, which settings to use for algorithms, and how to interpret the many metrics which can be derived from the data eye-trackers produce. All manufacturers know a lot about which cameras to use, the algorithms and filters used to process the video image, and the mathematics underlying gaze estimation, but too few staff of manufacturers are academics who publish research results. For a lot of the staff of manufacturers eye-trackers are products rather than research tools. Eyal Reingold of SR Research is a notable exception—with vast research experience in vision and eye movements he is a co-founder of the EyeLink system, arguably the most dominant eye-tracker on the market for the academic user group.

As competence in how to use eye-trackers for research still resides in the research laboratories, manufacturers will always need close collaboration with researchers who are well acquainted with the requirements and workflow of real eye-tracking studies when they make decisions about how to develop their hard- and software. When you talk to a manufacturer, try to find out how they gain access to scientists' experiences.

Sales people from the companies with a strong position in the academic world invariably say two things that are important to consider. First, they do not want to sell you something that you cannot use. This is a sign of normal business ethics. Second, if you want a property or functionality that they do not currently have, but which is interesting also for other customers, they will try to implement it and add it not only to the system that you buy, but to the future product line of the company. This is not only salespeople's talk; the authors of this book have witnessed several joint development projects with the three leading companies and researcher groups at universities. Customer requirements are a major factor behind product development, and it is important to evaluate a company in terms of how well they integrate requests (and bug reports) without compromising the integrity and overall consistency of their system.

If employees in a manufacturer company are often exchanged, there is a risk that their competence is lower than in a company where key staff have worked longer. Both technical development and customer relations are competencies that take time to form.

Evaluate the manufacturer's history of software upgrades. Are the upgrades coming at a reasonable rate? Do upgrades solve or address important issues? Is upgrading easy to perform? Are the upgrades done in such a way as to support comparability of results across software versions (for instance, when a new event detection algorithm is introduced)?

The manufacturer's *support line* is often the only remaining link between the manufacturer and the researcher. As an eye-tracker is a piece of equipment that often requires its owner to contact support, evaluate manufacturer support before acquiring the system. Be aware that the different manufacturers have very different reputations with respect to their support line. Some are extremely helpful, specific and quick, inviting discussion with a dedicated company representative who focuses on finding a solution that works for you as customer, and gives you feedback on how your request is being processed.

There is *no standardization* between systems. Terms for measurement quality differ, as well as the methods manufacturers use to calculate reported performance figures for their systems. Many of the concepts of recording and analysis also differ between manufacturers. Thus, comparing the webpages and specification sheets between several manufacturers and using that as a basis for acquiring an eye-tracker is often confusing and of little use.

*Technological transparency and openness* varies between manufacturer companies. There are several important aspects to this:

- Manufacturers vary in *how they record, calculate, and report performance values* such as precision. If you want to be sure, make your own tests.
- Some have developed *technically transparent recording software*, so that the user can see and control virtually everything. Several companies, for instance SMI, SR Research and ASL, have had this policy for a long time. The direct opposite is to hide the record-

ing settings, eye-video, and data viewers and only supply as little control as possible to the user. These opposing strategies obviously address the more versus less technical user groups. Technological transparency does require more of the user, in terms of getting over a competence threshold, but in the authors' experience, gives better and more easily comprehensible data.

- The *analysis software* of some manufacturers allows direct control over filter and algorithm settings, while other software has reduced access to or even makes settings inaccessible defaults. Again, having access to both allows and requires an understanding of the analysis tools, and increases your chances of performing a good, valid study.

Some customers would like their eye-tracking systems to be plug-and-play, with technical detail well hidden. Other customers are deeply suspicious of hidden details, knowing that the 'clean' data emerging may have lost the effect they looked for or introduced artefacts. In 2007, for example, a group of some 20 dissatisfied European academic users wrote a common letter to one manufacturer, demanding to have the source code for their fixation detection algorithm. In the long run, technological transparency and openness give better customer relations.

Still, relatively few users switch between manufacturers. *Long-term customer-manufacturer relationships* are common, for many reasons. The users are acquainted with the hardware, the stimulus and analysis software, and the way to work with them. In particular for users with lower technical competence, learning to use a second system, and having two manufacturers to talk to, is seen as a cost, which must be outweighed by the perceived improvement in the newly acquired technology, be it better precision or improved functionality.

To summarize so far, if you plan to use or even buy a particular eye-tracker:

- Find out who the manufacturers are and what competences the salespeople have that advise you to buy their system.
- Contact representatives of the manufacturer's customer group. Read their publications (or reports), talk to them at conferences, and visit their labs. If you need other things in your eye-tracker than what they make use of (or know about), check that these specifications are actually met. If their eye-trackers require other technical competences than those that you have, is it likely that you can gain the required skills?
- Take an academic course in eye-tracking methodology.
- Borrow a system and test it. The properties that we describe in Chapters 2–9 can mostly be turned into a test of hardware, algorithms, stimulus, and analysis tools.
- When talking to the manufacturers, bring the checklist in Table 2.2, and add your own points. Do not forget to ask them which variety of the algorithms and filters described in Chapters 2–9 they have implemented, and why.

### 2.3 Hands-on advice on how to choose infrastructure and hardware

In Tables 2.1 and 2.2, we provide condensed advice on the topics of the chapter. If you are setting up an eye-tracking laboratory, or if you let an established eye-tracking laboratory host your study, you might be interesting in checking Table 2.1. Table 2.2 is a list of which hardware and system properties addressed above should be checked before you buy an eye-tracker, or in any other way decide to use an eye-tracker in a study. Note that some properties concern only one type of eye-tracker, and that some are more fundamental than others.



**Table 2.1** Properties of the recording environment and skills of those who run it.

Property to check	Risks if you ignore the property
Cramped recording space	Uncomfortable participant
Lab availability	Difficult to get participants to show up
Sunlight, lamps and lighting conditions	Optic artefacts, imprecision, and data loss
Electromagnetic fields	Optic artefacts, inaccuracy, and data loss (magnetic headtracking systems)
Vibrations	Variable noise, low precision
Scientific competence of technical staff	Invalid, unpublishable results; time-consuming studies
Recording experience of staff	Data quality low
Programming experience of staff	Data analysis very time-consuming
Statistical experience of staff	Invalid, unpublishable results; confusion

## 2.4 How to set up an eye-tracking laboratory

An eye-tracking laboratory needs both physical space for the eye-tracker and the experiments, and an infrastructure that keeps the laboratory up to date and running.

### 2.4.1 Eye-tracking labs as physical spaces

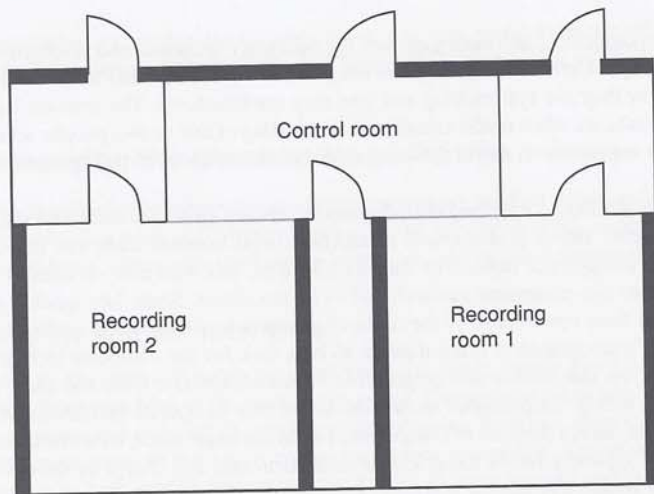
There is not one single solution for designing an eye-tracking laboratory. Every place where there are active people can be made into a place where researchers eye-track people. Take a car with a built-in eye-tracker and other measurement systems, or the mobile eye-trackers that we used in supermarkets for a study of consumer decision making. Neither are labs in the traditional sense. So, what is an eye-tracking lab, and how should it be designed? Most researchers work with single monitor stimuli, rather than real-life scenes. They then, in the authors' experience, prefer sound and light isolated rooms, minimizing the risk of distracting participants' attention from the task. They also tend to put their eye-tracker in very cramped locations (cubicles), where there is little room to turn around, let alone rebuild the recording environment for the needs of different studies.

In our lab, we found it useful to make the windowless recording rooms large enough (around 20–25 m<sup>2</sup>) to be able to rebuild their interior depending on the varying needs of different projects (see Figure 2.1). Many labs—including our own—have also built one-way mirror windows between recording rooms and a central control room. This allows the researcher controlling the experiment to leave the participant(s) alone with their task, whilst still being able to monitor both recording status on the eye-tracking computer, and the participant through the one-way mirror. Having several recording rooms allows for multiple simultaneous recordings. At our lab in Lund, this has proved valuable more than once, when large data collections are to be made in a short period of time.

It is useful to minimize direct and ambient sunlight (i.e. to have few or no windows), and to illuminate the room with fluorescent lighting (the best are neon lights), which both emits less infrared light and vibrates less than incandescent bulbs (the worst are halogen lamps). Figure 4.15(a) on page 126 shows what a halogen lamp can do—note, do not make the room too dark, as this makes the pupil large (and variable), affecting data quality for most eye-trackers. A bright room keeps the pupil small even with a variable-luminance stimulus, which generally makes the data quality better. Also, in darker rooms the participant may

**Table 2.2** Eye-tracker properties to ask manufacturers about.

Property to check	Risks if you ignore the property	Page
Manufacturer staff and openness policy	Poor support; strange errors in the system that are not explained to you	15
Manufacturer major user groups (publications; visits)	System properties that you need may be lacking	12
Software upgrade cycles and method	No improvement in software for years; a lot of hassle with software details	-
What eye movements can the system measure?	Study impossible to operationalize	23
Bi- or monocular	<i>Small</i> differences in fixation data go unnoticed	24
Averaging binocularity	Large offsets when one eye is lost	60
The quality of the eye camera	Noise (low precision)	38
Can the eye image be seen?	More difficult to record some participants; poorer understanding of system	116
The gaze estimation algorithm	Low data quality (precision and accuracy)	27
Frequency of infrared used	Poor data outdoors and in total darkness	-
Sampling frequency	You <i>may</i> need to record much more data; velocity and acceleration values invalid	29
Accuracy	Spatial (area of interest) analyses will be invalid	41
Drift (accuracy drops over time)	Constant recalibration; experimental design changes	42
Precision	Fixation and saccade data will be invalid; gaze contingency difficult; small movements not detectable	34
Filters used in velocity calculations	Fixation and saccade data will be imprecise	47
Headbox (remotes)	Data and quality loss when participant moves	58
Head movement compensation algorithm	Noise (low precision); spatial inaccuracy	-
Recovery time	Larger data loss just after participant moves or blinks	53
Latencies (in both recording and stimulus software)	Invalid results; gaze contingency studies impossible	43
Camera and illumination set-up	Data recording difficult or not possible with glasses	53
Robustness, the versatility for recording on more difficult participant populations	Data loss and poorer data for many participants	57
Portability of mobile system	Cannot be used out of laboratory	-
Connectivity	Difficult or impossible to add auxiliary stimulus presentations or data recordings	-
Tracking range	Data loss when participant looks in corners	58
Reference system for output coordinates	Data analysis very time consuming for some head-mounted systems	61
Parallax	Small and systematic offset in gaze-overlaid video data	60



**Fig. 2.1** Layout of one recording area in the Humanities Laboratory at Lund University. Each recording studio is 25 m<sup>2</sup>. Ante-chambers allow for reception of participants, storage, and a space for the researcher to work between participants. We also have several additional recording rooms for minor studies, technical maintenance, and storage.

see the infrared illumination reflected in the mirror, although this depends somewhat on the wavelength of light emitted from the illumination.

Sounds will easily distract your participant's visual behaviour, so it is advisable to use a soundproof room if you can. For sensitive measurements, place the eye-tracker on a firm table standing on a concrete floor. Do not allow the participant to click the mouse or type on the keyboard on the same table where the eye-tracker is located. Also minimize vibrations from nearby motion of people or outside traffic. If you are using magnetic systems for head-tracking, also minimize various sources of electromagnetic noise (lifts, fans, some computers) in the recording and neighbouring rooms. Stabilized electrical current is an advantage for some measurements, but not critical.

If you are recording eye-tracking data in fMRI systems, the strong magnetic fields require optimized eye-tracking equipment to be used, typically built to film the eye from a safe distance with long-range optics and mirrors near the face. With MEG systems, no auxiliary electromagnetic field may be introduced, and therefore long-distance eye-trackers are also used.

If possible, have your laboratory close to a participant population, or at least make it easy for your participants to reach your lab. That makes it easier to set up a 'production line' where participants arrive one after the other to large recordings.

#### 2.4.2 Types of laboratories and their infrastructure

There are labs that have done eye tracking for 20 years or more, and there are others that have just started. There are labs that only serve a few researchers around the owner of the system, and there are labs that actively invite others to use their equipment, against a cost. There are some eye-tracking companies that conduct studies on a commercial basis.

The largest *commercial* eye-tracking labs have 20–50 eye-trackers to test advertisement campaigns. They are connected to, and sometimes part of, the largest, well-known consumer

product companies, and have gathered the necessary technical and scientific competence in their groups. Unfortunately, they do not publish their work, and they are reluctant to talk about how they use eye tracking and how they are structured. The smallest commercial eye-tracking labs are often media consultants, consisting of one or two people, who often have no previous experience in any of the competence areas necessary to do high-quality eye-tracking work.

Many *academic* eye-tracking laboratories consist of a professor and one or two graduate students and/or post-doctoral researchers, who between them can mostly provide the *scientific* competence needed in their own studies, and who can—if needed—also read up on previous eye-movement research and its technicalities. Some labs quickly grasp the technology of their eye-tracker. If the research group is less technically inclined, the necessary *technical* maintenance is often thought to be a task for the computer technician in the department, the one who is also responsible for email, servers, web, and some programming. However, unless the computer technician learns how to operate and design studies with the eye-tracker, such a division of competences is, in our experience, an unfortunate organization of labs. It typically forces the graduate students to take full charge of the eye-tracker, solve technical issues, upgrade the software, and maintain contact with the manufacturer's support line. The graduate students do this because anyone who makes a change in hardware or the settings of the recording system must also understand that system well enough to be able to make a recording, and see that the data quality requirements for the next study in line are satisfactorily met. Since data quality issues are central throughout the research process, from data recording, over the various stages of data analysis, to the responses from reviewers to submitted manuscripts, satisfactory diagnosis and maintenance of an eye-tracker can only be done by a person confident in *all* aspects of this research process. It can be difficult to find an employee who is sufficiently competent in every one of these skills, and inevitably mistakes will be made as graduate students learn.

Ideally, a *larger lab* is headed by a person who has both technical and research background, someone who can bridge the competence gap that originates from the time when eye-trackers began to be manufactured and sold. This means knowing the recording technology in enough detail to know what a good signal is, to diagnose and remedy errors, to be able to record and analyse data, and follow the research process all the way from hypothesis formulation to reviewer comments and publication.

Since recording high-quality eye-tracking data requires expertise that they can only get from experience, it is important for the staff doing the actual recording work to take part in many recordings with different participants, stimuli, and tasks. As the quality of recorded data is important for subsequent data analysis, it is easier if the same person does both recording and analysis; it is better if the researcher with the highest incentive to get good data takes part in the recordings, so she can influence the many choices made during eye-camera set-up and calibration (pp. 116–134). The exception is when the analysis is subjective in nature and needs to be performed by a person naive to the purpose of the experiment. Any staff who meet and greet participants should have appropriate professionalism for the job; they should be able to answer questions relating to the experimental procedure which the participant is about to undergo.

It is very useful if laboratory staff are also knowledgeable in programming, both for stimulus presentation and for data coding/analysis. Matlab, R, and Python are the preferred software in our and many other labs. If you have scientific ambitions—following the standards of peer-reviewed journals, rather than having heat maps as deliverables—it is also very useful to have a dedicated methodological and statistical specialist in your laboratory.

Since it may be difficult to find all these qualities in one person, you may need several staff members in your lab. Finally, whether you alone carry the full responsibility of your

eye-tracking laboratory, or you share it with others, it is very useful to be part of a laboratory network, sharing experiences, knowledge, and software.

## 2.5 Measuring the movements of the eye

This section introduces the major eye-movement measuring method in use today, the pupil-and-corneal-reflection method. To better understand the principles of this measurement technique, we will begin with a very brief survey of the eye, and its basic movements.

### 2.5.1 The eye and its movements

The human eye lets light in through the *pupil*, turns the image upside down in the lense and then projects it onto the back of the eyeball—the *retina*. The retina is filled with light-sensitive cells, called *cones* and *rods*, which transduce the incoming light into electrical signals sent through the optic nerve to the visual cortex for further processing. Cones are sensitive to what is known as high spatial frequency (also known as visual detail) and provide us with colour vision. Rods are very sensitive to light, and therefore support vision under dim light conditions.

There is a small area at the bottom of Figure 2.2(a), called the *fovea*. Here, in this small area, spanning less than  $2^\circ$  of the visual field, cones are extremely over-represented, while they are very sparsely distributed in the periphery of the retina. This has the result that we have full acuity only in this small area, roughly the size of your thumb nail at arm's distance. In order to see a selected object sharply, like a word in a text, we therefore have to move our eyes, so that the light from the word falls directly on the fovea. Only when we *foveate* it can we read the word. Foveal information is prioritized in processing due to the cortical magnification factor, which increases linearly with eccentricity, from about  $0.15^\circ/\text{mm}$  cortical matter at the fovea to  $1.5^\circ/\text{mm}$  at an eccentricity of  $20^\circ$  (Hubel & Wiesel, 1974). As a result, about 25% of visual cortex processes the central  $2.5^\circ$  of the visual scene (De Valois & De Valois, 1980).

For video-based measurement of eye movements, the pupil is very important. The other important, and less known, element on the eyeball is the *cornea*. The cornea covers the outside of the eye, and reflects light. The reflection that you can see in someone's eyes usually comes from the cornea. When tracking the eyes of participants, we mostly want only one reflection (although in some systems two or more reflections are used), so we record in infrared, to avoid all natural light reflections, and typically illuminate the eye with one (or more) infrared light source. The resulting *corneal reflection* is also known as 'glint' and the '1st Purkinje reflection' (P1). One should also be aware that light is reflected further back as well—both off the cornea and the lens—as illustrated in Figure 2.2(b). The corneal reflection is the brightest, but not the only reflection.

Human eye movements are controlled by three pairs of muscles, depicted in Figure 2.3. They are responsible for horizontal (*yaw*), vertical (*pitch*), and torsional (*roll*) eye movements, respectively, and hence control the three-dimensional orientation of the eye inside the head. According to Donders' law (Tweed & Vilis, 1990), the orientation uniquely decides the direction of gaze, independent of how the eye was previously orientated. Large parts of the brain are engaged in controlling these muscles so they direct the gaze to relevant locations in space.

The most reported event in eye-tracking data does not in fact relate to a movement, but to the state when the eye remains still over a period of time, for example when the eye temporarily stops at a word during reading. This is called a *fixation* and lasts anywhere from some tens

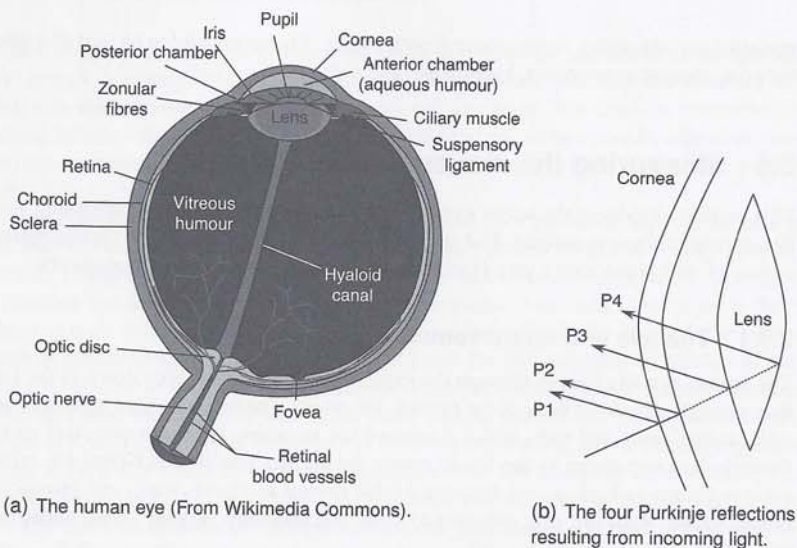


Fig. 2.2 For eye tracking, the important parts in the order encountered by incoming light are: the *cornea*, the *iris* and *pupil*, the *lens*, and the *fovea*.

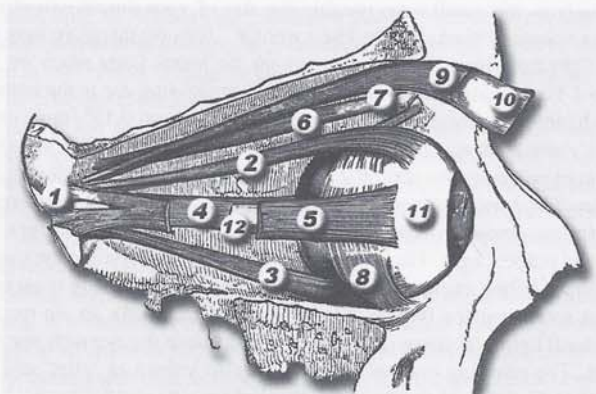


Fig. 2.3 The human eye muscles. From Grays (1918) *Anatomy of the Human Body* via Wikimedia Commons. The muscle pair (2)–(3) generate the vertical up-down movements, while (4)–(5) generate horizontal right-left movements. The pair (7)–(8) generate the torsional rotating movement. (9)–(10) control the eyelid.

of milliseconds up to several seconds. It is generally considered that when we measure a fixation, we also measure attention to that position, even though exceptions exist that separate the two.

The word 'fixation' is a bit misleading because the eye is not completely still, but has three distinct types of micro-movements: *tremor* (sometimes called *physiological nystagmus*), *microsaccades* and *drifts* (Martinez-Conde, Macknik, & Hubel, 2004). Tremor is a small movement of frequency around 90 Hz, whose exact role is unclear; it can be imprecise

**Table 2.3** Typical values of the most common types of eye movement events. Most eye-trackers can only record some of these.

Type	Duration (ms)	Amplitude	Velocity
Fixation	200–300	–	–
Saccade	30–80	4–20°	30–500°/s
Glissade	10–40	0.5–2°	20–140°/s
Smooth pursuit	–	–	10–30°/s
Microsaccade	10–30	10–40'	15–50°/s
Tremor	–	< 1'	20°/s (peak)
Drift	200–1000	1–60'	6–25°/s

muscle control. Drifts are slow movements taking the eye away from the centre of fixation, and the role of microsaccades is to quickly bring the eye back to its original position. These intra-fixational eye movements are mostly studied to understand human neurology.

The rapid motion of the eye from one fixation to another (from word to word in reading, for instance) is called a *saccade*. Saccades are very fast—the fastest movement the body can produce—typically taking 30–80 ms to complete, and it is considered safe to say that we are blind during most of the saccade. Saccades are also very often measured and reported upon. They rarely take the shortest path between two points, but can undergo one of several shapes and curvatures. A large portion of saccades do not stop directly at the intended target, but the eye ‘wobbles’ a little before coming to a stop. This post-saccadic movement is called a *glissade* in this book (p. 182).

If our eyes follow a bird across the sky, we make a slower movement called *smooth pursuit*. Saccades and smooth pursuit are completely different movements, driven by different parts of the brain. Smooth pursuit requires something to follow, while saccades can be made on a white wall or even in the dark, with no stimuli at all.

Typical values for the most common types of eye movements are given in Table 2.3. While these eye movements are the ones most researchers report on, especially in psychology, cognitive science, human factors, and neurology, there are several other ways for the eye to move, which we will meet later in the book.

Rather than mm on a computer screen, eye movements are often measured in *visual degrees* (°) or *minutes* (′), where  $1^\circ = 60'$ . Given the viewing distance  $d$  and the visual angle  $\theta$ , one can easily calculate how many units  $x$  the visual angle spans in stimulus space. The geometric relationships between these parameters are shown in Figure 2.4, and can be expressed as

$$\tan \frac{\theta}{2} = \frac{x}{d} \quad (2.1)$$

Note, however, that this relationship holds only when the gaze angle is small, i.e. when the stimulus is viewed in the central line of sight. For large gaze angles, the same visual angle  $\theta_1$  may result in different displacements ( $x_1$  and  $x_2$ ) on the stimulus, as illustrated in Figure 2.4.

If your stimulus is shown on a computer screen, you may want to use pixels units instead of e.g. mm. If  $M \times N$  mm denotes the physical size of a screen with resolution  $r_x \times r_y$  pixels, then 1 mm on the screen corresponds to  $r_x/M$  pixels horizontally and  $r_y/N$  pixels vertically.

When measuring eye-in-head movement, visual angle is the only real option to quantify eye movements, since the movements are not related to any points in stimulus space. Visual angle is also suitable for head-mounted eye tracking in an unconstrained environment, e.g. a supermarket, since the distance to the stimulus will change throughout the recording.

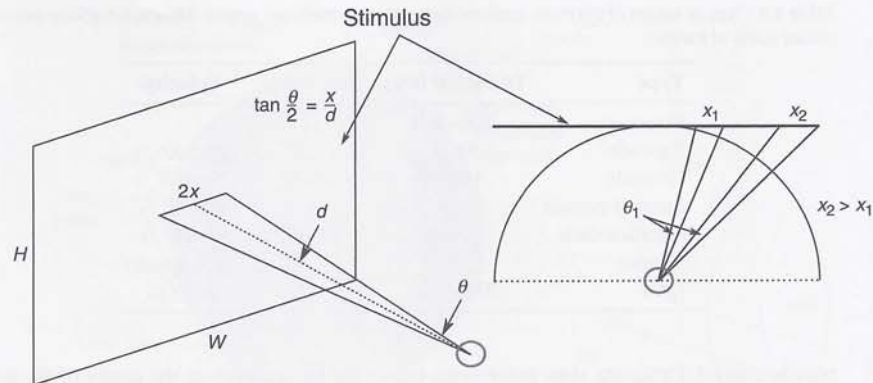


Fig. 2.4 Geometric relationship between stimulus unit  $x$  (e.g., pixels or mm) and degrees of visual angle  $\theta$ , given the viewing distance  $d$ . Notice that on a flat stimulus, the same visual angle ( $\theta_1$ ) gives two different displacements ( $x_1, x_2$ ).

### 2.5.2 Binocular properties of eye movements

An important aspect of human vision is that both eyes are used to explore the visual world. When using the types of movements defined in the previous section, the eyes sometimes move in relation to each other. *Vergence* eye movement refers to when the eyes move in directly opposite directions, i.e. converging or diverging. These opposite movements are important to avoid double vision (diplopia) when the foveated object moves in a three-dimensional space.

For most participants, both eyes look at the same position in the world. But many people have a dominant eye, and one which is more passive. If the difference is large, the passive eye may be directed in a different direction from that of the dominant one, and we say colloquially that the participant is *squinting*. The technical term is either *binocular disparity* or *disjucacy*.

In reading, disparity can be in the order of one letter space at the onset of a new fixation, and can occur in more than half of the fixations. Liversedge, White, Findlay, and Rayner (2006) found that disparity decreased somewhat over the period of the fixation, but not completely, and was of both crossed (right eye to the left of the average gaze position and vice versa) and uncrossed nature. In a non-reading 'natural' task, Cornell, MacDougall, Predebon, and Curthoys (2003) reported disparities of up to  $\pm 2^\circ$ , but also noticed that disparities of  $5^\circ$  were present (but rare) in the data. All these results were found for normal, healthy participants.

While it is commonly believed that the eyes move in temporal synchrony, binocular coordination varies over time. During the initial stage of a saccade, for example, the abducting eye (moving away from the nose) has been found to move faster and longer than the adducting eye (moving towards the nose) (Collewijn, Erkelens, & Steinman, 1988). At the end of the saccade this misalignment is corrected, both through immediate glissadic adjustments, and slower post-saccadic drift during the subsequent fixation (Kapoula, Robinson, & Hain, 1986).

### 2.5.3 Pupil and corneal reflection eye tracking

The dominating method for estimating the point of gaze—where someone looks on the stimulus—from an image of the eye is based on pupil and corneal reflection tracking (see Hansen & Ji, 2009, Hammoud, 2008, and Duchowski, 2007 for technical details and an



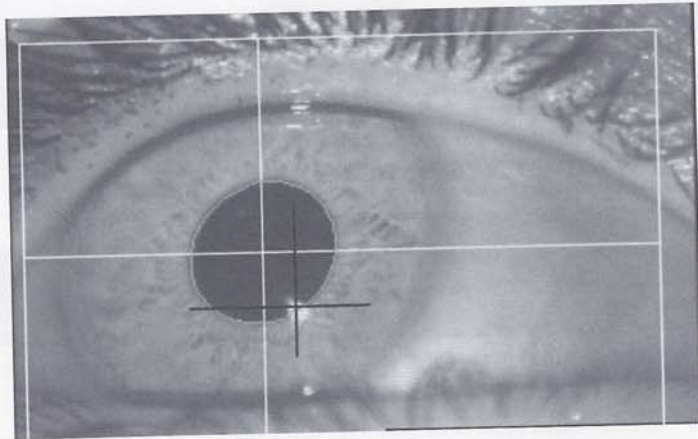


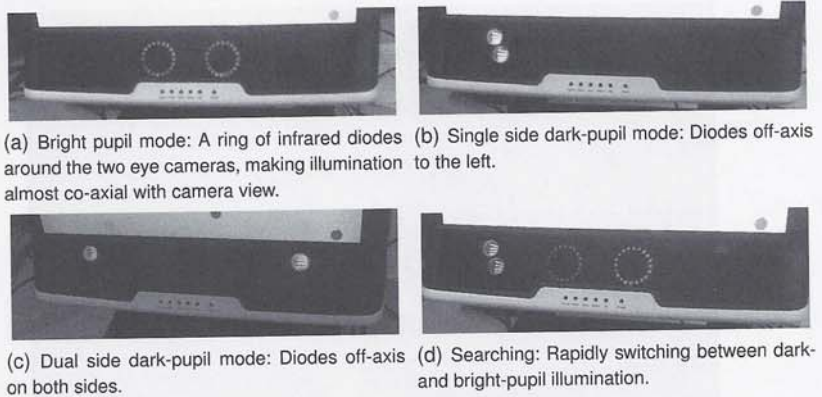
Fig. 2.5 A pupil–corneal reflection system has properly identified the pupil (white cross-hair) and corneal reflection (black cross-hair) in the video image of a participant's eye.

overview of other methods).

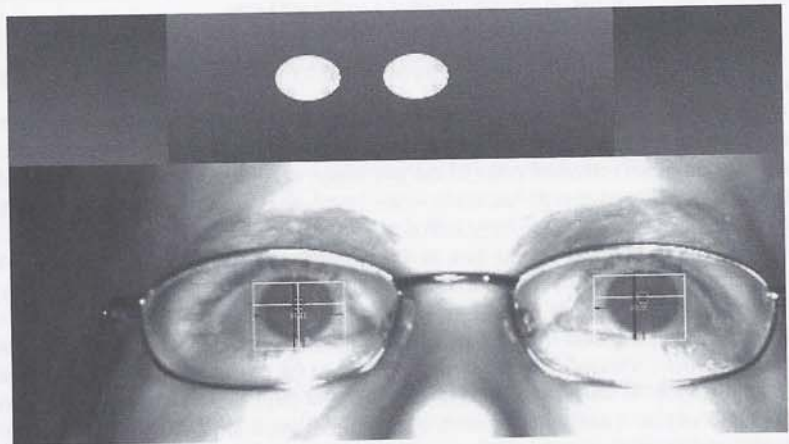
A picture of an eye with both pupil and corneal reflection correctly identified can be seen in Figure 2.5. While it is possible to use pupil-only tracking, the corneal reflection offers an additional reference point in the eye image needed to compensate for smaller head movements. This advantage has made video-based pupil and corneal reflection tracking the dominating method since the early 1990s.

The pupil can either appear dark in the eye image, which is the most common case, or bright, as with some ASL (Applied Science Laboratory), LC Technology, and Tobii systems. The bright pupil is bright because of infrared light reflected back from the retina, through the pupil. Such a system requires the infrared illumination to be co-axial with the view from the eye camera, which puts specific requirements on the position of cameras and illumination (Figure 2.6(a)). As long as the pupil is large, a bright-pupil system operates in approximately the same way as a dark-pupil system, but for small pupil sizes (when there is a lot of ambient light), a bright-pupil system may falter. The original motivation behind bright-pupil systems appears to have been to compensate for poor contrast sensitivity in the eye camera by increasing the difference in light emission between pupil and iris, but with new improved camera technology, contrast between iris and pupil is often also very good for most dark-pupil systems, as can be seen in Figure 2.5. At least one eye-tracker has been built that switches between bright and dark tracking mode, which requires the turning on and off of several infrared illuminators depending on how well tracking works in the current state (Figure 2.6). No studies have systematically investigated which of the two tracking modes gives better data quality over large populations, but in the authors' experience, data quality rather depends on the quality of the eye camera and other parts of the eye-tracker.

Noteworthy is that the methods used for image analysis and gaze estimation can vary significantly across different eye-trackers, both freely available and commercial. Therefore it may be difficult to compare systems between different manufacturers. To complicate the issue even further, some eye-tracking manufacturers keep many key technical details about the system secret from the user community. Sometimes the user is not allowed to see how the eye image is analysed, for instance, but only a very simplified representation of the position of the eyes is given. Figure 2.7 shows the eye image and the simultaneous simplified representation of the eyes, on a *remote system* (p. 51). If the recording software allows the operator to see



**Fig. 2.6** Four illumination states of the Tobii T120 dual mode remote eye-tracker. This particular eye-tracker changes to another tracking mode when tracking fails in the current mode.



**Fig. 2.7** Eye image in bottom half; and simplified representation of the eyes at the top.

the eye image, it is easier to set up the eye camera to ensure that tracking is optimal. Access to the eye image also makes it easier to anticipate and detect potential problems before and during data collection (pp. 116–134).

Figure 2.8 shows a schematic overview of a video-based eye-tracker, where the operations required to calculate where someone looks have been divided in three main blocks: *image acquisition*, *image analysis*, and *gaze estimation*.

In the acquisition step, an image of the eye is grabbed from the camera and sent for analysis. This can usually be done very quickly, but if head movement is allowed (as in remote eye-trackers), the first step of the analysis is to detect where the face and eyes are positioned in the image, whereafter image-processing algorithms segment the pupil and the corneal reflection from a zoomed-in portion of the eye. Geometrical calculations combined with a calibration procedure are finally used to map the positions of the pupil and corneal reflection to the data sample  $(x, y)$  on the stimulus.

While the pupil is a part of the eye, the corneal reflection is caused by an infrared light

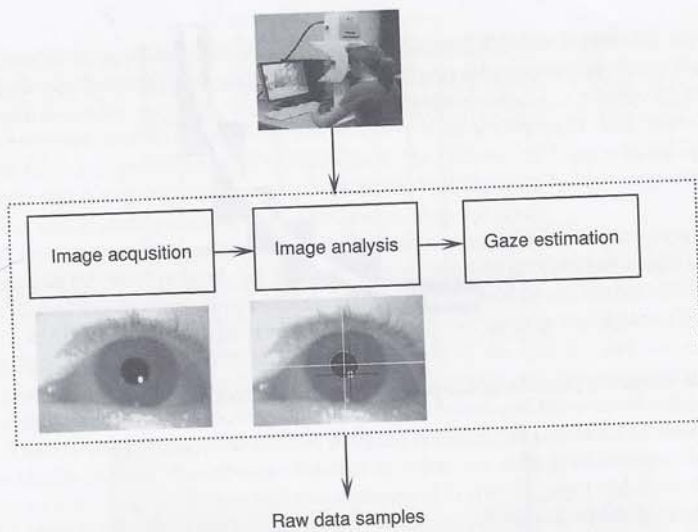


Fig. 2.8 Overview of a video based eye-tracking system.

source positioned in front of the viewer. The overall goal of image analysis is to robustly detect the pupil and the corneal reflection in order to calculate their geometric centres, which are used in further calculations. This is typically done using either *feature-based* or *model-based* approaches. The feature-based approach is the simplest where features in the eye image are detected by criteria decided automatically by an algorithm or subjectively by the experimenter. One such criterion is thresholding, which finds regions with similar pixel intensities in the eye image. Having access to a good eye image where the pupil (a dark oval) and the corneal reflection (smaller bright dot) are clearly distinguishable from the rest of the eye is important for thresholding approaches. Another feature-based approach looks for gradients (edges, contours) that outline regions in the eye image that resemble the target features, e.g. the pupil.

To increase the precision of the calculation of geometric centres, the algorithms typically include sub-pixel estimation of the contours outlining the detected features. The principal calculation is illustrated in Figure 2.9.

The major weakness of feature-based pupil–corneal reflection systems is that the calculation of pupil centre may be disturbed by a descending eyelid and downward pointing eye lashes. Lid occlusion of the pupil may cause—as we will see on pages 116–134 on camera set-up—*offsets* (incorrectly measured gaze positions) and increased *imprecision* in the data in some parts of the visual field. Figure 2.10 shows a participant with a drooping eyelid and downward eyelashes. The pupil is covered and cannot be identified, while the corneal reflection is dimly seen among the lashes.

A second weakness of feature-based systems concerns extreme gaze angles, at which the corneal reflection is often lost, but as we explain on pages 116–134, this can often be solved by moving the stimulus monitor, eye cameras, or infrared sources.

A third and mostly minor weakness is that the measured gaze position may be sensitive to variations in pupil dilation. In recordings where accuracy errors are not tolerated—as in the control systems for the lasers used in eye surgery—another technology called limbus-tracking is used. The limbus is the border between the iris and the sclera. It is insensitive to variations

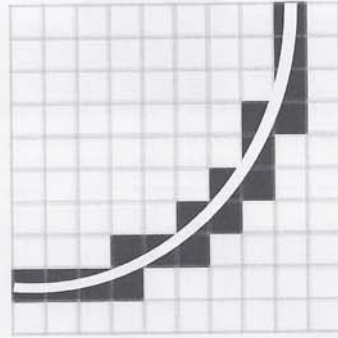


Fig. 2.9 Increasing precision by sub-pixel estimation of contours.

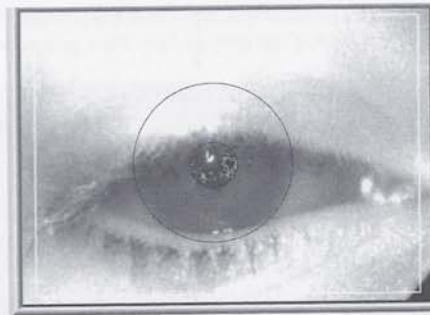


Fig. 2.10 In model-based eye tracking, the recording computer uses a model of the eye to calculate the correct position of the iris and the pupil in the eye video, even if parts of them are occluded by the eyelids. Rings indicate features of the model.

in pupil dilation, but very sensitive to eyelid closure, and is therefore fairly impractical except in specific applications such as laser eye surgery.

Model-based solutions can alleviate the weakness of feature-based pupil–corneal reflection systems by using a model of the eye that fits onto the eye image using pattern matching techniques (Hammoud, 2008). A useful eye model would assume that both iris and pupil are roughly ellipsoidal and that the pupil is in the middle of the iris. For example, using an ellipsoidal fit of the pupil would prevent the calculated centre of the pupil moving downward when the upper eyelid occludes the top part of the pupil, something that happens in the beginning and at the end of each blink, or when the participant gets drowsy.

Another model assumption is that the pupil is darker than the iris, which is darker than the sclera. When the model knows this, it can position the iris and pupil circles onto the most probable position in the image.

Moriyama, Xiao, Cohn, and Kanade (2006) implemented and tested a model-based iris detector that could be used for eye tracking. Although it does solve many eyelid problems known from feature-based pupil–corneal reflection systems, the model-based eye-tracker still sometimes misplaces the iris due to shadows in the eye image. While having the potential to provide more accurate and robust estimations of where the pupil and corneal reflection centres are located, model-based approaches add significantly to the computational complexity since they need to search for parts of the eye image that best fit the model. Without a good initial

guess of where the pupil and corneal reflections are located, the time it takes the algorithm to find the indented features (often called *recovery time*) may be unacceptably long. Fortunately, a full search is needed only in the first frame, since feature positions in subsequent frames can be predicted from previous ones. Note that recovery time is very individual, since the algorithm will find some participants' eyes much faster than others. As feature-based approaches can provide this first guess, eye-tracking approaches that combine feature- and model-based approaches will probably become even more common in the future.

Now assume that the centres of both pupil and corneal reflection have been correctly identified, that the head is fixed and that we have a complete geometrical model of the eye, the camera, and the viewing set-up; then the gaze position could be calculated mathematically (Guestrin & Eizenman, 2006). However, this is usually not done in real systems, mainly due to the difficulty in obtaining robust geometric models of the eye. Instead, the majority of current systems use the fact that the relative positions of pupil and corneal reflections change systematically when the eye moves: the pupil moves faster, and the corneal reflection more slowly. The eye-tracker reads the relative distance between the two and calculates the gaze position on the basis of this relation. For this to work, we must give the eye-tracker some examples of how points in our tracked area correspond to specific pupil and corneal reflection relations. We tell this to the eye-tracker by performing a *calibration*, which typically consists of 5, 9, or 13 points presented in the stimulus space that are fixated and sampled one at the time. The practical details of calibration are described on pages 128–134.

While using one camera and one infrared source works quite well as long as the head is fairly still, more cameras and infrared sources can be used to relax the constraints on head movement and calibration. Using two infrared sources gives another reference point in the eye image, and is in theory the simplest system that allows for free head movement, which is desirable in remote eye-trackers. Using multiple cameras and infrared sources, it is theoretically possible to use only one point to calibrate the system (Guestrin & Eizenman, 2006). However, using another light source complicates the mathematical calculations.

The most common commercially available eye-trackers are those with one or two cameras and one or multiple infrared light sources that work best with 5, 9, and 13 point calibrations. More on different types of eye-camera set-ups can be found on pages 51–64.

## 2.6 Data quality

Data quality is a property of the sequence of raw data samples produced by the eye-tracker. It results from the combined effects of eye-tracker specific properties such as sampling frequency, latency, and precision, and participant-specific properties such as glasses, mascara, and any inconsistencies during calibration and in the filters during and after recording.

Some eye-trackers also output pupil and corneal reflection positions, calculated directly from the eye image prior to gaze estimation. We will not talk about the quality of this position data as they are not used very often; however, the reader should observe that they have properties common to the sequence of raw data samples.

Data quality is of the utmost importance, as it may undermine or completely reverse results. Already McConkie (1981) argues that every published research article should list measured values for the quality of the data used, but this has not yet become the standard.

### 2.6.1 Sampling frequency: what speed do you need?

The sampling frequency is one of the most highlighted properties of eye-trackers by manufacturers, and there is a certain competition in having the fastest system. You do need some

speed in your system to be able to calculate certain eye-tracking measures, but high-speed eye-trackers are typically more expensive, more restrictive for the participants, and also produce larger data files.

Sampling frequency is measured in *hertz* (Hz), a unit we will frequently use throughout the book. A 50 Hz eye-tracker records the gaze direction of participants 50 times per second. This may sound like often enough, but a 50 Hz eye-tracker is generally considered a slow system.

So what speed do you need in your eye-tracker? For oscillating eye movements, we can use the Nyquist–Shannon sampling theorem to argue that a sampling frequency at least twice as large as the recorded movement is needed. For instance, for a tremor at 150 Hz, the necessary minimal sampling rate for detection is 300 Hz. Generally, the required sampling frequency depends on what you need to detect or measure, and how precisely you want to measure it. The faster the eye movement, the faster your system has to be.

However, most statements on required sampling frequency are not based on scientific or mathematical investigation. For instance, the border between low speed and high speed is often considered to be around 250 Hz. Why 250 Hz? There are only a few, little-cited studies to support this, and they all deal with the calculation of saccadic peak velocity. Nevertheless, it has gradually become accepted that the statistical effect sizes most studies report would be undermined by using a system which records at a frequency less than 250 Hz. But does this mean that data recorded with systems slower than 250 Hz will suffer from a temporal inaccuracy that renders statistical conclusions invalid? This depends on your outcome measure/dependent variable and your desired effect size. In the literature on reading research, differences below 20 milliseconds in fixation durations are claimed; at the limit of your system's capabilities one might reasonably question the validity and replicability of these effects.

Existing eye-trackers cover a spectrum of sampling frequencies from a few Hz up to more than 1000 Hz.

**25–30 Hz** These are the slowest systems sold, and typically record data only as gaze-overlaid videos (p. 61). The sampling rates 25 and 30 Hz (or more precisely 29.97 Hz) originate respectively from the European television PAL-standard and the NTSC-standard used in the United States. Only web-cam based eye-trackers are slower.

**50–60 Hz** Many remote systems and head-mounted eye-trackers run at this speed, because it was the most common frequency in camera technology for a long time.

**120 Hz** This range of sampling frequencies gradually became more common from around 2007.

**250 Hz** The low end of the higher speed systems, set here because this was the speed of the 1990s eye-tracker SMI EyeLink I, running at 250 Hz.<sup>5</sup>

**500 Hz** Midsection of sampling frequencies that was reached by pupil-corneal reflection eye-trackers around the year 2000. Not many manufacturers provide this speed, and those that do typically offer eye-trackers that are *tower-mounted contact systems* (defined on p. 51).

**1000–2000 Hz** The highest sampling frequencies available in 2010. Before these high-speed video-based eye-trackers arrived around 2006, only coil-based and dual Purkinje systems had this speed.

The higher the speed of the eye-tracker, the more infrared illumination is needed, since each eye camera sample is collected for a shorter interval (like camera shutter speed and ISO in traditional photography). Also notice that sampling may be different between the two eyes. Some dual camera systems such as the remote eye-tracker EyeFollower from LC

<sup>5</sup>Up until August 2001, SMI manufactured and sold the EyeLink eye-trackers.

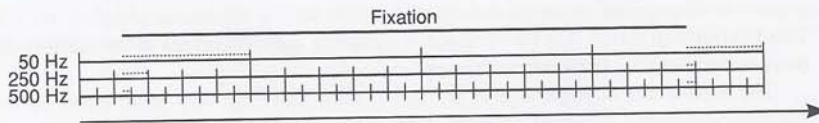


Fig. 2.11 A hypothetical fixation recorded at sampling frequencies 50, 250, and 500 Hz. At each small peg along the time line, the eye camera photographs the eye, and a gaze position is calculated, and we have a sample. As the fixation starts and ends anywhere between samples, and is not recorded until the subsequent sample, we will have errors on the calculated fixation duration at both start and end. Errors—indicated with dashed lines—will be larger for slower systems, but when averaging over many durations, these errors to a large extent become equal (Andersson, Nyström, & Holmqvist, 2010).

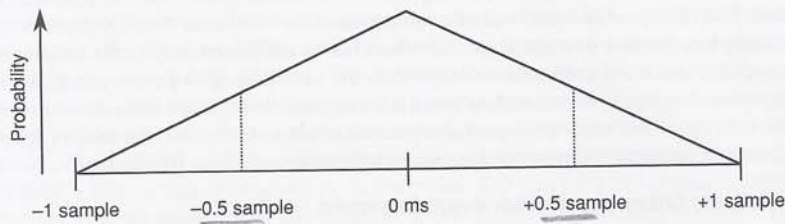


Fig. 2.12 The probability function for fixation duration measurement error at a given sampling frequency.

Technologies are particular in that they can output 120 Hz data, but every sample is taken alternately from the opposite eye, requiring only 60 Hz sampling from cameras.

### Samples, and what happens between them

When a *single* photo is taken of the eye and processed by gaze estimation, it results in a *sample*. For a 50 Hz system, there are 50 samples recorded per-second. Each sample is ideally momentous; but even in real 50 Hz systems, it is not 20 ms long. It is the intermediary window of no sampling that is 20 ms long. Figure 2.11 illustrates samples and the intermediary windows of no sampling, along a time line for three different sampling frequencies.

A fixation that we recognize in a sample can have started anytime between the previous sample (when we saw there was no saccade) and the current sample. If you have a 50 Hz system, you have 20 ms between samples, so the saccade can start anywhere within that window of no sampling (of 20 ms). If instead you have a 500 Hz eye-tracker, the window of no sampling is only 2 ms. This means that with a higher sampling frequency, we can more precisely measure the start and end ('on-' and 'offset') of saccades, fixations, and other events of the eye.

### Fixation durations and other event durations

Given a low-speed eye-tracker, can we reliably measure, e.g. fixation durations? Andersson *et al.* (2010) quantified the effect of sampling frequency on event durations such as fixation durations in a series of simulations and tests on real data. Since the probability of error in sampled event durations follows the distribution shown in Figure 2.12, the central limit theorem can be used to deduce the relationship between sampling frequency ( $f_s$ ), number of data points (e.g. fixation durations) ( $n$ ) and the resulting error,

$$\bar{\epsilon} \sim N\left(0, \frac{1}{18nf_s^2}\right) \quad (2.2)$$

This relationship shows that the variance in sampling error decreases as the number of fixations or the sampling frequency increases.

The simulations in Andersson *et al.* (2010) show that even very small effect differences can also be detected at slower sampling frequencies if the number of recorded data points is large enough. There is a quadratic relation between the sampling frequency and the number of fixations needed. This means that, for example, if you are choosing between a 50 and a 250 Hz system, if all else is equal, you will need 25 times  $((\frac{250}{50})^2)$  as many data points with the 50 Hz system to achieve the same low variance in the average as with the 250 Hz system. In other words, for most event durations, it is always possible to compensate for the effect of a low sampling frequency with more data. Equation (2.2) is generally true for all eye-tracking duration measures with a sampled onset and a sampled offset. In practice, however, not all sampling-related errors can be compensated for with more data. Event-detection algorithms, like fixation detection, may introduce biases and uncertainty in the estimation of the durations, and these error may be larger than the sampling-related error calculated from the sampling frequency. Also, in many cases it is not possible to record more data to compensate for this, e.g. if you work with special populations such as babies and animals, or only get one chance to correctly estimate the duration of a fixation, as in gaze interfacing.

### Saccadic latency and other event latencies

Andersson *et al.* (2010) distinguish between event durations, which were described in the previous section, and event latencies. Latencies consist only of a single sampled onset or offset, and the other point is given by a sampling-independent point in time, such as the start of a trial. The fact that latencies only consist of a single sampled point has the effect that the temporal error caused by the limited sampling frequency does not even out, even given large amounts of data. Fortunately, this error is predictable given enough data, and has the expected mean of half a sample of time. That is, if the sampling frequency is 50 Hz, then the expected error is  $\frac{1000}{50} = 20$  ms. This error is either an overestimation or an underestimation, depending on whether the sampled point is the onset or the offset of the event. For a latency where the latency is counted from the start of the trial, a sampling-independent point, then the latency is over-estimated. For a measure where the event offset is given, e.g. by a trial end, then the latency is underestimated.

### Saccadic velocity and acceleration

A high sampling frequency is required to accurately capture fast eye movements such as saccades. In reading, saccades are around 30–40 ms, which means the motion will be registered only by 1–2 samples with a 50 Hz system.

The smaller the saccades, the higher the required sampling frequency. For instance, Enright (1998) suggests that saccadic peak velocity can be estimated using 60 Hz pupil–corneal reflection eye-trackers, but only if the saccades are larger than  $10^\circ$ . If the saccades are shorter than  $10^\circ$ , e.g. saccades from reading research, then the peak velocity calculation will not be accurate from 60 Hz systems. Juhola, Jäntti, and Pyykkö (1985), who used EOG- and photoelectric eye-tracking systems to study  $20^\circ$  saccades, provide evidence that sampling frequency should be higher than 300 Hz in order to reliably calculate the maximum saccadic velocity. Inchingolo and Spanio (1985), using a 200 Hz EOG system, found that saccadic duration and velocity data were equivalent to data recorded with a 1000 Hz system, albeit only if the saccades are larger than  $5^\circ$ . The lowest bid is given by Wierds, Janssen, and Kingma (2008), who argue that a 50 Hz eye-tracker can be accurately used to measure peak velocities as long as saccades are at least  $5^\circ$ .

Acceleration values are even more sensitive to sampling frequency than velocity. A 50 Hz eye-tracker cannot provide accurate peak acceleration/deceleration values (Wierds *et al.*, 2008).



**Table 2.4** Sampling frequencies and the number of microsaccade studies having used them according to the overview in Martinez-Conde *et al.* (2009).

Sampling frequency	Number of studies
<200	1
200–300	9
500	25
1000	2

### Microsaccades

As the average microsaccade has the same dynamic characteristics as a saccade, only around a factor 50 smaller (Engbert, 2006), correctly measuring it requires a higher sampling frequency than does measuring saccades. The durations of microsaccades are only somewhat smaller than saccadic durations, however, which means that the requirement on sampling frequency would only slightly exceed that for saccade measurements. There appear to be no systematic investigations of these requirements, but in practice eye-trackers used in microsaccade research have sampling frequencies no lower than 200 Hz. Table 2.4 presents sampling frequencies from microsaccade studies in the overview by Martinez-Conde, Macknik, Troncoso, and Hubel (2009).

### Gaze contingency

In gaze-contingent experiments, the stimulus display changes online in relation to how the eyes move (p. 49). The stimulus is typically manipulated during a saccade when visual intake is significantly impaired. To maximize the time available for such a computationally costly stimulus manipulation, it is important to quickly and accurately detect the offset and, in particular, the onset of a saccade. In this case a high sampling frequency is very desirable.

### 2.6.2 Accuracy and precision

While the *accuracy* of an eye-tracker is the (average) difference between the true gaze position and the recorded gaze position, *precision* is defined as the ability of the eye-tracker to reliably reproduce a measurement. The difference between accuracy and precision is illustrated in Figure 2.13. Obviously, a good eye-tracker should have both high accuracy and high precision. Beware that these two properties of eye-trackers are often confused.

This section only deals with *spatial* precision. There is also *temporal* precision, which we describe on page 43.

Other much-used eye-tracking concepts that draw on the definition of accuracy and precision include:

**Offset** Formally, angular distance between calculated fixation location and the location of the intended fixation target, i.e. an operational definition of accuracy. Informally, an acceptable precision in combination with a poor accuracy, exemplified in the left part of Figure 2.13.

**Drift** A gradually increasing offset, common in older eye-trackers.

**System-inherent noise** Best possible precision you can get with a given eye-tracker, also known as *spatial resolution*. This is typically measured with artificial eyes, which are absolutely still, so we know for certain that spatial variance comes from the eye-tracker itself.

**Oculomotor noise** Traditionally refers to the fixational eye movements tremor, microsaccades, and drift, even though microsaccades have been linked to cognitive functions (Martinez-

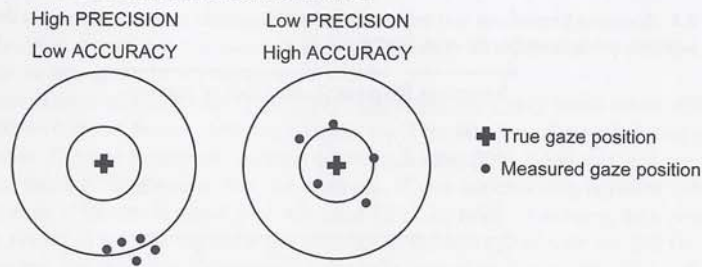


Fig. 2.13 Precision and accuracy.

Conde *et al.*, 2009). Oculomotor noise is often called *jitter* (Martinez-Conde *et al.*, 2004; Jacob, 1991).

**Typical precision** The average precision as measured from a large participant population with a wide spectrum of eye physiologies and iris colours. The typical precision is the precision that can be expected of a system in a standard eye-tracking experiment.

**Optic artefacts** False, i.e. physiologically impossible, high-speed movements, often with a much larger amplitude than other types of noise, and caused by interplay between the optical situation (such as glasses, contact lenses, additional reflections and shadows, and varying ambient light conditions) and the gaze estimation algorithm.

**Environmental noise** Variation in the gaze position signal caused by external mechanical and electromagnetic disturbances in the recording environment.

**Noise (general)** The combination of system-inherent, oculomotor, and environmental noise and sometimes also including optic artefacts.

*Resolution* is a common term related to precision, usually referring to 'the smallest movement that can be detected'. It is sometimes defined as the standard deviation of the pupil position in the eye video, but to the vast majority of users, this is less relevant than precision in gaze position.

Both accuracy and precision can—after some training—be evaluated in data visualizations such as scanpath plots with raw data, and space-time diagrams where data samples are plotted against time.

### Precision and how to measure it

Spatial precision is one of the most important technical properties of an eye-tracker. Precision is important for everyone who wants to calculate fixation or saccade measures, and perhaps unexpectedly even for your heat map visualizations. Also, if you need to measure the very small fixational eye movements known as tremor, drift, and microsaccades, you need very high precision. While precision is vital for such measurements, accuracy is not of critical importance. Several types of gaze-contingent studies require both a very high precision and a high accuracy.

Precision should be calculated from data samples recorded when the eye is fixating on a stationary target, such that sample variation originating from eye movement is excluded as large an extent as possible. The only way to completely disregard eye movement is to use an artificial eye positioned in front of an eye-tracker. There are two common ways to calculate precision: standard deviation of data samples and root mean square (RMS) of inter-sample distances.

The most straightforward way to calculate precision for  $n$  data samples is perhaps to use the standard deviation,

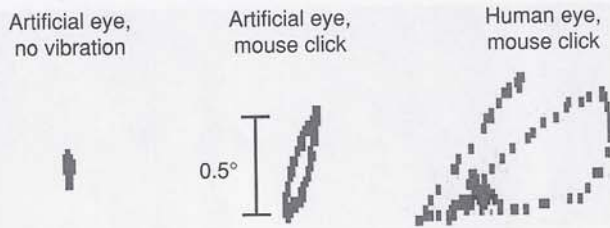


Fig. 2.14 Raw data samples in (x,y)-space. The recording with an artificial eye on this 250 Hz remote eye-tracker has an RMS of  $0.02^\circ$  when everything is still, but clicking a mouse on the same table where the eye-tracker is standing causes vibrations in the data large enough to be mistaken for small saccades. The RMS will remain small across all these low-frequency vibrations, but standard deviation will respond to the movement.

$$s_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.3)$$

This can be computed separately in the horizontal and vertical dimensions, and measures how dispersed samples are from their mean value.

Precision can also be calculated using angular distances  $\theta_i$  (in degrees of visual angle) between successive data samples  $((x_i, y_i)$  to  $(x_{i+1}, y_{i+1}))$ . It is then typically computed as the root mean square (RMS) of such distances,

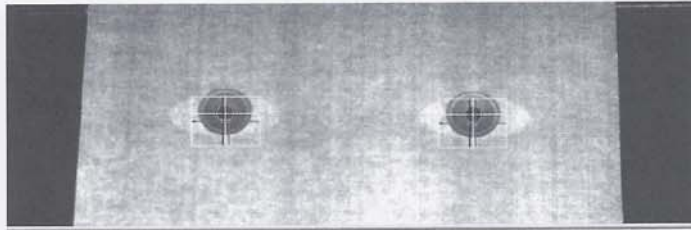
$$\theta_{\text{RMS}} = \sqrt{\frac{1}{n} \sum_{i=1}^n \theta_i^2} = \sqrt{\frac{\theta_1^2 + \theta_2^2 + \dots + \theta_n^2}{n}} \quad (2.4)$$

Note that precision calculated from data samples directly and that calculated from inter-sample distances capture variation in the eye-movement signal in slightly different ways; since inter-sample distances only compare temporally adjacent samples, they are less sensitive to a large overall spatial dispersion of the data. Figure 2.14 gives an example where standard deviation and RMS will differ significantly in the rightmost measurement, but not in the measurement to the left.  $\theta_{\text{RMS}}$  seems to be the choice of most eye-tracking manufacturers and practitioners to quantify precision, and it is also the measure we will use in this book.

Poorer eye-trackers have RMS values up to  $1^\circ$ , while manufacturers of high-end eye-trackers typically report a precision that is better than  $0.10^\circ$ , although precisions down to  $0.01^\circ$  are sometimes reported.<sup>6</sup> A  $0.01^\circ$  RMS would mean that the average sample to sample movement due to noise in the eye-tracker is around  $0.0001^\circ$ , far below the amplitude of microsaccades, even below the level of the oculomotor noise originating from eye muscles. Not all manufacturers report precision, and only some calculate the reported value using an artificial eye, which exhibits no physical movements at all, so that all measured movements can be attributed to the system-inherent noise of the eye-tracker.

For microsaccade and gaze-contingent studies, a rule of thumb is that the eye-tracker should have an RMS lower than about  $0.03^\circ$ . For very accurate calculation of fixation durations and saccadic measures, see to it that your eye-tracker has a RMS lower than around  $0.05^\circ$ . Any further increase in RMS always introduces noise in your measures of all these events.

<sup>6</sup>For instance, SR Research (2007) and [www.smivision.de](http://www.smivision.de)



(a) Artificial eyes as seen from the scene camera of the SMI RED X remote.



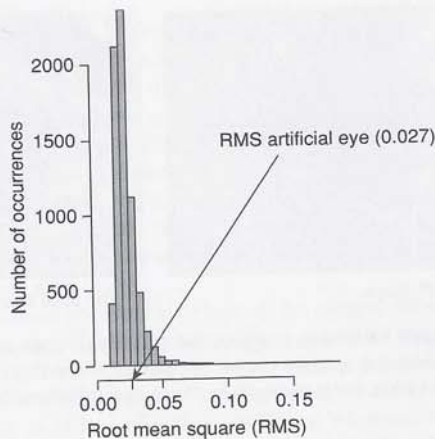
(b) Set-up of artificial eyes in precision test of SMI RED X remote eye-tracker. The eyes are on the grey patch on the black computer to the left. Make sure the eyes are properly attached and do not vibrate.  
 (c) Set-up of an artificial eye in a precision test of SMI HED X.

**Fig. 2.15** Artificial eyes used for testing the (spatial) precision of eye-trackers.

Figure 2.15 shows how we use a pair of artificial eyes to test the precision of two SMI systems. The procedure is simple: first of all, calibrate on a human so that you can get coordinate data.<sup>7</sup> Then put one or a pair of artificial eyes where the human eye(s) would have been, and make sure the artificial eyes are properly attached (Figure 2.15(b)). Beware of vibration movements from the environment, which should not be part of your precision measurement. See to it that the gaze position of the artificial eye(s) is somewhere in the middle of the calibration area, and then start the recording. Export the raw data samples, use trigonometry and the eye-monitor distance with the physical size and resolution of the monitor to calculate sample-to-sample movement in visual degrees. Then select a few hundred samples or more where the gaze position appears to be still, and calculate the RMS from these samples.

Testing only with an artificial eye may be misleading, however. The artificial eyes do not have the same iris, pupil, and corneal reflection features as human eyes, and may be easier or more difficult for the image analysis algorithms to process. Also, in actual eye-tracking research, real eyes tend to exhibit a large variation in image features that cannot be simulated with artificial eyes. Therefore some manufacturers complement the artificial eye test with a precision test on a human population with a large variation in eye colour, glasses, and contact lenses, as well as ethnic background, having them fixate several positions across the stimulus monitor. The distribution of precision values from such a test, exemplified in Figure 2.16, is an important indicator of what precision you can expect in actual recordings, and its average defines the *typical precision*. Its drawback is that this data includes oculomotor noise, and

<sup>7</sup>Calibration on a human *may* introduce some small noise, so if you have a system where you can get data without first calibrating, you may do that, but be aware that the RMS will not be comparable to systems that require calibration before data recording.



**Fig. 2.16** Histogram over RMS values calculated from 165 people looking at 13 validation points just after calibration and again after 10–20 minutes of reading, using a high-end tower-mounted eye-tracker. The arrow pointing to  $0.027^\circ$  indicates the RMS value calculated from data recorded with an artificial eye. This indicates that the eye-tracker is optimized for human eyes rather than for our artificial eye.

therefore both human and artificial eyes are needed.

Also note that different artificial eyes give slightly different RMS values for the same eye-tracker. We found RMS values of  $0.021^\circ$  and  $0.032^\circ$  on the same eye-tracker when using two different artificial eyes. Also different specimens of the same eye-tracker may exhibit different RMS values, which would indicate a difference in build quality.

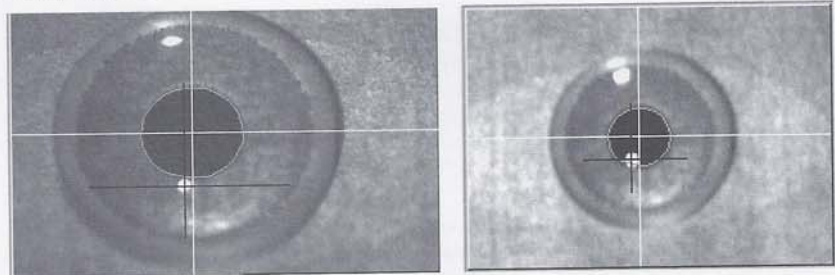
It is easy to do the test on human eyes yourself. Calibrate the system, and record the human as he is staring at a single point as steadily and for as long as possible. Export the raw data, and use samples in the beginning of the long fixation, blinks excluded, to calculate the RMS value. In our tests, an artificial eye often gave on average data 2–10 times as precise as a real human eye during fixation, but exceptions such as in Figure 2.16 were also found.

### Factors that influence precision

Precision is influenced by the eye-tracking hardware and software, participant-specific properties, and the recording environment. In the following section, we will discuss the influence of these factors in relation to our own observations from measuring RMS values from almost 20 different eye-trackers from several manufacturers, using both artificial eyes and real eyes<sup>8</sup> making prolonged fixations. We only made 1–5 measurements per eye-tracker and setting/condition, but found a strong consistency in RMS values across similar recordings. These values are presented so as to explain the properties of the hardware, and should not be seen as an absolute property of a specific eye-tracker or a specific manufacturer. Precision varies with a number of factors, so these values may deviate somewhat from the ones in your recording.

It is possible to improve precision by modifying or adjusting the hardware components of the eye-tracker. Precision is for example closely related to the *resolution of the eye camera*. In particular, the camera should view the pupil and corneal reflection with many pixels, in high quality, and be able to robustly segment them from their backgrounds in order to have a high precision output (see Figure 2.17). As the number of pixels spanned by the eye image

<sup>8</sup>The same person (one of the authors), with no mascara, downward eye lashes, glasses or contact lenses, and good lighting conditions.



(a) The pupil spans 77 pixels.

(b) The pupil spans 39 pixels.

**Fig. 2.17** The measured RMS value in eye-tracker (a) is 5–10 times better than the RMS value for eye-tracker (b). In general, the resolution of the eye camera is one major factor behind precision. The more pixels used for the pupil, and in particular for the corneal reflection, the better.

decreases as the camera moves further away from the eye, long-range eye-trackers such as those used in fMRI studies typically have lower precision than tower-mounted eye-trackers that film the eye from a shorter distance. One way to artificially increase the size of the corneal reflection is to *reduce the focus of the eye camera*. This should theoretically increase precision, but the measurements we have made indicate the opposite. It is unclear why. The free head motion in remote eye-trackers is often combined with *autofocusing eye cameras*, i.e. the focus of the eye video changes back and forth, which is very likely to give a precision that varies with time. High-end eye-trackers sometimes allow for binocular recording with a single camera by zooming out to film both eyes simultaneously. According to the above argumentation, this should reduce the precision.

Eye cameras differ both within and between manufacturers, and the competitive market sometimes forces many to use less expensive cameras, which may have difficulties clearly distinguishing iris and pupil for a large spectrum of participants with differing pigmentation. For instance, Kammerer (2009) showed that the eye camera of a common remote eye-tracker gives significantly poorer data quality for participants with bright-coloured irises (and that glasses give poorer data than contact lenses). Another reason why low-quality eye cameras increase noise levels in data is the slower pixel updating, which makes pixels retain some of the brightness of the passing corneal reflection, leaving a bright trace behind the real reflection (Droege & Paulus, 2009). A camera with a high-quality sensor simply gives better data (when all else is equal).

When manufacturers choose what eye cameras to put in their eye-tracker, there is often a tradeoff between precision and sampling frequency. Since sampling frequency is currently the most pronounced sales argument, it is often prioritized over precision. This appears to be the case both for remote eye-trackers and tower-mounted systems. Furthermore, for a longer sampling time, that is, if the eye camera is open for a longer period for each sample, the movement of the eye during that period smears the sampled eye image as a natural lowpass filter, which may additionally increase precision for lower sampling frequencies.

There are several software-related factors that influence precision. Simultaneous binocular and separate recordings of two artificial eyes on those remote eye-trackers that allowed it gave on average an 8-fold increase in RMS, compared to calculating a single average from the two eyes. If you have a system with a higher than necessary sampling frequency, it is possible to trade a lower sampling frequency against higher precision. For instance, replace every four samples in a 2000 Hz recording with a single average of those four, and you will have a more precise 500 Hz recording.

Precision can be increased by *filtering*, and many manufacturers indeed filter the data in the system they sell. Turning off the default filter in one of the most precise remote systems increases the low RMS  $0.01^\circ$  value to  $0.03^\circ$ . Check what filters there are and note that some filters improve precision at the cost of an increased latency.

In practice, the eye image is processed so that the border between the pupil and the iris is not represented by the pixels we see in Figure 2.17 but by the *sub-pixel estimation* exemplified in Figure 2.9. A sub-pixel estimation of the border curve responds more smoothly to very small movements of the eye. The sample-to-sample motion in the data stream will be smaller, and the eye-tracking software delivers better precision levels than a pixel-based border calculation would give.

Another way to increase precision is to turn off the corneal reflection and make a *pupil-only* recording, which is possible with some of the systems we tested. The increase in precision when recording without corneal reflection is usually dramatic, but it comes at a cost of poorer accuracy and lower tolerance to head movements, and is therefore best suited for studies where accuracy is of less relevance (studies of fixational eye movements such as microsaccades, for instance), and for systems that have a fixed head-camera position (head- and tower-mounted contact systems).

When the head is allowed to move freely relative to the eye camera, as in remote eye tracking, an additional layer of *head position calculations* is added to the gaze estimation algorithm. Consequently, precision drops (Kolakowski & Pelz, 2006). The precise calculation and update frequency for head position, seen as eye position in the eye camera, varies with manufacturers but is not generally revealed.<sup>9</sup> We know that many of the possible technical solutions decrease precision, however. It is quite possible that remote eye-tracking can only give a high precision if the positions of the head and eye are *measured* using optical reflectors or magnetic sensors, but few remote systems have been produced that add sensors to the participants.

*Bite bars* lock the distance and position between eye and camera. This can be particularly useful when studying very small eye movements, such as microsaccades, but bite bars are hardly ever used today. For the remote system with the poorest precision, the RMS value decreased after we put the participant in a head support system.

The poorest RMS value we noted in our investigation was  $1.03^\circ$ , measured on a human eye in one of the most popular 50 Hz remote eye-trackers. The best precision we found was  $0.0012^\circ$ , using an artificial eye in a high-end tower-mounted system from one of the manufacturers with an academic user group. In comparison, the Dual-Purkinje eye-trackers were reported to have precision of about  $0.005^\circ$  (Crane & Steele, 1985). Table 2.5 lists the range of RMS values found when measuring precision on different systems with both human and artificial eyes.

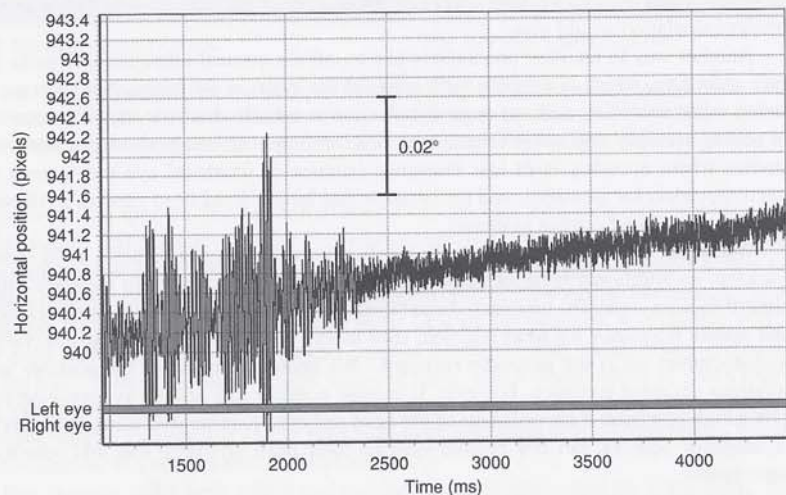
While tower-mounted and head-mounted eye-trackers in our test scored very well, remote eye-trackers were more variable. Also, the *exact position* of the participant affects precision in remote eye-tracking. In one of these, we got RMS values as low as  $0.01^\circ$  when we positioned artificial eyes so as to minimize the online cursor movement in the recording software. When instead we positioned the artificial eyes without this position adjustment, RMS increased to  $0.08^\circ$ . Figure 2.35 (p. 59) shows how precision varies with distance from the monitor for another commercial remote eye-tracker.

Note that the manufacturer value reported—irrespective of brand of eye-tracker—may very well be for an eye position that gives the lowest precision value possible. Real partici-

<sup>9</sup>If you fear that change in pupil size is part of the calculation of head position (which would degrade data quality) and want to test this, put a participant in a chinrest at a fixed distance and vary light conditions so that the pupil changes, and see if the eye-tracker reports distance changes.

**Table 2.5** Ranges of RMS values in the three classes of eye-trackers (described on page 51), according to our measurements. Included eye-trackers were manufactured by Tobii, SMI, SR Research, ASL, and SmartEye.

Set-up	Artificial eye	Human eye
Remote	0.0100°–0.3060°	0.0300°–1.0300°
Head-mounted	0.0013°–0.0067°	n/a
Tower-mounted	0.0012°–0.0300°	0.0100°–0.0500°



**Fig. 2.18** Vibration noise from walking past the table where the eye-tracker is placed, followed by system-inherent noise. 1 vertical unit corresponds to less than  $0.02^\circ$  of visual angle. Raw x coordinates over time as recorded with a high-end eye-tracker on an artificial eye.

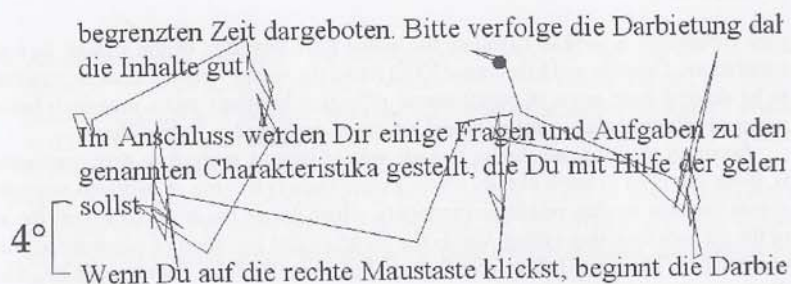
pants may sit very differently, and move around. Therefore, place your artificial eyes at different positions in front of the eye-tracker, so you can measure the full spectrum of RMS values throughout the region known as the 'headbox' (p. 58).

Variation in the amount of light that hits the eye changing the contrast in the eye video is a likely factor for precision decrease.

Noise may also result from movements in the immediate vicinity of the eye-tracker. Figure 2.18 shows data from a pupil-only recording of an artificial eye. At the beginning of the plot, there is a series of sinusoidal patterns that resulted from a person getting up from a chair next to the table where the eye-tracker was placed. Having participants click a mouse on the same table where the eye-tracker is placed can induce low-frequency noise with amplitudes up to  $1^\circ$ , as shown in Figure 2.14. For sensitive recordings of microsaccades and drifts, movement vibrations in the eye-tracker should be avoided, whether resulting from people moving next to the eye-tracker, clicking a mouse or typing a keyboard, or from heavy lorries driving by.

In some remote systems with poor eye cameras, precision can be so low that even *fixation and saccade analysis* should not be done without first reviewing the data and removing sections with very poor data. Figure 2.19 shows eye movements recorded by a popular commercial eye-tracker; precision is so poor that even a coarse fixation-saccade analysis does not





**Fig. 2.19** Poor precision as seen in the raw data scanpath (gaze replay) view. Notice that there are (probably) 4–5 fixations here, but that within each possible fixation, there are movements as big as many normal saccades. RMS within fixations (saccades and blinks excluded) for this recording of a real eye is  $2.92^\circ$ . Recorded 2009 with a common remote 50 Hz system and standard settings. Participant without glasses, contact lenses, and other noise-increasing physical properties.

**Table 2.6** Factors that influence precision, assuming all else equal, and effects found (\* indicate our own investigation).

Factor	Effect on RMS
Eye position in camera	Large variations *
Filtering	Up to 4 times better *
Averaging data from two eyes	Up to 7 times better *
Eye camera resolution	Large, higher is better *
Sensor refresh time in eye camera	n/a
Head compensation method	Large
Autofocus eye camera	Probably worse
Reducing focus in eye camera	Small *
Binocular recording	Small *
Pupil-only recording	Up to 3 times better *
Participant eye colour	Can be large
Bite bars and head support	Generally better *
Long range (fMRI eye-trackers)	Worse

seem worthwhile.

A summary of factors that influence precision can be found in Table 2.6.

### Accuracy

*Accuracy* is calculated as the average angular offset (distance)  $\theta_i$  (in degrees of visual angle) between  $n$  measured fixation locations and the corresponding locations of the fixation targets,

$$\theta_{\text{Offset}} = \frac{1}{n} \sum_{i=1}^n \theta_i \quad (2.5)$$

Although accuracy is straightforward to calculate given the position of a fixation, there is little consensus on exactly how to define the fixation (see Chapter 5). It is also not clear where on the monitor accuracy should be measured.

The technical key to high accuracy lies in a robust *gaze estimation algorithm*. Given the geometric set-up and the position of the pupil and corneal reflection from the eye image,

gaze estimation is used to calculate the actual gaze direction of the viewer. In a series of simulations, Guestrin and Eizenman (2006) found the main sources of error in gaze estimation to be derived from noise in pupil/corneal reflection locations and a mismatch between the assumed spherical shape of the corneal reflection compared to its true shape.

Accuracy is vital in all studies that use area of interest analysis or gaze contingency, that is, those that need to know exactly where a participant is looking. A pertinent example of this comes from the reading research community, where debate has revolved around the accuracy of the saccade targeting system versus the accuracy and precision of commercially available eye-trackers. The quote below from Rayner, Pollatsek, Drieghe, Slattery, and Reichle (2007) illustrates this point clearly:

...there can be a discrepancy between the word that is attended to even at the beginning of a fixation and the word that is recorded as the fixated word. Such discrepancies can occur for two reasons: (a) inaccuracy in the eye tracker and (b) inaccuracy in the eye-movement system. ... Kliegl, Nuthmann, and Engbert (2006) attempted to rule out the first hypothesis by using only fixations in which their eye-tracking device agreed that both eyes were on the same word. (p. 522)

In contrast, accuracy is often less important for the calculation of stimulus-independent events such as fixations, saccades, and microsaccades.

Theoretically, accuracy is limited by the size of the fovea, the area of high visual acuity, which spans some 1.5–2° of the visual field. At a standard recording distance of 70 cm distance, this area corresponds to around 2 cm. You might think that accuracy cannot be better. However, accuracy is also decided by the precise position of the eye during calibration (p. 128). If eyes are directed so that each calibration point is projected onto the same position of the fovea for each point, then accuracy can be at the level of < 0.5°, or half a centimetre at a 70 cm distance, which is what manufacturers state in their product fliers. Beware that many researchers report accuracies such as “the eye tracker is binocular, sampling at 50 Hz with 0.5° accuracy”, which is virtually always a reiteration of information from manufacturer fliers rather than accuracy measured in the researcher’s own experiment.

Accuracy tends to be poorest in the corners of the stimulus monitor, it exhibits a systematic variation across the entire field of recording, and furthermore it depends strongly on the particular characteristics of the individual participant (Hornof & Halverson, 2002).

Accuracy, furthermore, depends on the particular system used, with tower-mounted eye-trackers giving the best accuracy, followed by head-mounted systems, followed by remote eye-trackers. *Drift* means that the measured data samples move slowly away from the true gaze position, as physical conditions change after calibration. The experimenter is then forced to do regular recalibrations during the experiment. Today, most eye-trackers are less likely to drift, but even some high-end eye-trackers were very drift-prone long into the 2000s.

However, accuracy is also influenced by a number of factors that the operator of the eye-tracker must learn to work with:

- What happens during and just after calibration. For instance, many participants are alerted by the importance of the calibration, opening their eyes, tensing a little, and this is the state that you calibrate. If your participant later relaxes, changes position, and perhaps closes his eyes a little, accuracy may soon drop. Setting up the eye camera correctly before calibrating is the practical key to high accuracy. In some systems, typically the remote ones, the eye–video set-up is done automatically, while other systems rely on manual set-up. Whether your system has an automatic set-up or not, it is very useful to understand how a particular eye–video configuration relates to recorded data quality, so we will delve deeply into this on pages 116–134.

- All the individual participant and environmental properties that degrade eye-tracking data also influence accuracy: glasses, contact lenses, eye colour, varying eye physiologies, varying levels of sunlight, tears etc. all introduce an added inaccuracy averaging to 0.1–0.3° but sometimes much larger. Even variations in pupil dilation due to changes in stimulus brightness can increase inaccuracy with up to 1.5°.
- Head movements (e.g. due to an active task requiring speech or arm movements) may decrease accuracy in some remote eye-trackers, but the forehead and chin support of the high-end systems appear to stabilize the head enough to retain very good accuracy in all tasks.

Reported measured accuracy values range from 0.3° to around 2°. For instance, Jarodzka, Balslev, *et al.* (2010) found a 0.3° average accuracy in a tower-mounted eye-tracker, while Komogortsev and Khan (2008) accepted data with inaccuracies below 1.7°. Other researchers simply re-calibrate until the desired level of accuracy is reached. Tatler (2007) and Foulsham and Underwood (2008), for instance, do not begin to record data until the measured accuracy is below 0.5°. Van Der Geest and Frens (2002, p. 193) remark that results from video-based eye-trackers “should be treated with care when the accuracy of fixation position is required to be smaller than 1°”. In comparison, Deubel and Schneider (1996) report an accuracy better than 0.1° on their Dual-Purkinje-Image eye-tracker. Practical within-participants comparisons of accuracy in commercial eye-trackers have been made by Komínková, Pedersen, Hardeberg, and Kaplanová (2008) (SMI RED and HED 50 Hz) and Nevalainen and Sajaniemi (2004) (ASL 501, ASL 504 and Tobii 1750).

Several approaches have been suggested to maintain high accuracy after the initial calibration. They include online monitoring of data quality so as to trigger recalibrations when accuracy is low (Hornof & Halverson, 2002). Einhäuser, Rutishauser, and Koch (2008), for example, performed an additional drift correction when participants failed to look within 1.4° of a centrally located fixation cross within five seconds. Manual or automatic offset compensation can also be performed during the analysis phase, although this should generally be avoided (p. 224).

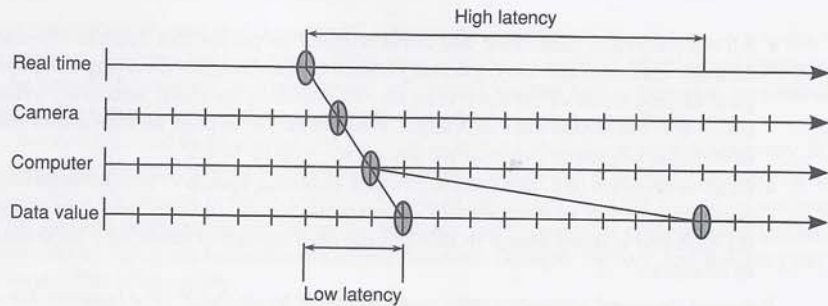
### 2.6.3 Eye-tracker latencies, temporal precision, and stimulus-synchronization latencies

Latency and temporal precision are two properties that both have to do with the minute timing of the samples recorded by an eye-tracker. Although important to all realtime, gaze-contingent studies, and for studies that require precise synchronization to external equipment such as EEG, fMRI, and motion tracking, these timing issues are remarkably little discussed. Latency and synchronization issues are also crucial when recording auxiliary data channels which complement eye movement recordings (pp. 95–108, pp. 134–139, and Chapters 9 and 13 address the combination of auxiliary data with eye-tracking).

If your study only concerns showing stimuli and recording data for offline analysis, the latency between the stimulus presentation and the eye-tracker is a much more important danger to your results; nevertheless, this source of temporal error is also little discussed in the literature.

#### Eye-tracker latency

*Eye-tracker latency* is defined as the average end-to-end delay from an actual movement of the tracked eye until the recording computer signals that a movement has taken place. Having a low eye-tracker latency is a crucial property in gaze-contingent research, but in most other types of study, the stimulus-synchronization latency of page 45 is much more important.



**Fig. 2.20** The time from eye movement ('real time') until the data value is available from the eye-tracker defines the latency. Capturing the eye image, transferring it to the computer, and calculating the gaze direction should not take longer than three samples to achieve a low latency. However, factors such as heavy CPU loads on the computer can lead to higher latencies, as indicated at the top of the figure.

Eye-tracker latency can be measured by having the recording computer (of the eye-tracker) trigger a movement in an artificial eye, and measure the time until a change occurs in the data sample output from the gaze estimation algorithm (and filters). In order to avoid additional latencies due to mechanical movements, produce an artificial movement in the artificial eye by switching off the real infrared illumination and switching on a second infrared illumination somewhere else. When the altered position of the corneal reflection is seen by the eye camera and travels through the gaze estimation algorithm, it will be registered by the recording software as a change in sample data. The time from change in infrared illumination until change in sample data is the eye-tracker latency.

An alternative but less exact measuring method is to use a mirror and a high-speed camera. The mirror is positioned next to the participant, and the high-speed camera films the participant's eye and, through the mirror, the recording software where the current raw data sample is displayed. The monitor used should have the lowest possible refresh time, and the spatial precision of the eye-tracker should be reasonably good.

Theoretically, a high-quality eye-tracking system should have a constant latency of less than three samples (i.e. 3 ms on a 1000 Hz system). This means that the eye image is being analysed and transformed into a data sample in less than three samples. The first sample is for stabilizing the eye image in the camera, the second sample is for moving the eye image to the computer buffer, and the third is for calculating the gaze data. Figure 2.20 shows the principle of latency and the path over time from the oculomotor event to the data coordinate value in the recording software. Measuring as described above, SMI and SR Research report similar average latencies for their high-end systems: slightly less than 2 ms for 1000/1250 Hz.

When working with offline analysis, if you have a *constant* latency and can measure or calculate it, you can mostly move or subtract it from the timestamp in your data, and correct for it. Unfortunately, latencies are often variable.

### Temporal precision

*Temporal precision* is defined as the standard deviation of eye-tracker latencies. A high temporal precision means that even if the samples arrive with a latency, the interval between successive samples remains almost constant. If the temporal precision is low, you have a variable latency and a variable delay in the synchronization to external software such as stimulus programs or auxiliary recordings such as EEG.

```

Recording date: 09.12.2008
Recording time : 12:40:37:484 (corresponds to time 0)
Study: XXXXXXX_52_gazereplay
Subject: G_grt-a_90
Recording: G_grt-a_90
Screen resolution: 1280 x 1024
Coordinate unit: Pixels

```

Timestamp	Found	GazepointX	GazepointY
...			
7261	Both	516	451
7281	Both	513	448
7301	Both	560	452
7660	None	-1280	-1024
7680	Both	533	454
7700	Both	528	465
...			

Fig. 2.21 After timestamp 7301, there are 360 ms not accounted for. Either timestamps are erroneous, which causes a negative latency, or samples are just not registered, which only means a loss of data. Excerpt from raw data file originating from a common 50 Hz eye-tracker used in actual research.

The cause behind variable latencies is usually that the recording computer allows processes—such as hard disk operations—to take up processor time and even get priority over the gaze data calculations. Again, for gaze-contingent tasks such as boundary crossing or simulated scotoma experiments, it is enough that the processor is occupied for a short instance, and the participant will detect an anomaly in the experiment.

The authors have recently tried a 50 Hz remote eye-tracker from a major manufacturer where latencies increased from 40 ms up to 2300 ms when increasing the load on the processor (as in the 'high latency' path in Figure 2.20). Latencies then decreased over a period of a few seconds until it was back to an acceptable value. Variable latencies of this size undermine several of the eye-tracking measures in Part III, including all duration measures (such as fixation duration, saccade duration, dwell time etc.). If these measures are important to you, and you have a system with variable latency, there is not much you can do, except to acquire a better system.

Equally important to eye-tracker latency is that samples are not lost from the data file. Figure 2.21 shows a loss of 360 ms of sample time, noticed by accident in the raw data mode, and not visible after fixation analysis has been done.

High-end eye-trackers have built in technical solutions to the latency problem. Absolutely foolproof solutions include placing the image and gaze estimation calculations on a dedicated computer board or in a real-time operating system, so they can have a processor entirely to themselves. Another solution often implemented in commercial systems, and which works for all post-processing of data, is to keep all eye images and their time code in a buffer. This makes the gaze data timestamp in the data file correct, even if the processor needs to prioritize other calculations. If you are doing gaze-contingent studies, such a system may build up the size of the buffer, and feedback to the participant may be too late. However, system tests of a high-end eye-tracker that the authors have carried out, show that such latencies are so rare that they are likely to play no role in your results.

### Stimulus-synchronization latencies

Another type of much larger latencies arises in the interplay between stimulus presentation and recording software. Stimulus presentation software sends synchronization signals to the

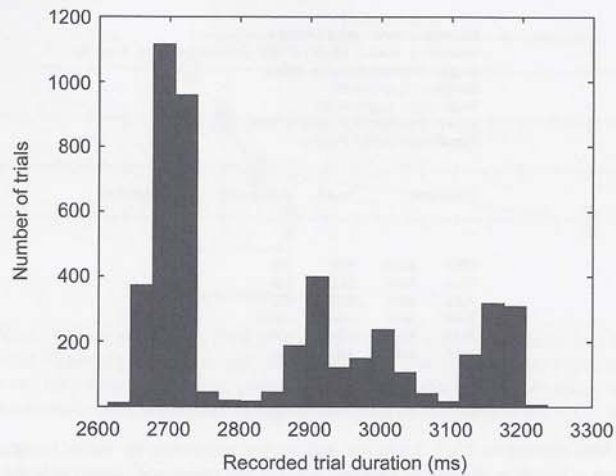


Fig. 2.22 Histogram of actual trial durations in a film viewing task. Bin size 30 ms. The films all had a 3300 ms duration. Could be related to the latency problem in Figure 2.21, or caused by latencies in the video playback. Recorded 2008 with a very common remote 60 Hz eye-tracker and manufacturer presentation software. From Wiens, Moniri, Kerimi, and Juth (2009).

recording software in order to keep the two in synchrony, typically at the onset of a new stimulus. However, clocks on the two computers may run differently, and signals may be delayed at ports for a variety of reasons. Running both stimulus and recording software on the same computer increases the danger of latencies, because the processor and hard disk must share time between recording and demanding operations such as video replay and internet browsing.

When using video stimuli, there is a latency not only at the onset of the video, but at every single frame. This is because video players typically run slightly faster or slower than the recording of data samples, so that at every frame in the video, the data sample resulting from a participant looking at that frame is in fact stored earlier or later in the data file (than the time of the frame in the video). Sending regular synchronization messages throughout the video playback gives a certain control over these variations. Another way is to use hardware time-locks, which however require advanced low-level programming.

Poor synchronization is fairly common, and can be disastrous to a study. Figure 2.22 shows data from an actual study where the stimulus presentation program failed to present videos in real time. Such timing errors then propagate through the subsequent analysis steps. This particular synchronization error was discovered as the expected statistical effect at the beginning of the film appeared to be significant even *before* the onset of the stimulus (Wiens *et al.*, 2009). Several video-playing stimulus programs have this problem. When playing videos with a commercial presentation software, we have seen latencies up to 1500 ms for 20 second videos, relative to the recorded eye-tracking data.

In some cases, inadequacies in software may cause large latencies. For example, in tests we have made with a specific software version of a very common recording software, after a few recordings, the software lagged behind (due to memory leaks, probably). This can be seen in the sluggish response to mouse clicks. In data we recorded just before this behaviour, we found latencies up to 3 seconds between onset of stimulus images and recording. Since eye-tracking software is always developing and is shipped in small series that few people give

feedback on, latencies like these may suddenly appear in data after a software upgrade, and disappear again after the next upgrade.

#### 2.6.4 Filtering and denoising

Filtering and denoising of eye-tracking data is a little-discussed issue, but most manufacturers do it to decrease variations that derive from sources other than eye movements themselves.<sup>10</sup> All types of filtering have an effect on subsequent analysis; in particular event detection, which is discussed in Chapter 5.

There are two places where filters can be found:

- Data are often already filtered while recording, and the recording software typically has settings for filtering, which are not always easy to understand for the typical user (and often not even for the experienced user). Filtering in real time during recording imposes a certain latency, which is typically around 1–2 samples (1–2 ms in a 1000 Hz system).
- There is usually a second (hidden or visible) filtering option in the analysis software used to calculate velocities and accelerations.

Beware of what types of filter your analysis software uses, since the choice can significantly affect your results. If possible, try to understand what the filters do; test a few filter settings in your software and see how they affect velocity and acceleration data.

#### Denoising and artefact removal

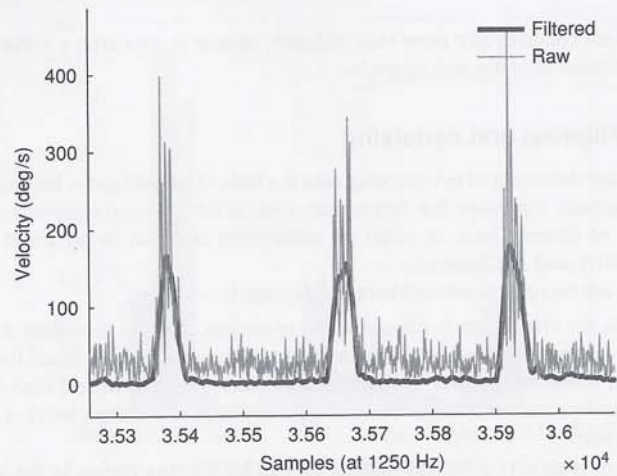
Noise reduction optimally aims to remove all variation in the recorded data that does not derive from true eye movement. It can be done online during recording or offline after all data are recorded. For some applications, such as gaze interaction, online analysis is the only option since data are used in real time to control an interface. In other applications, offline processing of data is done in preparation for subsequent analyses.

One type of noise among data samples is the *optic artefacts*. These can derive from recording imperfections due to, e.g. downward eyelashes or an erroneously detected pupil or corneal reflection. These unphysiological movements often appear as sudden spikes in the data and can rather easily be identified and removed. Stampe (1993) distinguishes between *impulse* and *ramp* noise. The former is characterized by a one sample spike, whereas the latter comprises a plateau with two deviating samples. Stampe (1993) proposes a heuristic filter design for detecting and replacing such artefactual samples with neighbouring sample values. Since access to the next sample is necessary to decide whether the current sample is impulse noise, the filtering process adds a one sample delay. Similarly, detection of ramp noise adds a two sample delay. The amplitudes of the artefacts are typically checked against a threshold such that only samples deviating more than a threshold value from their neighbouring samples are removed.

Another type of noise is the low-amplitude, high-frequency noise that occurs due to eye-tracker imprecision as well as oculomotor noise. Filters targeting precision are harder to design, because the noise and real eye movement are tightly intertwined, and the filter thus stands a risk of removing authentic eye movements.

A challenge in the design of filters for eye-movement data is to retain high-frequency information necessary to accurately describe saccadic waveforms, while removing similar high-frequency information from fixations. Kumar, Klingner, Puranik, Winograd, and Paepcke (2008) proposed a solution where fixation samples were detected and lowpass filtered online,

<sup>10</sup>Some manufacturers let you choose what filters to use; others filter for you without telling what they do.



**Fig. 2.23** The effect of filtering on velocity. 'Raw' velocity is generated from sample-by-sample differences of adjacent data samples, whereas the 'filtered' velocity represents the same data after lowpass filtering.

leaving saccade samples unprocessed. They found that this improved interaction in systems where gaze was used as an input.

#### Filtering when calculating velocity and acceleration values

Filtering is important when calculating velocity and acceleration. Velocity calculation is done by a process called *numerical differentiation*, which in its simplest form finds the eye velocity  $\dot{\theta}$  by calculating the angular distance  $\theta$  between two adjacent pairs of data samples, and multiplies this distance by the sampling frequency of the eye-tracker  $f_s = 1/\Delta t$ . Formally, this can be expressed as

$$\dot{\theta} = \frac{\theta}{\Delta t} \quad (2.6)$$

This way we get the velocity in its most common representation: degrees of visual angle per second (denoted  $^\circ/s$ ). Acceleration can be calculated by performing the same operations on the velocity samples. Notice that each time we perform a differentiation using this simple method, the noise will be magnified. Unless the precision of the eye-tracker is exceptionally high, filtering is required to produce velocity and acceleration data that are of use in subsequent analyses.

Figure 2.23 illustrates the unfiltered, noisy velocity curve, and the much smoother lowpass filtered curve, which is what you typically see in your manufacturer's software, and which is used for detecting events such as fixations and saccades. It is still possible to see the saccades in the unfiltered version, but separating fixations from saccades by means of thresholding becomes difficult.

There is a range of filters that can be used when generating velocity and acceleration data from raw data samples. The most careful investigations on this issue were made by Inchingolo and Spanio (1985) and Larsson (2010), who showed how saccade parameters (e.g. duration and peak velocity) change as a function of filter type and threshold. Many of the design criteria of filters seem to be guided by heuristics, or 'rules of thumb' motivated by visual inspection of the data (e.g. Stampe, 1993). Be aware that pattern matching filters, such as Stampe (1993) and Duchowski (2007) amplify parts of the eye-movement signal with similar appearance as the filters while attenuating other portions. Investigating the effect of filters on



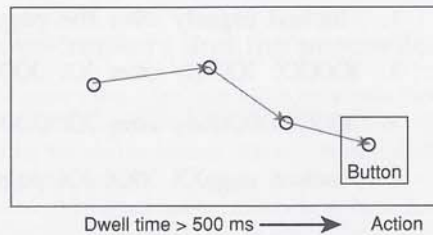


Fig. 2.24 Gaze-sensitive button. If looked at (dwelled upon) for more than e.g. 500 ms, an action is performed: changing stimulus, starting music, etc.

eye-movement velocity and acceleration, Larsson (2010) concluded that the Savitzky–Golay filter used by Nyström and Holmqvist (2010) and the differential filter by Engbert and Kliegl (2003) produced the most physiologically reasonable values. Unlike the pattern-matching filter, these two filters make no strong assumptions on the overall shape of the velocity curve.

The application imposes constraints on the design of a filter. Gaze-contingent experiments, for example, require short filters that do not introduce excessive latencies in the data, whereas offline analyses can use longer, more complex filters.

### 2.6.5 Active and passive gaze contingency

Gaze contingency means that the stimulus display changes depending on where or at what a participant is looking. There are two different ways to do this: *(inter)active gaze contingency*, which is technically easier, and *passive gaze contingency* which demands more of your system.

*Active gaze contingency* refers to the process of actively and consciously controlling an interface by means of gaze input. Figure 2.24 illustrates the principle for selecting an item, and therefore initiating an action. The gaze position from the eye-tracker basically replaces the mouse position, and this allows the user to perform actions such as open menus, click on buttons, select music, or operate an entire interface just by looking at appropriate items. Items are typically selected when they are looked at for longer than a certain duration, but can also be combined with enveloping menu hierarchies to allow for easy undoing and avoidance of the so-called Midas touch problem (Jacob, 1991). A 50 Hz eye-tracker is enough, since this is not a time critical process, but precision and accuracy must be high, so that data samples remain inside the button area for as long as the user looks there.

During *passive gaze contingency*, in contrast, participants are not required to actively control the appearance of a stimulus display. In fact, most gaze-contingent experiments of this kind assume that participants are not consciously aware that the display is updated contingent on where they look. In the typical situation, the display has a high level of detail only directly where the participant looks, whereas peripheral parts of the display are reduced in detail, but not so much that it can be detected by the participant. The display is then updated during each saccade, such that the display change is completed before visual intake begins at the beginning of the next fixation.

*Passive gaze contingency* has been implemented more for theoretical than applied purposes; it has long been used in research on vision, reading, and psycholinguistics. In reading alone, there are several types of gaze contingency, as Figure 2.25 shows. Reading researchers design gaze-contingent manipulations in order to investigate how much information we pick up from the word to the right of the fixated word. All such manipulations require very good timing on the part of the eye-tracker. In the boundary paradigm, for example, a word must be

1. looked eagerly over the pages
2. XXXXX XXXrly over XX XXXX
3. XXXXXXXXXXXrly over XXXXXXXX
4. looked eageXX XXX XX pages
5. looked eagerly over the pages  
looked eagerly over the fence

**Fig. 2.25** Different varieties of passive gaze contingency in reading research: 1. Gaze cursor over text. 2. Moving window with spaces visible. 3. Moving window without visible spaces. 4. Foveal mask with spaces visible. 5. Boundary paradigm. An invisible boundary is placed in the stimulus. When gaze passes the boundary, the word or picture on the other side changes, and the eye lands on a word contrary to what was visible before the eye movement was directed there. (Note, the small black circles indicate fixation position, shown to illustrate the technique; this may or may not be shown on screen to participants depending on your choice of gaze-contingent paradigm)

changed instantly when the saccade to it has just started. If we take the current sentence as an example; while you were fixating 'current', the next word was still 'sentence', but with a boundary paradigm, as soon as you start to move your eye, we can change 'sentence' to some other word, such as 'technology', which is what your eye will land on.

Since saccades are very short during reading, only some 20–40 ms, the gaze data must be first calculated (which takes up to three samples, see Figure 2.20), and then fed back to the stimulus program very quickly, within no more than some 10–15 ms, so that the stimulus program can update the monitor image (preferably CRT, so refresh time is low) before the end of the saccade and the saccadic blindness.

The foveal mask, case 4 in Figure 2.25, is often called 'simulated scotoma' (compare 'macular scotoma'); it artificially creates a blind spot in foveal vision which resembles the scotoma caused by physiological disorders of the eye such as macular degeneration. In healthy participants this blurring of foveal vision can help answer questions both about visual processing in physiological disorders which cause scotoma, and about visual processing per se in areas such as reading and scene perception.

The ability to make passive gaze-contingent studies of the demanding kind depends on how quickly a data sample, or the beginning saccade, can be fed back to the stimulus program so that the stimulus can be changed without the participant noticing an anomaly. Foveal masks and moving windows move in close to real time with the eye, which means that both latency and stimulus update time of the system must be very low. Only if the whole system is very fast does the illusion work, that the monitor changes without the participant noticing that a change has taken place. It is enough for the participant to notice once for his behaviour to change—once the gaze-contingent manipulation is spotted the participant will constantly remain aware of it. Thus, the eye-tracking system must have a high accuracy, a low and constant latency, a high sampling frequency, online saccade detection, and a very tight connection between recording system and stimulus presentation.

## 2.7 Types of eye-trackers and the properties of their set-up

Even if they all use the same video-based pupil-to-corneal reflection measurement technology, eye-trackers are very different among themselves. If you want to do both gaze-contingent reading research and study the eye movements of soccer players during games, you will most likely need two different eye-trackers.

The eye-trackers shown in this section are *examples of basic hardware set-ups* and their consequences on data. In real life, manufacturers offer *hardware combinations* that extend each of these basic types to more functions or allow the user to build more than one type of eye-tracker from the same set of basic hardware, but the basics are always the same.

### 2.7.1 The three types of video-based eye-trackers

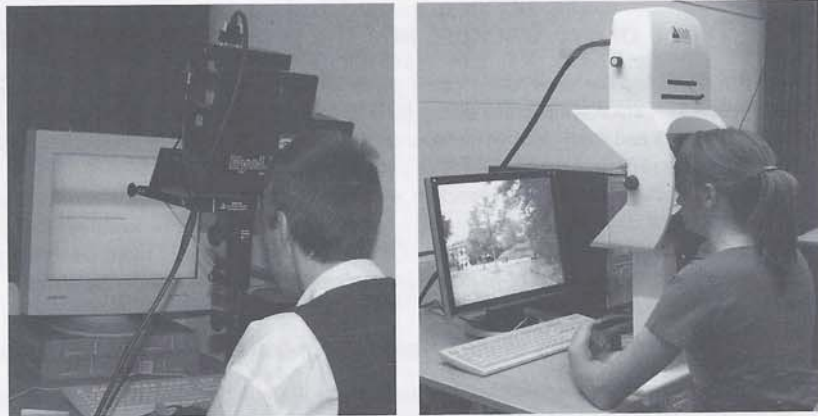
Basically, a video-based eye-tracker has an infrared illumination and an eye video camera, and typically an additional scene camera for head-mounted eye-trackers. Illumination(s) and camera(s) can be put on a table in front of the participants, or on their heads. Sometimes head-tracking is added to head-mounted systems. This gives us three types of eye-trackers that differ not only with respect to the position of cameras and illumination, but more importantly in the type of data they produce and how we can analyse the output.

1. The most common set-up is the *static eye-tracker*, which puts both illumination and eye camera on the table, in front of the participant. There are two sub-types; *tower-mounted* eye-trackers that are in close *contact* with the participant, restraining head movements, and those that view from a distance, known as *remote* eye-trackers, with nothing or very little attached to the head. In practice, stimuli are almost always presented on a monitor, although wall projections and real scenes can easily be used with static eye-trackers.
2. Another common set-up is the *head-mounted eye-tracker*, which has put both illumination and cameras on the head of the participant, mounted on a helmet, cap, or a pair of glasses. A scene camera takes the role of recording the stimulus—the scene of view.
3. The third type of set-up adds a *head-tracker* to the head-mounted eye-tracker in order to calculate the *position of the head in space*. For reasons soon to be explained, this addition makes the analysis of data from head-mounted systems much easier. Not many manufacturers offer this combination, however.

These three different ways to combine illumination and camera give eye-trackers with very different properties. We will use the terms 'remote', 'tower-mounted', 'head-mounted', and 'head-tracking' throughout the rest of the book. We will now describe the three types in detail.

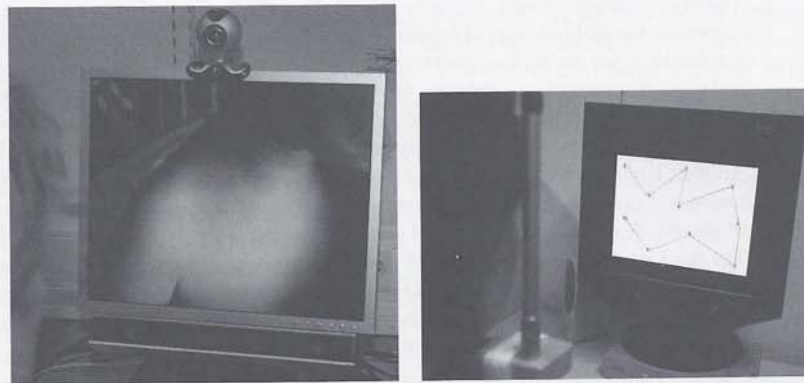
*Static eye-trackers* come in two varieties; those that restrict the participant's head less, and those that restrict it more. For a number of reasons, you get better data if you restrict the participant's head more. Previously, bite-bars were used to immobilize participants' heads. Today, the video-based eye-trackers with the best precision have forehead and chin rests that gently restrict the participant's head movements, like at the optician, as in Figures 2.26(a) and 2.26(b). The camera and illumination are hidden inside the box on top of the eye-tracker. The gentle head restriction is the price you pay for high precision and accuracy.

Eye-trackers that restrict the head less place the camera near the stimulus (monitor), without contact to the participant. These 'remote' eye-trackers are capable of viewing the participant's eye from a distance, and even keep track of the eye as it moves within a certain volume (Figure 2.27). Because of imperfections in gaze estimation models during head movement, and because the eye is typically filmed at a lower resolution, data from remote systems are



(a) The SR EyeLink 1000 Hz tower-mounted eye-tracker. (b) The SMI 1250 Hz HiSpeed tower-mounted eye-tracker.

**Fig. 2.26** Two types of static tower-mounted contact eye-trackers with high sampling frequency, precision, and accuracy. Both film the eye via a mirror.

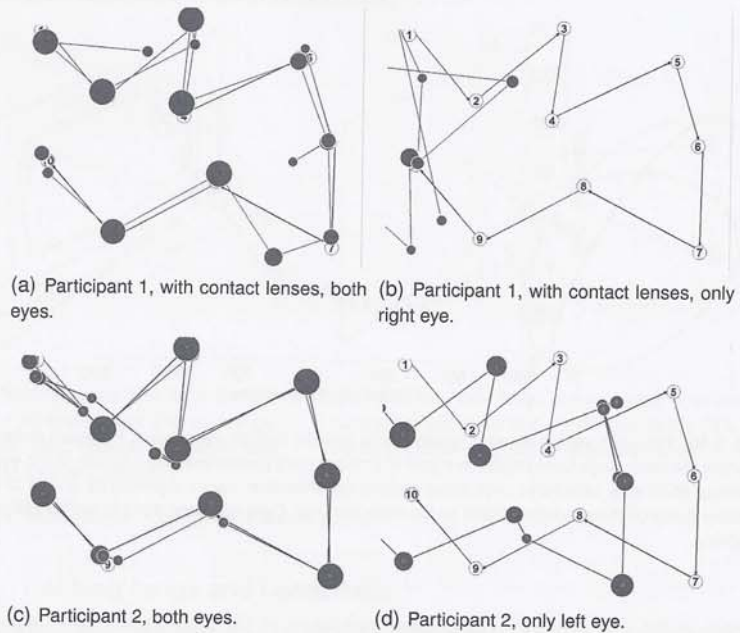


(a) SMI RED 4, a 50 Hz remote system. Here with an additional web camera to observe participants' facial expressions during diagnosis of neuropaediatrics cases. (b) Tobii 1750 remote system at 50 Hz. One of the most sold eye-trackers in the mid 2000s. The participant is using a chin rest to increase precision.

**Fig. 2.27** Remote eye-trackers: illumination and eye camera are hidden in the dark ledge below each monitor. Nothing is attached on the participant.

almost invariably of a poorer quality than data from those eye-trackers that restrict or measure the participant's head movements. Research is intense in improving the data quality of remote eye-trackers, however.

Knowing the position of the head is again the key to sufficient accuracy and precision. One particular variety of remote eye-trackers solves the geometrical problem by putting an infrared reflector on the forehead of the participant and measuring exactly where he is. Others that use magnetic head tracking have also been on the market. Their low market share indicates that many users prefer not having to add any markers or sensors to the participant's



**Fig. 2.36** The effect on accuracy of concealing one eye. Data from two participants looking at the numbers 1 to 10 in order, recorded 2009 on a very common but slightly outdated remote 50 Hz eye-tracker. Solid circles represent detected fixations. The small fixations next to the large ones are artefacts from low precision in combination with a dispersion-based fixation algorithm.

culating averages between the two data samples, or using the eye with currently better data quality. The precision increase can be considerable (Cui & Hondzinski, 2006). Conversely, if an averaging remote system temporarily loses track of one eye, there can be a dramatic loss of accuracy in the data, as Figure 2.36 very clearly shows. If you plan to buy a remote system, this is one property to test.

### 2.7.5 The parallax error

A head-mounted system with a scene camera will exhibit a larger or smaller *parallax error*; that is: the gaze cursor, which we saw as a cross in Figure 2.39, will be misaligned with the actual line of gaze for certain distances from the object looked at. Figure 2.37 shows the principle for parallax error. The size and direction of the parallax error varies systematically with the distance to the object looked at. If the fixated object is at the same distance as the calibration plane, when the participant was calibrated, then the parallax error will be zero. At larger or smaller viewing distances the parallax error grows, being at its worst at very close distances.

In the gaze-overlaid scene video for the right eye, the parallax error always follows a line, slightly tilted as the scene camera is above eye level, with far away errors to the left and close object errors to the right, assuming the head-camera configuration in Figure 2.37 (the reverse case for the left eye). The error is so systematic that if you are coding the resulting gaze-overlaid videos manually, it is possible—with a little training and an understanding of the principle behind parallax—to directly estimate the parallax error, and subtract it from the



**Fig. 2.28** A four-camera SmartEye system set-up in a car simulator. Infrared illumination is seen as white hexagons on either side of the steering wheel, and two of the cameras are seen as dark silhouettes on top of the dashboard.

head, even if it is at the cost of a somewhat lower data quality.

Remote eye-trackers are generally easier to operate, however, and the participants more easily tend to forget that the eye-tracker exists. They are the only practical alternative to record on infants, and allow additional measuring equipment to be added to the participant's head without too much interference. Remote systems also exist in multi-camera versions, which can be built into workplaces (such as cars and flight simulators) where participants need to turn their head a lot. Figure 2.28 shows a four-camera set-up.

There is a hypothesis—sometimes used as a sales argument—that participants behave more naturally in a remote system than in tower- and head-mounted systems, which would mean that data would be ecologically more valid. Such a superiority of remotes remains to be proven, however. In the authors' experience, participants feel only moderately restrained even in tower-mounted systems: in our lab we have had participants speak, type with the keyboard, and even play intense first-person-shooter games, and we have still recorded highly accurate data from participants who claim to have behaved naturally.

When a participant moves his eyes out of reach of the eye camera, and then returns to his initial position, the remote eye-tracker typically takes some time—known as *recovery time* and sometimes 'pickup time'—to resume tracking, which is of extra importance if you have very mobile participants, such as infants. Recoveries are also made after long blinks, and tend to be longer with remotes than with systems that know where the eye is. If your system has a long recovery time you lose a corresponding portion of data during each recovery.

Static eye-tracking systems, whether tower-mounted or remote, can be used with one stimulus plane, typically a monitor, but also a magazine, or just the monitor-less field of view above the camera housing. A static eye-tracker gives a data file with coordinates in the coordinate system of that plane, i.e. that monitor magazine or field of view. This coordinate system is defined and positioned as part of the set-up and calibration, which we will look at in Section 4.5. Giving coordinates in a fixed coordinate system is an important property that makes analysis much simpler, and it is by no means a self-evident one. The reason that it works is that the eye camera, the illuminations, and the stimulus (the monitor looked at) are fixed spatially, and the head of the participant is fixed or at least fairly well measured.



(a) SMI HED head-mounted system with Polhemus headtracking. The device generating the magnetic field is located in the upper part of the figure. Sensors are mounted on the helmet.



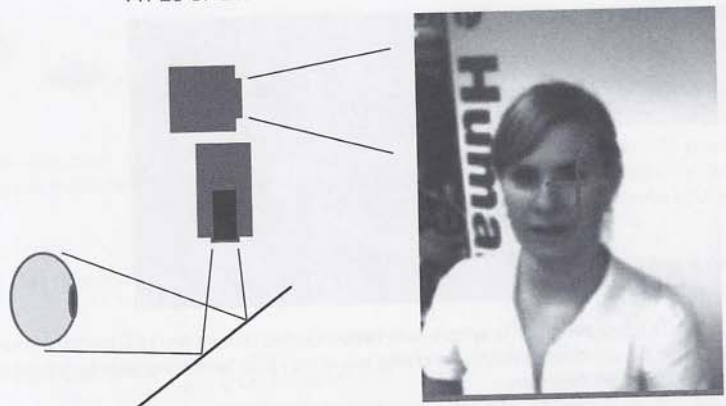
(b) SMI HED-mobile system with recording computer in rucksack (from 2008). Photo used courtesy of Gunnar Menander and Petra Francke.

Fig. 2.29 Head-mounted systems with (a) and without (b) head tracking.

*Head-mounted systems* as in Figure 2.29(a) have cameras and illuminations mounted on top of a helmet, cap, head-band, or a pair of glasses. They allow the participants maximum mobility, in particular if the recording computer is small and lightweight. If the mounting is steady, the participant can take part in many different real life activities, such as driving, riding a bicycle, buying food in a supermarket, playing tennis etc. Traditional head mounted systems are likely to be the most versatile eye-trackers when it comes to glasses, contact lenses, drooping eyelids, mascara and steep viewing angles. This is because the eye camera angle towards the eye can—in principle—be shifted and adapted to the individual participant and task—an important property that we will cover in detail on pages 116–134. A notable exception are the Tobii eye-tracker glasses from 2010, where no parts can be adapted.

Head-mounted systems also have a scene camera mounted on the helmet, filming in the line of sight, as in Figure 2.30. The recording computer overlays the gaze coordinate onto the scene video, which is endowed with a moving marker that shows where the participant is looking. The result is known as a *gaze-overlaid video*, and for reasons that we will discuss on page 61, this is (mostly) all you can use. Even if there is a data file, the coordinates of the data file will refer only to positions in the scene video, not to positions in the surrounding world. We will have a closer look at the consequences of this property on data analysis in pages 60, 175, and 227.

In order to associate objects in stimulus space with collected data samples, some head-mounted eye-trackers can be combined with magnetic head tracking. The head-tracking system calculates all (absolute) motion of the head, and adds that onto the (relative) motion of the eye in the head. The resulting combined head–eye gaze vector is expressed in the coordinate system of the recording environment. Surfaces in the environment can be measured up, and given their own reference systems. With such a combined system, the eye-tracker will give both *gaze-overlaid video* and a *very useful data file*, that allows for automatized data analysis, while still allowing for very large head movements.



(a) Head-mounted eye-trackers have an additional scene camera; here placed above eye camera and illumination. The scene camera films forward, while the eye camera in this case films through a mirror.  
 (b) Scene camera view with gaze-overlaid haircross cursor at the gaze position.

Fig. 2.30 The scene camera of head-mounted systems.

Another method to get data coordinates that are meaningful in analysis is to attach (large enough) object markers with specific patterns onto the stimulus. These markers are captured as part of the scene video, and four markers define the corners of a rectangular object. When the participant is within a certain range, the recording software finds the markers, and can relate gaze to the objects even with lots of head movements. Putting markers in a natural stimulus risks altering the participant's visual behaviour, however.

There are also head-mounted eye-trackers with a more restricted version of head tracking. The most well known is the SMI EyeLink I 250 Hz system and its follow-up, the SR EyeLink II 250/500 Hz, which has a small camera on top of the headgear and four light emitting diodes (LEDs) at the corners of the single plane (see Figure 2.31). On the basis of the positions of LED corners in the camera image, the software calculates the position of the head relative to the monitor, and thus compensates for head movement. This works fairly well, but the plane must largely be perpendicular to the line of gaze and cannot for instance be tilted to lie down on a table. Also, with this technical solution, it is not possible to get data samples from more than one plane. In fact, this type of eye-tracker much resembles tower-mounted eye-trackers in its high precision, but still allowing for some degree of head movements. Accuracy, however, sometimes has the tendency to drop over time, if the eye-tracker slips relative to the head. Data consists of a file of co-ordinates in one plane, just like the static eye-trackers.

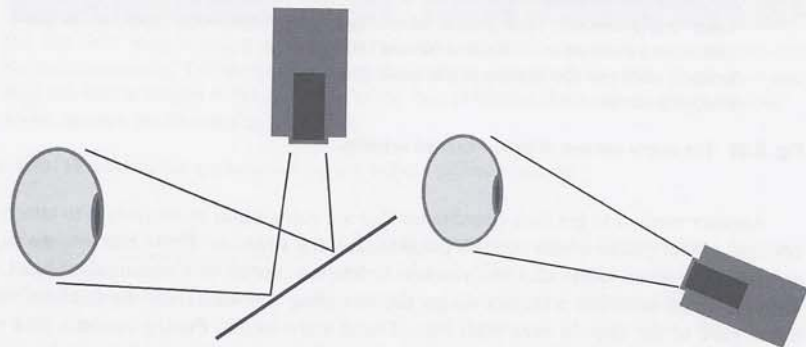
There are other human interfaces, such as fMRI (functional magnetic resonance imaging) eye-trackers, VR (virtual reality) goggles with built-in eye tracking, and even primate eye-trackers. For the purpose of this section, it suffices to say that they are all versions of static eye-trackers: The stimulus is fixed in relation to the eye camera and the participant's head.

In addition to gaze direction, many of these eye-trackers output the pupil and corneal reflection positions in the eye-video, which can—assuming there is software to support this—in principle be used to extend the area from which data is collected to beyond the area calibrated. Some systems even output these gazes outside of the calibration plane as coordinates, but for a variety of reasons, the quality of this data is typically much poorer than within the calibration





**Fig. 2.31** The EyeLink I 250 Hz system, with head-mounted camera and LED markers around the stimulus to define the coordinate system. For clarity, two of the LEDs have been circled. Picture used with kind permission from SR Research.



(a) Contact and head-mounted eye-trackers with mirror. (b) Remote eye-trackers and mirrorless head-mounted eye-trackers.

**Fig. 2.32** Eye camera viewing angle towards the eye. The grey and black boxes are the infrared illumination and camera.

plane.

Also note that contact eye-trackers and (most) head-mounted eye-trackers film the eye through a mirror, while the remote eye-trackers film directly, as in Figure 2.32. On pages 116–134, when we describe how to set up the eye camera to get an optimal viewing angle and the best possible data quality, we will often manipulate the camera angle towards the eye to adjust it to our participant's individual physiology, glasses etc.

All current high-end systems have mirrors, but the ideal eye-tracker is mirrorless, because the mirror introduces edges in the visual field of the participant. This may change the wavelengths of the light reaching the participant's eye and introduce reflections from overhead illumination. Nevertheless, the advantages of having camera and illumination in a fixed position to the participant's eye, and the increased data quality this means, still outweigh the disadvantages of introducing a mirror. Eye image quality is high with current mirrors, and with the stimulus display at a distance much larger than that of the mirror, in our experience, participants are no more disturbed by the mirror than by wearing glasses.



(a) Eye-video from participant aged 48 days successfully tracked with dark-pupil remote eye-tracker.



(b) Eye-video of an orangutan from Lund University Primate Research Station at Furuvik Zoo, recorded using a remote dark-pupil system.

Fig. 2.33 Two representatives of uncommon groups of participants recorded with the same remote eye-tracker.

## 2.7.2 Robustness

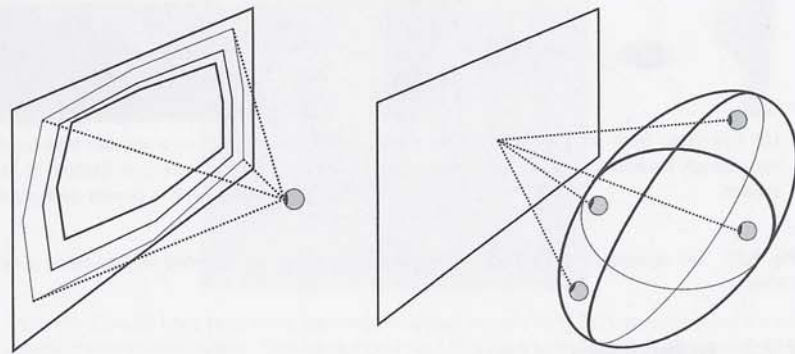
*Robustness* (also known as *versatility*) refers to how well an eye-tracker works for a large variety of participants. Poor robustness can lead to frequent data loss of up to a third of the participants (p. 141) and poorer data quality. Some eye-tracking systems have a hard time with glasses, and contact lenses are also a problem that some eye-trackers handle better than others. Participants also vary in their eye physiology: some eye-trackers may produce poorer data for certain eye colours, and another common obstacle to good data recording is drooping eyelids, which of course vary between participants.

In order to handle as many participants as possible, your eye-tracker should allow you to vary the angle of the eye camera to the eye. As we will see on pages 116–134, adjusting the eye-camera angle solves many problems. This is often more difficult on remote systems than on head-mounted ones, but the design of the head-mounted system also plays an important role to its robustness. If you can also vary the position of the infrared illumination, your system will be even more versatile, especially for people wearing glasses. The major problem with contact lenses can easily be solved, as we will later see, if you can adjust the focus of the eye camera. The current manufacturer trend is to remove these options from the operator of the eye-tracker, and thereby prioritize ease of usage over robustness.

Technological components that manufacturers use to automatize robustness reflect what the researchers would manipulate: increased resolution and sensitivity of the eye camera(s), number of eye cameras, the quality of the camera sensors, the number of and precise positioning of illuminators, and the implementation of image processing algorithms, which together have a major impact on the robustness of an eye-tracking system whether or not the researcher has access to camera settings and positions of illumination.

Infants are a special class of participants, for whom remote eye tracking is most suited. Infants may move about a lot, and can hardly be restricted using chin rests or bite bars. Before they can sit, they can be placed lying in a tight hammock that gently restricts sideward head movements, with the eye-tracker and stimulus monitor overhead. Light should be turned down in the room, so the stimulus monitor appears more salient to the infant. When sitting in a parent's lap, holding their head gently minimizes large head movements. Figure 2.33(a) shows a very young human participant recorded on a remote eye-tracker.

Our primate relatives would disassemble the eye-tracker if they could, so they have historically always been more or less restricted when being eye-tracked. New attempts are made to track through protective glass, and remotes are then the common choice among video-based trackers. The orangutan in Figure 2.33(b) was tracked through a 20 mm polymethyl methacrylate glass at 50 cm distance.



(a) Tracking range: how far to the side can a participant look and we still get (good) data. Towards the edge, data will be gradually poorer.

(b) Head box: how large is the volume in which the participant can move and we still get (good) data. Towards the edge, data will be gradually poorer.

Fig. 2.34 Illustration of tracking range and headbox.

### 2.7.3 Tracking range and headboxes

The *tracking range* (also known as *visual field of recording*) is a measure of how far to the side your participant can look and you still get data. The *headbox* is the volume relative to the eye-tracker in which a participant can move without compromising the quality of recorded data. Figure 2.34 illustrates both concepts. There is currently no generally accepted definition of either that could be used to compare different eye-trackers.

Most eye-trackers will have problems with extreme *gaze angles*, that is when the line of gaze deviates very much from the direction of the head, because the corneal reflection degrades or either pupil or corneal reflection are covered by the eyelid (p. 131).

The tracking range is particularly important with large stimuli. If you have a broadsheet newspaper as a stimulus, and an eye-tracker that does not allow your participant to move his head, then looking at the texts in the corners of the pages will require him to turn his eyes very much to the side. If your participant is a car driver, he will make very large eye (and head) movements when alternating between the rear-view mirror, pedestrians on each side of the car, the GPS on the dashboard etc. If instead you have a very small stimulus, such as a cellular telephone, your participant does not have to make very large eye movements to look at all of it.

Single-camera video-based eye-trackers can measure gaze on a stimulus within a horizontal gaze span of around  $40^\circ$ , and around  $25^\circ$  in the vertical direction, relative to the head direction. For a static eye-tracker (the common one-camera remote, for instance) at a viewing distance of 70 cm, the horizontal  $40^\circ$  corresponds to a width of approximately 50 cm. Tracking range is not an absolute property of the eye-tracker, but depends on participant physiology, with borders of decreasing data quality in all extreme gaze angles (p. 116–134).

A large headbox is important when participants move around in front of a remote eye-tracker, for instance infants, monkeys, and some clinical groups. For many remote systems, data quality shows considerable variation within the headbox, and may be much worse towards the extremes, as the precision measurement in Figure 2.35 exemplifies.

Multi-camera remote systems can measure gaze in the entire  $360^\circ$ , because another camera takes over when the first one falters. This is very useful when participants rotate on an office chair in front of a large control board, or if they drive a car. The increased headbox often

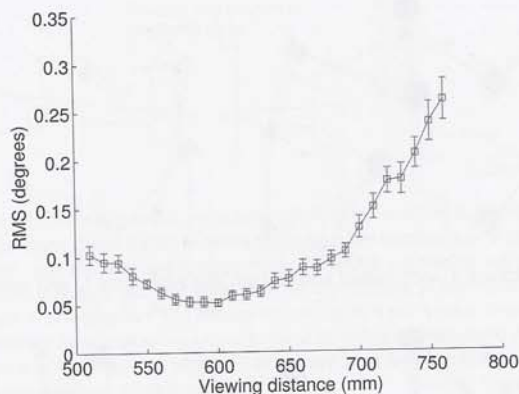


Fig. 2.35 Precision versus viewing distance in a specific remote eye-tracker. For each distance from the monitor, 24 recordings were made on a pair of artificial eyes positioned in the middle of the eye video, and average RMS was calculated over those recordings. Precision varies significantly and is at its best at a smaller distance than recommended by the manufacturer. Data collected 2011 in collaboration with Pieter Blignaut.

comes at the cost of a time consuming calibration of the environment and head of the participant, and therefore suits experiments with few participants and long trials. There are also remote eye-trackers with single eye cameras that mechanically move with the participant's head, also creating larger headboxes.

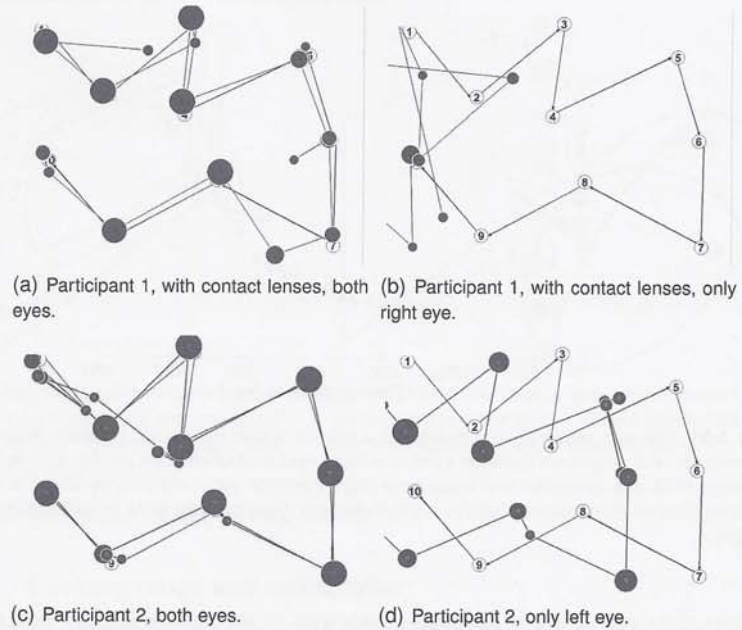
### 2.7.4 Mono- versus binocular eye tracking

Monocular eye-trackers record from one eye only, while a binocular eye-tracker takes data from both eyes. The vast majority of eye-tracking research is done monocularly, for mainly two reasons. First, it is commonly believed that both eyes make the same movement at approximately the same time, and also look at roughly the same position. It would therefore give no additional value to measure both eyes simultaneously. As we saw on page 24, however, this is not always a valid assumption. Second, monocular eye-trackers are cheaper to acquire. These two reasons combined make it tempting to buy a system that can measure only from one eye at a time.

Binocular eye tracking may be particularly relevant in some situations:

- When recording from children, who have a larger distance between gaze positions of the left and right eyes—known as *disjugacy* or *disparity*—than adults.
- If you have an experiment where double vision due to the misalignment of the two eyes relative to each other—known as *diplopia*—may occur.
- If you plan to perform clinical studies on participants with neurological dysfunctions affecting vergence (see Leigh & Zee, 2006, p. 367).
- If small differences in saccade measures matter to your study.

The purpose of recording binocular data often differs between high-end and low-end eye-trackers. High-end eye-trackers output one data stream for each eye with a quality high enough to address the situations above. Low-end eye-trackers typically give you only one data stream (a cyclopean view), even if both eyes are used. In this case, binocularly is used only to increase the accuracy and precision of the data from a remote eye-tracker by cal-



**Fig. 2.36** The effect on accuracy of concealing one eye. Data from two participants looking at the numbers 1 to 10 in order, recorded 2009 on a very common but slightly outdated remote 50 Hz eye-tracker. Solid circles represent detected fixations. The small fixations next to the large ones are artefacts from low precision in combination with a dispersion-based fixation algorithm.

culating averages between the two data samples, or using the eye with currently better data quality. The precision increase can be considerable (Cui & Hondzinski, 2006). Conversely, if an averaging remote system temporarily loses track of one eye, there can be a dramatic loss of accuracy in the data, as Figure 2.36 very clearly shows. If you plan to buy a remote system, this is one property to test.

### 2.7.5 The parallax error

A head-mounted system with a scene camera will exhibit a larger or smaller *parallax error*; that is: the gaze cursor, which we saw as a cross in Figure 2.39, will be misaligned with the actual line of gaze for certain distances from the object looked at. Figure 2.37 shows the principle for parallax error. The size and direction of the parallax error varies systematically with the distance to the object looked at. If the fixated object is at the same distance as the calibration plane, *when the participant was calibrated*, then the parallax error will be zero. At larger or smaller viewing distances the parallax error grows, being at its worst at very close distances.

In the gaze-overlaid scene video for the right eye, the parallax error always follows a line, slightly tilted as the scene camera is above eye level, with far away errors to the left and close object errors to the right, assuming the head-camera configuration in Figure 2.37 (the reverse case for the left eye). The error is so systematic that if you are coding the resulting gaze-overlaid videos manually, it is possible—with a little training and an understanding of the principle behind parallax—to directly estimate the parallax error, and subtract it from the

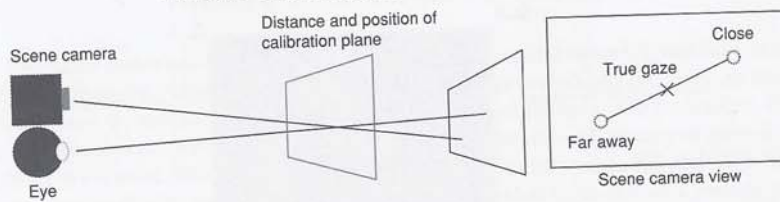


Fig. 2.37 Parallax error. The grey plane is the position of the original calibration screen. The dark plane is the true stimulus. The scene camera view shows a frame from the overlaid scene video. The cross marks the true gaze position, and the two dotted rings mark where the eye-tracker will put the overlaid gaze marker in relation to the true gaze position, depending on the distance to the object looked at. A far away stimulus pulls the error in one direction, and a close stimulus pulls it in the other direction. The gaze cursor is only perfectly positioned for stimuli at the same distance as the calibration plane during calibration. This figure assumes the scene camera is mounted above the eye level between both eyes, and the displayed error is true for measurements on the right eye.

gaze position shown while coding.

The cause of the parallax error is the fact that the scene camera and the eye look at the scene from slightly different angles. The two gaze lines of the eye and the scene camera cross at the distance at which the calibration screen was, but are misaligned elsewhere. There are several hardware solutions to parallax:

- A few eye-trackers are built so that the line of sight coincides perfectly with the scene camera direction, but at the cost of difficult and endurable mechanical solutions.
- Adding head tracking and measuring the position of the eye and scene camera allows a mathematical solution to parallax, but head tracking limits the mobility of the participant.
- Binocular head-mounted eye-trackers can use a depth calibration and linear depth correction.
- Another alternative with a binocular eye-tracker is to simply average data samples from the two eyes.

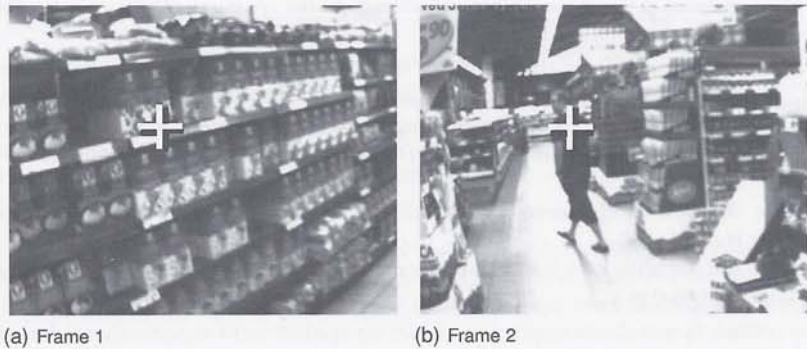
When these solutions are not available, the eye and scene cameras should be placed as closely together as possible on the headgear.

### 2.7.6 Data samples and the frames of reference

The vast majority of all eye-tracking research today consists of showing participants sequences of still images (possibly with sound), and having the observing participant sit more or less still in front of a static (remote or contact) eye-tracker. Virtually all analysis software is written for this particular set-up. The data file that you get will be a sequence of coordinates with time stamps, and the coordinates will be meaningful in the coordinate system of your stimulus. For instance, if you show a face, as in Figure 2.38, any raw data sample with coordinates ( $x = 262, y = 291$ ) will *always* be the coordinate of the pearl earring in that face image, because everything is still and kept in place. This association between the coordinate of a raw data sample and a semantic object is necessary later, in the analysis software. It makes it possible to calculate the duration that a participant gazes at the object (known as 'dwell time') and how many times he looks there (known as 'entries'). Also fixation and saccade analysis becomes easier. The association between sample coordinates and stimulus positions is established during calibration, typically quite a fast set-up carried out for each participant individually (p. 128).



**Fig. 2.38** The coordinate system in eye-tracking studies typically has the origin ( $x=0$ ,  $y=0$ ) in the upper left corner. For as long as this picture is shown, the data coordinate ( $x=262$ ,  $y=291$ ) will be a point on the pearl earring; that is, we have a link between a coordinate and a semantic object. The stimulus is the painting 'Het Meisje met de Parel' by Johannes Vermeer, circa 1665.



**Fig. 2.39** Head motion and motion in the stimulus dissociate the link between gaze coordinates and semantic objects. Pictures show how the two data samples with identical coordinates indicate two entirely different objects at two instances separated by only a couple of seconds. Two frames from a gaze-overlaid video recorded with the SMI HED mobile system.

There are two large obstacles to the desired connection between gaze direction and a meaningful portion of the stimulus viewed: either *the participant's head moves*, or *the scene changes*, in ways the system cannot measure (we will overlook for now the errors stemming from moving the infrared emitters or the camera, adding glasses to the participant after calibration etc.). Suppose the researcher has a participant walking in a supermarket with a mobile head-mounted eye-tracker. Coordinates in the data file would indicate different positions at different times, because the participant moves. The coordinates (262,291) may at one moment be a box of pasta on the product shelf, and only seconds later be one of the other customers in the supermarket (see Figures 2.39(a) and 2.39(b)). Since the coordinate reference (262,291) changes in meaning all the time, adding up gaze data at this position over a whole trial (a round in the supermarket) would simply not give a value that refers to one single object, and many types of analyses would make no sense.

The same dissociation of coordinates and stimulus content occurs if you have animated stimuli: web pages with scrolling, video clips, and various multimedia products. As an object moves across the screen, its coordinate values change continuously with its position. First of all, this stimulus motion may make coordinate based analysis such as area of interest (AOI) analysis much more difficult, unless you use dynamic AOIs (see pages 209 and 227). Second, when viewing slowly moving objects, the eye makes what is known as a *smooth pursuit motion*, and this makes fixation analysis invalid with some of the current methods for event detection (p. 169).

There are notable technical solutions to parts of the problem with gaze coordinates lacking a permanent association to semantic objects.

- If you have a head-mounted system, the simplest solution is to *lock the head* relative to the stimulus. It works for single monitor stimuli, making a tower-mounted system out of the head-mounted system.
- If you do not want to fasten the head-mounted system (and the participants) in a tight construction, you may instead *measure the precise positions and movements of the head*, and add that measurement to the eye-tracker position to get a resulting *gaze vector* that is meaningful in the environment of the head-tracker. Both magnetic and camera-based head tracking can be used, but magnetic is the common choice.
- If you do not use a head-mounted system, but your coordinate problem is due to animation in the stimulus, the optimal solution is to *automatically detect and extract the coordinates of objects from the animated stimulus*. This is difficult in general, but feasible when the target object is known and very different from other objects in the environment, such as an overhead information board at a train station or an airport. Scrolling on web pages was a problem until two manufacturers made their own browsers that could read the scroll distance, and add it to the gaze coordinate. Placing markers on the stimulus and filming them with a scene camera (Figure 2.31) can give coordinates in the space spanned by the markers. In a similar manner, you may add radar and GPS measurements onto your car, so it calculates the precise position of your own and neighbouring cars, of pedestrians and buildings, and immediately calculate what object the driver's gaze hits.

With a head-mounted or a remote eye-tracker, if you record only gaze-overlaid video, but want to calculate numbers for the data, for instance how much time each participant looked at each type of pasta, then you have to extract this information manually from the video. If you have 20 minute recordings on video, and want to code fixation durations and dwell time (two central measures for how long people look), throughout the recording—for instance by requiring the gaze marker to be still for three frames which equal 120 ms for a 25 Hz video system—you may spend several days coding each participant. If your task requires you to code only portions of the recordings, an eye-tracker that only delivers gaze-overlaid video is more acceptable. Skipping large portions and coding only small time windows is reasonable in face-to-face interaction studies as well as consumer visual behaviour in supermarkets, assuming that the question is “What did they look at during the 10 seconds before they selected what to buy?”, or “What did they look at during the 5 seconds after the interlocutor uttered a pronoun?”. See pp. 227–229 for different methods for coding gaze-overlaid video data.

In conclusion, an eye-tracker produces streams of data samples, either in the coordinate system of the scene video camera attached to the participant's head, or in the coordinate system of the stimulus. As we noted, an area of interest can be placed on the stimulus to give a portion of the  $(x,y)$ -values a semantic value like ‘the pearl’. The recorded eye-movement data can then be expressed as a sequence of semantically meaningful areas, such as PPMPPH, where P, M, and H refer to the pearl, mouth, and headscarf in Figure 2.38. This is also known



as an AOI string. An alternative to letting the researcher decide what values  $(x,y)$  coordinates should be given, is to use values calculated from the  $(x,y)$  points in the stimulus. For instance, we can calculate the luminosity at each position and get a string of numerical luminosity values, or produce a string of pasta prices by replacing each gaze hit on a pasta package by the price of that particular pasta. These are called feature values and feature strings, because luminosity and price are features of the stimulus. While semantic AOIs can only be used when the researcher defines the AOIs in accordance with the experimental design, feature values can be extracted algorithmically from the stimulus. Chapter 6 explains these uses of eye-movement data in detail.

## 2.8 Summary

This chapter introduced properties and varieties of video-based eye-trackers, which produce an output consisting of:

- often the position of the pupil and corneal reflection from the eye camera, *αντανάλυση κερρατικού*
- always raw data samples with time stamps and  $(x,y)$  coordinates,
- sometimes velocity and rarely acceleration values,
- sometimes gaze-overlaid videos, *βίντεο-υπερθέση*
- in many eye-trackers, pupil size is a free extra. *μέγεθος της κόρης*

Eye-trackers are built for different purposes and have a number of hardware and software related properties that should be taken into account when designing an experiment. The resolution of the eye camera and sampling frequency are examples of important hardware properties that influence what types of eye movements that can and cannot be measured. Software accompanying the eye-trackers contain algorithms that perform, for example, image analysis to find pupil and corneal reflections in the eye video, gaze estimation, and calculation of eye-movement velocity and acceleration from raw data samples.

Together with the participants and the recording environment, such properties decide the quality of the recorded data, and thus largely constrain the research questions that can be addressed, and the type of analyses that can be performed on the data. It is therefore important to have a basic understanding of how your eye-tracker works, in order to successfully design an experiment, record data, and analyse the recorded data. ✓