

## 13 Latency and Distance Measures

---

Although latencies and distances are central metrics in the vast majority of eye-tracking measures covered thus far in this book, we reserve this chapter for 'time' and 'space' specifically as it pertains to unitary eye-tracking events *in relation to other events*.

*Latency* is a measure of time delay, that is the time from the on- or offset of one event to the on- or offset of another (p. 286). Eye-voice latency, for instance, is the time between looking at a word or picture and saying it. Several of the latency measures can be used as reaction time measures.

*Distance measures* return the distance from one point to another (for instance gaze and mouse cursor) at one and the same time, either as a single value (e.g. saccadic gain) or as a continuous measure (e.g. smooth pursuit gain). These distance measures differ from saccadic amplitude for example, where the distance refers to the movement of *one single point* over a period of time.

Both latencies and distances are often called 'spans', and eye-voice span is identical to eye-voice latency. Spatial distance and temporal latency are tightly coupled aspects of relative motion between the two events. Because of this, latency and distance are often used interchangeably: eye-voice latency can either be measured in the temporal domain (milliseconds), or the spatial domain (number of words, or pixels). Measures that are predominantly 'time-based' have been put in the latency section of this chapter, while measures that are more 'space-based' are found in the distances section.

Because they have a true zero-point, namely the zero milliseconds latency when both events are simultaneous, all latency measures are of the ratio type, which allows for the use of a wide variety of parametric tests for statistical calculations (p. 90). The exception is the complex latency of the proportion over time measures, for which we present a separate statistical section.

Be aware that these latencies are dependent measures, and not sources of error like the system latencies on page 43. However, for values of latency measures to be correctly measured, it is crucial that your stimulus program and recording system are temporally accurate; if the flash or new stimulus picture is shown 48 ms later than the mark shows in your eye-tracking data file, due to slow loading and rendering, your latency value will be 48 ms too high. Also, your *detection algorithm* must calculate the onset of the eye-movement event correctly, which is easier for fast saccades and more difficult for slow smooth pursuit. In addition to these hidden and variable errors in the stimulus software and detection algorithms, the *sampling frequency* on average causes a constant error of half a sample (e.g. 1 ms for a 500 Hz eye-tracker) for all latency measures. Also note that using a stimulus monitor with, for instance, an 8 ms refresh time will reduce the efficient speed to 125 Hz for these measures, even if a much faster eye-tracker is being used. Historically, cathode ray tube (CRT) monitors have lower latencies and refresh times than the more modern thin film transistor (TFT) monitors. Measuring your monitor with a photodiode is the only way to make sure that it does not buffer the image, and shows it more seldom and later than your experiment presumes. In the remainder of this chapter, it is assumed that all these technical issues have been addressed or can be neglected, and therefore do not add additional system latencies to the collected data.

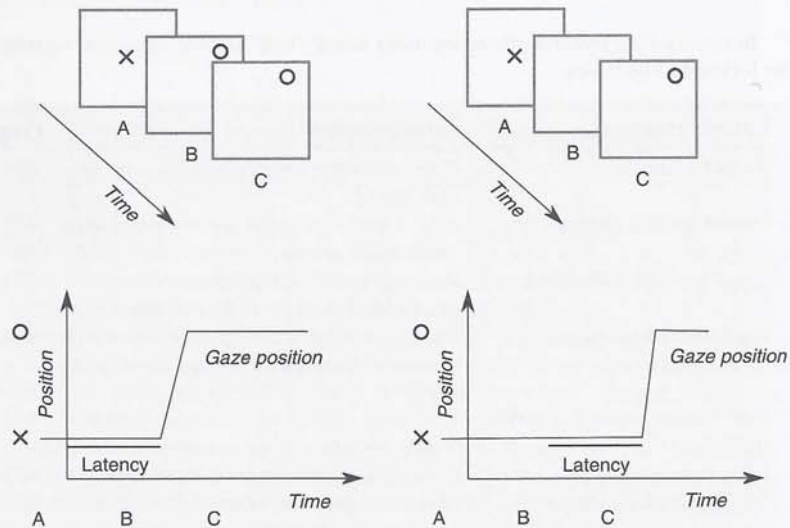
In eye-tracking research, there are many latency and distance measures, summarized in the following two tables.

Latency measure	Target question	Page
Saccadic latency	<i>How soon after target onset does the saccade start?</i>	430
Smooth pursuit latency	<i>How soon after target motion onset does smooth pursuit start?</i>	432
Latency of the reflex blink	<i>How soon after onset of an event which causes blink does the blink commence?</i>	434
Pupil dilation latency	<i>How soon after onset of an event which causes dilation does the pupil start to dilate?</i>	434
Eye fixation related potential (EFRP)	<i>How soon after the eye started looking at X does the ERP component show?</i>	436
Entry time	<i>How soon after onset is the AOI entered?</i>	437
TX: Thresholded entry time	<i>How soon after onset have X % of participants visited the AOI?</i>	438
Proportion of participants over time	<i>What proportion of the participants look or have looked at an AOI at a specific point in time?</i>	440
Eye-voice latency	<i>How soon after the eye started looking at X does the participant verbalize X?</i>	442
Eye-hand span	<i>How soon after the eye looked at X does the hand perform the corresponding action?</i>	445
The eye-eye span (cross-recurrence analysis)	<i>How soon, on average, does a listener look where the speaker looks?</i>	447

Distance measure	Target question	Page
Eye-mouse distance	<i>What is the distance between the point of gaze and the mouse position?</i>	448
Disparity	<i>What is the distance between the points of gaze of left and right eye?</i>	449
Smooth pursuit gain	<i>What is the velocity ratio between point of gaze and the target?</i>	450
Smooth pursuit phase	<i>How far behind or ahead is the eye with respect to the target?</i>	451
Saccadic gain	<i>What is the distance between saccadic ending point and target?</i>	452

### 13.1 Latency measures

Latencies measure the difference in *time* between two events, for instance from the onset of a stimulus until the onset of the first saccade. The majority of latency measures operate with absolute time in milliseconds, but some of the longer-duration measures—for instance entry time—alternatively count time in units of number of fixations or number of unique AOIs.



**Fig. 13.1** On the left side, the *principle of saccadic latency*: A fixation point (X) holds the participant's gaze until onset of target point (O) between time A and B. In the *gap* condition to the right, the fixation point (X) is removed some time before the target point (O) is onset, and for a period called the gap, there is no item that could lock the participant's attention.

### 13.1.1 Saccadic latency

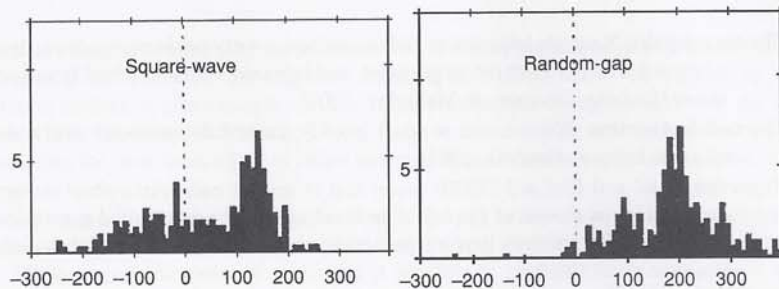
Target question	<i>How soon after onset does the saccade start?</i>
Input representation	<i>A saccade and an onset time of the stimulus</i>
Output	<i>Latency (ms)</i>

Saccadic latency is a measure of reaction time to stimuli. It is defined as the time it takes for the brain to program and launch the saccade to the onset target point. Since saccades have very sharp and easily pinpointed onsets, it is relatively easy to calculate the latency accurately.

The specific event that triggers the saccade is known as the 'onset' of a stimulus. The onset can be a sudden flash in the periphery, or a complete change of a stimulus picture. Saccadic latency is a widely used measure, and there are several named experimental paradigms that differ in how stimuli are presented. The most well known is the 'gap' paradigm (see also p. 68) in which visual attention is released for a short while (the gap) between the offset of a fixation point and the onset of a saccade target; commonly an empty white screen is presented within this interval. Figure 13.1 illustrates both the basic and 'gap' conditions. Saccadic latencies are also regularly reported from studies in the *anti-saccade paradigm* described on pages 305–307.

Simultaneously flashing two scenes with animals or distractors for 20 ms, Rousselet, Fabre-Thorpe, and Thorpe (2002) showed that human participants can reliably make saccades to the side containing an animal within 120 ms (Kirchner & Thorpe, 2006). For comparison, the differential ERP effect starts at 150 ms, while manual button pressing has latencies starting at 180 ms for visual stimuli (Welford & Brebner, 1980).

Saccadic latencies have repeatedly been shown to have a *bimodal distribution*. In monkeys, for instance, there are an abundance of saccades with a latency of 80 ms and also many



**Fig. 13.2** On the right side, saccadic latency histogram for a (random) gap paradigm with one participant. Note the tendency for a bimodal distribution with two peaks. On the left side a predictable square wave stimulus where a large part of the latencies are negative. Bin size 10 ms. With kind permission from Springer Science+Business Media: *Experimental Brain Research*, A short-latency transition of visually-guided and predictive saccades, 76(1), 1989, A.C. Smit.

saccades with a latency of 120 ms, but only few saccades with a 100 ms latency (Fischer & Boch, 1983). The faster saccades in the first peak of this distribution have been named 'express saccades' while the slower ones are thought to be regular saccades. If the superior colliculus, a brain area important to saccades, is damaged, express saccades disappear from the latency distribution. Because of this, express saccades became important in the study of the neural structures that drive rapid oculomotor responses (Schiller & Tehovnik, 2001).

When designing your experiment or interpreting results, take into account the existing research on what makes an upcoming saccadic latency short:

**Releasing attention** Removal of the fixated point—and hence *releasing attention* to it—before onsetting the peripheral target reduces saccadic reaction times from the typical 200 ms down to 120–150 ms. This is known as the 'gap effect', because of the temporal gap between the removal of the fixated object and the onset of the target (see Figure 13.1). The opposite is sustained attention to the fixated point by showing it even after the target has been onset, which prolongs latencies.

**Distractors** When the stimulus image contains a simultaneous distractor, saccadic latencies will increase by 20–40 ms compared to trials where no distractor is present. If the distractor is instead flashed as little as 50–100 ms before the target, latencies are reduced. The distractor then works as a facilitating warning signal (Born & Kerzel, 2008).

**Splitting attention** Directing a saccade to one hemifield (i.e. one half of the visual field) and performing a manual reaction to an object in the opposite hemifield—*splitting attention* in two directions—results in prolonged saccade latencies as well as manual reaction times (Shepherd, Findlay, & Hockey, 1986).

**Anticipation** When the task is sufficiently predictable that it allows the participant to anticipate the direction and amplitude of the next saccade, the latency may be negative, i.e. the saccade shifts the eye before the target appears (Smit & Van Gisbergen, 1989).

**Type of stimulus** Natural stimuli elicit much lower saccadic latencies than classical lab stimuli consisting of a black background with a few white dots (Trottier & Pratt, 2005; White, Stritzke, & Gegenfurtner, 2008). This suggests that the second latency peak in the bimodal distribution, at 180–200 ms, may be a laboratory artefact, and the 'express saccades' with a latency of 120–150 ms the natural latencies that occur with everyday stimuli and environments.

**The target point** Saccade latencies to Gabor patches, a wavy luminance pattern, decrease as a function of contrast at the target point, and increase with the spatial frequency of the waves (Ludwig, Gilchrist, & McSorley, 2004).

**The task instruction** When asked to react quickly, participants respond with a decreased saccadic latency (Kapoula, 1984).

**Expertise** Land and McLeod (2000) found that in cricket batsmen, a short latency to the expected bounce point of the ball after bowling, distinguishes good from poorer batsmen. Also, elite shooters have a significantly shorter saccadic latency than controls (Di Russo *et al.*, 2003).

**Participant age** Saccadic latency decreases significantly by around 25 to 60 ms when participant age increases from 8 to 19 years (Salman, Sharpe, Eizenman, *et al.*, 2006). Saccadic gain, however, and peak velocity, remain constant. As participants grow older, latency increases once more (Moschner & Baloh, 1994).

**Impaired processing** Increased saccadic latencies appear with a large number of impairing factors, such as schizophrenia, melancholic depression, alcohol levels above 0.5‰ and total sleep deprivation (Bocca & Denise, 2006; Winograd-Gurvich *et al.*, 2006; Buser, Lachenmayr, Priemer, Langnau, & Gilg, 1996), just to mention a few. Some drugs (e.g. nicotine, amphetamine) appear to *decrease* these increased latencies, but never below the original baseline.

**Transcranial magnetic stimulation** A number of studies have shown that saccadic latencies are prolonged by transcranial magnetic stimulation of the frontal and parietal areas of the brain, which are thought to regulate the control of eye movements. The combined technique then became important in testing neurological models for eye-movement generation (Müri, Hess, & Pierrot-Deseilligny, 2005; Kapoula *et al.*, 2001; Muri, Vermersch, Rivaud, Gaymard, & Pierrot-Deseilligny, 1996; Zangemeister, Canavan, & Hoemberg, 1995; Priori, Bertolasi, Rothwell, Day, & Marsden, 1993).

In other words, to be sure of 'typical' saccadic latencies, your participants' attention should not be unduly locked to a fixation point, the stimulus should be natural and the target salient. He should be around 20 years of age, be well practiced in the task, know that he has to be quick, and not be impaired by neurological deficiencies, transcranial magnetic stimulation, alcohol, drugs, or sleep deprivation.

Release of attention from a target is also studied using the non-eye-tracking paradigm 'attentional blink': Rapidly showing e.g. letters with ~100 ms interval between them, participants are asked to identify a subset of the letters and at the same time report if there is an X in the subsequent 12 letters (Hari, Valta, & Uutela, 1999). As the identification of the first letter locks participants' attention for 200–600 ms, X:s that appear in that period are seldom identified even when appearing in the foveal projection of the participant.

### 13.1.2 Smooth pursuit latency

Target question	<i>How soon after onset does smooth pursuit start?</i>
Input representation	<i>Onset time of smooth pursuit and stimulus</i>
Output	<i>Latency (ms)</i>

*Smooth pursuit latency* (also known as 'latency of pursuit initiation') is a measure of reaction time, where the triggering event is the onset of a smoothly moving target object. Be aware that smooth pursuit initiation consists of two parts: the earliest 100 ms phase in smooth pursuit, so-called 'open loop pursuit', is considered to be a ballistic acceleration towards the

estimated target direction. The second phase, 'closed-loop pursuit', starts before 300 ms after stimulus onset. The target moving away from the gaze point, known as 'retinal slip' owing to target motion, is continuously and much more accurately compensated for in the second phase (Wallace, Stone, & Masson, 2005). There is also an offset latency in smooth pursuit, measuring the time from offset of target movement until the eye starts to decelerate, but this variety is rarely utilized.

Both gap and no-gap paradigms are used. In the gap condition, to release the participants' visual attention, the target disappears for a short period, then reappears, and starts to move.

While the precise onset of a saccade is possible to calculate even with a fairly simple detection algorithm, detecting smooth pursuit onset is much more difficult and latencies have often been calculated by manual inspection: "Latencies for smooth eye movements were estimated visually from the eye velocity records for individual trials" (Lisberger & Westbrook, 1985). Beginning in the mid 1990s, several detection algorithms for smooth pursuit onset appeared, that calculate them slightly differently (p. 178).

Latencies for smooth pursuit onset are around 100–200 ms when the direction and velocity of the target motion are unpredictable to the participant, but as soon as the participant has a possibility to predict the upcoming target motion, latencies may drop to under 0 ms; that is the smooth pursuit motion starts before the target motion (Burke & Barnes, 2006) in an anticipatory manner. Offset latency of smooth pursuit is in the same range as onset latency (Becker & Fuchs, 1985).

The following factors are known to influence smooth pursuit latency:

- Anticipation** When the visual system has information on the future direction of smooth pursuit, latencies are not only shorter, but even negative, i.e. the eye starts to move before the target does (De Hemptinne, Lefevre, & Missal, 2006; Burke & Barnes, 2006).
- Chromatic isoluminant stimuli** Latencies of pursuit initiation are prolonged by 50 ms when the stimulus is isoluminant—that is, luminance of the background equals that of the moving target—and hence appears to move more slowly (Braun *et al.*, 2008).
- Single distractors** When there was a distractor moving in the opposite direction, smooth pursuit latency of rhesus monkeys was prolonged by about 150 ms, Ferrera and Lisberger (1995) found when empirically testing a neurological model for smooth pursuit. However, distractors only affected the latency, not the continued tracking.
- Moving stimulus background** A dotted background—a whole field of distractors—moving in the same direction as the target decreases smooth pursuit latency, and increases it if moving in the opposite direction (Spering & Gegenfurtner, 2007).
- Schizophrenia** Some studies report shorter smooth pursuit latencies in schizophrenia patients, although this effect is debatable and other studies have found the reverse (O'Driscoll & Callahan, 2008).
- Participant age** Sharpe and Sylvester (1978) report longer latencies for older participants (mean 67 years of age) than for younger ones (mean 42), which make age an important factor to take into account when selecting participants and analysing data. Morrow and Sharpe (1993) used a variety of unpredictable smooth pursuit initiations, and noticed an age-related degeneration in smooth pursuit acceleration.

Many studies that use the smooth pursuit latency measure are designed with the explicit goal of understanding the neurological underpinnings of smooth pursuit, and the extent to which they differ from the systems that drive saccades.

**Table 13.1** Data from an early investigation on blink latencies (Rushworth, 1962). Latency values refer to the first reaction recorded with needle electrodes in the orbicularis oculi muscles.

Stimulus	First peak	Second peak
Glabella tap	12–18 ms	25–45 ms
Electrical stimulation of supraorbital nerve	10–13 ms	25–38 ms
Corneal irritation	-	25–40 ms
Loud click next to ear	-	23–33 ms
Bright flash close to eye	30–58 ms	60–90 ms

### 13.1.3 Latency of the reflex blink

Target question	<i>How soon after onset does the eye blink?</i>
Input representation	<i>Onset of blink and stimulus</i>
Output	<i>Latency (ms)</i>

The *latency of the reflex blink* is the time from stimulation onset until eye blink onset, or in some varieties until maximum eyelid closure. Eye blinks occur after a variety of stimulations: bright flashes, auditory beeps, puffs of air, tapping on the glabella (the space between the eyebrows and above the nose), irritation of the cornea, etc. Stimulation of one eye causes blink reflexes in both eyes.

The reflex blink is a reflex and not a reaction, because the initiation of the blink takes place without any involvement of cognitive or deliberate action. The reflex blinks can therefore be much shorter than reaction times such as saccadic latency.

As blinks are movements with a rapid acceleration, detection of blink onset has never been considered difficult. Note however that pupil and corneal reflection eye-trackers detect onset of blink only when part of the pupil is covered, rather than at onset of eyelid movement. Blink measurements have previously been made with needle electrodes in the blink muscles, and with electromagnetic sensors placed on the eyelid.

The reflex blink has a bimodal distribution, and the precise values vary with the type of stimulation. The very low values indicate that the neurological route between stimulation and reaction is short and does not involve cognitive decisions. Table 13.1 summarizes results from the seminal investigation by Rushworth (1962).

Factors influencing the blink reflex latency often relate to basic alertness states. For example, Grillon, Ameli, Woods, Merikangas, and Davis (1991) found that when participants were feeling anxious about the risk of sudden electrical shock, their blink reflex latencies (to acoustic stimuli) were shorter compared to when they felt safe.

### 13.1.4 Pupil dilation latency

Target question	<i>How soon after onset does the pupil start to dilate?</i>
Input representation	<i>Onset of pupil dilation and stimulus</i>
Output	<i>Latency (ms)</i>

*Pupil dilation latency* is defined as the time elapsing between the onset of increased luminance (or other stimulus) and the beginning of pupil dilation. In addition to high sampling frequency and few or no system latencies, pupil dilation latency measurements are benefitted from the use of a high-resolution eye camera with a fixed distance to the eye, a uniformly

luminous stimulus and—if the task allows it—a fixation point that the participant must look at (p. 392)

It has been known since the late 19th century that pupil dilation varies with both light variation and sensory, cognitive, and emotional events (Beatty & Lucero-Wagoner, 2000). Furthermore, the magnitude of change is large with light stimuli, but quite small when the change is due to internal events, down to 0.01 mm. Note that cognitive and affective pupil reactions are *only indirectly linked* to the stimulus or internal state, and the connection is not necessarily causal (Beatty & Lucero-Wagoner, 2000, p. 143). More precisely, small movements due to mental effort and arousal are superimposed onto two larger effects from the light reflex and the accommodation of the lens. To detect these small movements in pupil size—given the much larger sensory movements—it is necessary to average over many trials to reliably tell whether the effect is indeed present. Since pupil size is individual and sensitive to the environment, pupil size is measured against a baseline obtained during some interval prior to onset (Van Gerven, Paas, Van Merriënboer, & Schmidt, 2004). If the pupil is already involved in a dilation movement, latencies are longer (220–385 ms) than if it is stabilized before onset (about 180 ms) (Young & Biersdorf, 1954). Therefore always allow time to stabilize the pupil before latency measurements.

Changes in pupil dilation are comparatively slow with respect to onset, and as with smooth pursuit latency, difficult to pinpoint exactly. Small-magnitude changes are less easy to detect than the large ones owing to changes in light. Detection by manual inspection of the pupil dilation curve was the only available alternative until the mid 1900s. Young and Biersdorf (1954); Lee, Cohen, and Boynton (1969) used curve-fitting techniques to model the dilation curve, and defined the latency as the time delay giving the most accurate fit. Later studies define the moment of onset as the time when *pupil acceleration is maximally negative*. Onset is only looked for *inside a finite time window* of for instance 200–450 ms after stimulus onset, so that variations other than the reflex are excluded. According to Bergamin, Zimmerman, and Kardon (2003) “This time window is also very easy to define with software analysis ... and also showed greater diagnostic power than the most commonly used contraction” (p. 110).

The following summary of reported latencies to different stimuli describes the overall trend:

**Light** Latencies to light are about 150–400 ms for control participants. Age, several illnesses, personality disorders and drug abuse strongly alter the latency, and hence much clinical research as well as practical diagnosis use pupil dilation latency as one diagnostic criterion.

**Auditory signals** Pupil dilation latency is around 600 ms when judging single tones (Beatty, 1982).

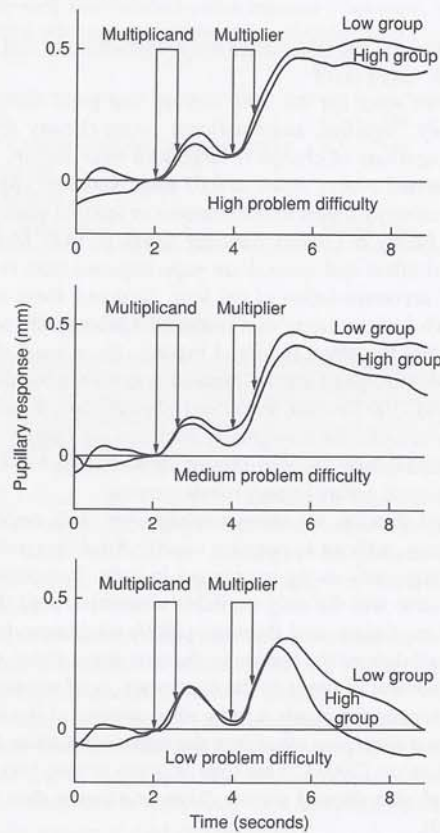
**Pain** Chapman *et al.* (1999) found that pupil dilation responses to pain begin at 330 ms, and peak at 1.25 seconds after stimulus onset.

**Mental multiplication** When Ahern and Beatty (1979) presented spoken integers to be multiplied, the pupil reacted within less than 300 ms, and the reaction was stronger for more difficult multiplications and also stronger for students weaker in mathematics, as shown in Figure 13.3.

**Social stimuli** Latencies reach 600–800 ms when control participants react to the emotional state—and in particular the pupil—of another person (Harrison *et al.*, 2006).

Bearing the above findings in mind when working with pupil dilation latency, also note that inter-eye asymmetry in control participants may range between 8 and 35 ms (Bergamin, Schoetzau, Sugimoto, & Zulauf, 1998).





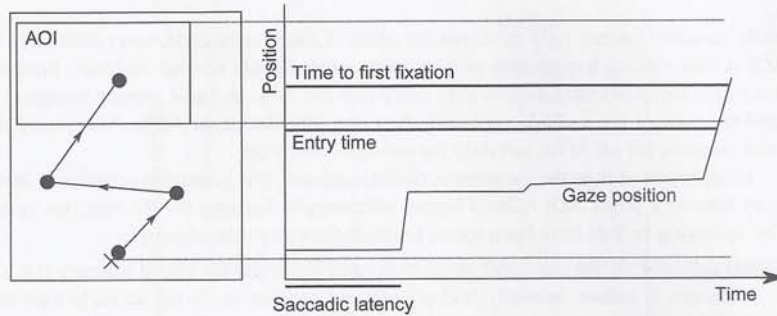
**Fig. 13.3** The pupillary response to spoken integers that were to be multiplied by the participant, the on- and offset of which are shown. Three levels of difficulty were used, the most difficult task at the top. The two groups of participants are either high or low in psychometric measures of intelligence. Overall latencies from onset of multiplicands and multipliers to pupil response are around 300 ms. Reprinted from Ahern and Beatty (1979) with kind permission from the AAAS (American Association for the Advancement of Science).

### 13.1.5 EFRPs—eye fixation related potentials

Target question	<i>How soon after the eye started looking at X does the ERP component show?</i>
Input representation	<i>Onset of fixation and ERP event</i>
Output	<i>Latency (ms)</i>

Eye-fixation related potential simply means that the ERP component in question is *related to the onset of an eye-fixation* (the EF in the acronym) rather than to the onset of the stimulus per se (p. 288).

This is a relatively new measure that arises from the combination of technologies for eye tracking and measuring brain activity. Apart from the distinction in onset, an EFRP-latency is just like any other ERP latency, and thus not really an eye-movement measure. Simola,



**Fig. 13.4** Saccadic latency, entry time, and time to first fixation. Entry into the AOI is reached only during the third saccade—left. Saccadic latency is the time until the first saccade is launched—right. Time to first fixation in AOI is always longer than entry time.

Holmqvist, and Lindgren (2009) measured EFRPs to investigate parafoveal preview benefit, and observed a right visual field advantage associated with an occipital EFRP component which differentiated between processing words and non words.

### 13.1.6 Entry time in AOI

Target question	<i>How soon after onset is the AOI entered?</i>
Input representation	<i>Gaze samples or fixations, AOI location, and onset time of stimulus</i>
Output	<i>Entry time in AOI (ms) or number of events until AOI entry</i>

The term *entry time* is closely related to, but not synonymous with 'search time' in Rötting (2001), 'time to first hit' in Krupinski *et al.* (2003), 'distraction time' in Casey and Richards (1988), and 'time to first fixation on target area of interest' in Jacob and Karn (2003). Entry time is defined as the duration from onset of stimulus until the AOI is first entered, whether entry is made via a saccade or via smooth pursuit. Time to first fixation includes the time period from entering the AOI until the first fixation is made, and is therefore always somewhat longer (Figure 13.4).

As time can be represented either at sample level, or as given by the sequence of a selected event, the precise operational definition of the duration from trial start to first entry has three varieties, each of which can be absolute or relative:

1. *Time in milliseconds* was used by Giorgetti, D'Amato, Pagani, Cavarzeran, and Tagliabracchi (2007) for studying the effects of alcohol. The entry time value can also be made relative to the trial duration.
2. *Number of fixations* from trial start until first entry, used by Bojko (2006) for studying web page design, and Kundel and La Follette Jr (1972) in studying experience of radiologists. The number of fixations until entry can also be divided by the total number of fixations in the trial to give a relative value.
3. *Number of unique AOIs* visited before first entry, or this absolute number divided by the total number of unique AOIs visited during the trial. The latter method thus results in a relative value.

Entry time is a latency measure, that can be used in reaction time studies, but it differs from the other latency measures, for instance saccadic latency, in a number of ways. First,

while saccadic latency only concerns the onset of the first saccade, entry time runs until the AOI is first visited, irrespective of how many saccades are needed en route. Second, while entry time measures the duration until entry into the AOI, saccadic latency measures only up until the start of the saccade (not until the entry into the target AOI). Also, entry time is a valid measure for all AOIs, not only for one selected target.

Irrespective of how the measure is operationalized, it is generally considered that a short entry time to a target AOI reflects higher efficiency in locating for the stimulus in question. The following factors have been found to affect the entry time measure:

**Visual saliency** In the low-level sense of Itti and Koch (2000), visual saliency (i.e. discontinuities in colour, intensity, and orientation) appears *not* to attract early fixations *when the participant has a specific purpose* when inspecting an image (Underwood, Foulsham, Van Loon, Humphreys, & Bloyce, 2006). These authors' conclusions are based on the fixation count definition of the entry time measure. In fact, there is evidence that saliency does a poor job of attracting early attention even during more neutral tasks when there is no specific purpose in mind (Nyström & Holmqvist, 2008).

**Out of context objects** Objects that violate the 'gist' of a scene, that is they have a low probability of appearing in this context, are likely to attract early fixations (Loftus & Mackworth, 1978). However, this finding has been disputed in later work (De Graef *et al.*, 1990).

**General conspicuity** When the target AOI is made to stand out this leads to shorter entry times. For instance, changing the warning labels on cigarette advertisements to be more visible and direct decreases entry time significantly (Krugman, Foxer, Fletcher, Fischer, & Rojas, 1994). Also, redesigning a web page to highlight important elements can lead to significantly shorter entry time to target links (Bojko, 2006).

**A preview or memory of the scene** Previews lead to faster target fixation (short entry times in target AOIs), according to Hollingworth (2009), even for short preview presentations of 500 ms during which the target is absent.

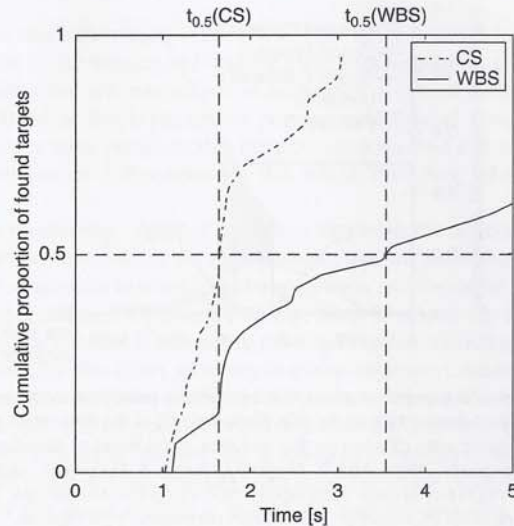
**Alcohol** Ethanol consumption slows down entry time for at least 150 minutes after intake, compared to controls (Giorgetti *et al.*, 2007).

**Expertise** Experts in quickly and accurately identifying pathologies in radiological images can be identified by their lower entry times (Krupinski, 1996; Nodine, Kundel, Lauer, & Toto, 1996). Higher age, often correlated with less experience with user interfaces leads to longer entry times for the interface in question, in comparison to the shorter entry times of younger users with more experience (Obrist, Bernhaupt, Beck, & Tscheligi, 2007).

**Spoken language** Hearing speech very often leads the listener to earlier entries into the corresponding semantic element to which the speech refers. First utilized by Eberhard *et al.* (1995) to measure point-of-disambiguation effects, this use of the entry time measure has now developed into proportion over time-curves (pp. 197–205 and p. 440).

### 13.1.7 Thresholded entry time

Target question	<i>How soon after onset have X% of the participants looked at the AOI?</i>
Input representation	<i>Gaze samples or fixations, AOI location, and onset time of stimulus</i>
Output	<i>Latency (ms)</i>



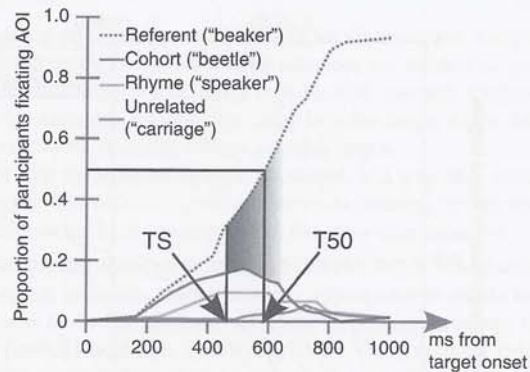
**Fig. 13.5** Cumulative proportion of targets found for two groups, one group with Williams Beuren syndrome and a control group. With a 5 second trial time, only 67% of the WBS participants found the target, but 99% of the CS participants. T50 are shown for both curves. Reprinted from *Neuropsychologia*, 45(5), Montfoort, I., *et al.*, Visual search deficits in Williams-Beuren syndrome, 931–938, Copyright (2007), with permission from Elsevier.

While previous latency measures concerned single events, the *thresholded entry time* is the time (in milliseconds) until a specified proportion ( $X$ ) of all participants have looked at a particular AOI. The T50, for instance, is the time that it takes until 50% of the participants have looked at an AOI. The  $X$ , thus, signifies the threshold, and can be adjusted according to the purpose of the study.

The originators of the measure, Montfoort, Frens, Hooge, Lagers-Van Haselen, and Van Der Geest (2007), point out, however, that calculating average entry time on data where not all participants actually found the target can be misleading. For the data in Figure 13.5, the average entry times for both sets would be very similar. This is why it is advisable to use the median rather than the mean, and thus compensate for incomplete data sets within trials.

When a single participant does many trials, T50 is a better measure than average entry time for the trials where the participant found the AOI, because T50 yields performance (percentage correct) and timing in one number. When multiple participants do a single trial, T50 is equivalent to a cumulative proportion over time graph (p. 197 and below), as it uses a curve delineating the cumulative proportion of AOIs inspected within a fixed time period.

As Figure 13.5 illustrates, the T50 method has been used to identify the visual search processes underlying inefficient visual scanning in Williams Beuren syndrome (Montfoort *et al.*, 2007).



**Fig. 13.6** Two common types of latencies in a momentous proportion over time graph: T50 is the time when 50% of the participants look at the AOI 'Referent'. TS is the time when the curves for 'Referent' and 'Cohort' are significantly different for the first time since onset of stimulus. Modified from *Journal of Memory and Language*, 38(4), Paul D. Allopenna, James S. Magnuson, and Michael K. Tanenhaus, Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models, pp. 413–39, Copyright (1998), with permission from Elsevier.

### 13.1.8 Latency of the proportion of participants over time

Target question	<i>At what time has a specific proportion of the participants looked at an AOI?</i>
Input representation	<i>Proportion over time graphs</i>
Output	<i>Latency (ms)</i>

A proportion over time graph (pp. 197–205) visualizes how the proportion of participants that look at each AOI develops over time. It is used as a measure of the timing and development of cognitive processes over time. It is a relatively complex measure, but also an intuitive visualization. It is often a very sensitive measure, which can be exploited to the researcher's benefit. Typically, this takes the form of a presentation of the visual stimuli, and then introduces some task-critical information which causes the participant to start allocating attention in a new pattern and thus changing the proportion over time curve. We now have a gaze pattern before the critical information bit as a baseline, and we can contrast this to the gaze following the introduction of the critical information. The researcher now has both baseline and treatment data for every unique trial.

Different properties of this curve change can then be extracted and used as predicting variables, but as Figure 13.6 illustrates, it is typically the latency—the time from onset to when a given amplitude is reached (T50 in Figure 13.6), or until a significant difference between two AOIs or two conditions are reached (TS in Figure 13.6)—that is focused on.

#### Statistical analysis of proportion over time curves

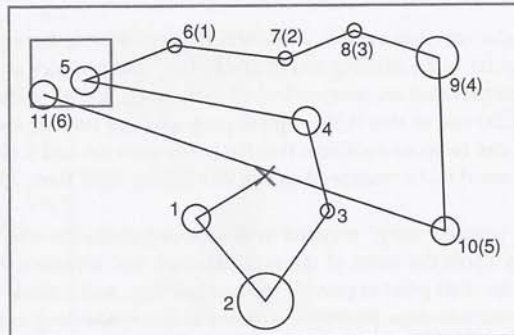
The analysis typically starts with some form of preprocessing of the data. Often the continuous time line gets divided up into several 'bins' which will then be used as a predictor, depending on the particular analysis method. There are two reasons for this binning of time. The first is that the data become more manageable with only a few bins rather than raw data samples. This may also be a requirement by some statistical tests where time cannot be used as a continuous predictor. The second reason is for fulfilling any assumptions of indepen-

dence. A particular raw data sample at point  $t_1$  is very likely to have similar properties to a gaze sample at point  $t_2$ . By binning and treating many data samples as a whole, we are more likely to have samples that are independent of each other. A rule of thumb would be to have bins of at least 200 ms, as this is the typical programming time of a saccade. So, with 200 ms or more, we can be more confident that the participant has had a chance to reallocate his attention if he wanted to, i.e. independence in this setting. (See Barr, 2008 for a discussion of this and more).

The second 'preprocessing' we need to do is to determine the window of analysis. Most often we already know the onset of the manipulation, and we know that we need at least a window before the onset point to provide the baseline data, and a window after the onset point to provide the treatment data. However, we need to know how long our expected effect is to capture it in an analysis window. If we select a too small window, the effect may peak after the window and we may miss the effect, resulting in a false negative. Contrarily, if we select too large a window, the effect will have dissipated before the window ends, and we include more noise than necessary, also risking a false negative as a result. The approach most commonly encountered involves two parts. The first part is selecting a very large baseline window and a very large treatment window. Then, the data from these curves are collapsed and a proportion over time curve for this grand average is plotted. As both baseline data and treatment data are included in this curve and the researcher does not know which hump in the line comes from what data, it is assumed to be safe to visually inspect the curve and determine window intervals from this. Additionally, it is common to divide up the treatment window into several bins, for example one bin covering the onset of the critical manipulation, a bin to cover the growth of the effect, and a bin to cover the peak of the effect. If the decline or dissipation of the effect is also of interest, then further bins can be used after the peak of the effect.

In the past, ANOVA has been used for the analysis of proportion over time curves. In short, this analysis consists of calculating for each bin whether there is a significant difference between the average proportions of participants who looked at, for instance, the target AOI, and the proportion of participants who looked elsewhere. However, there are several reasons why this analysis is not optimal for this design. The choice of bin size, for instance, is an arbitrary choice, but smaller bins imply more analyses, thereby affecting the experiment-wise error rate (p. 94). Furthermore, proportions are bound by the limits of 0 and 1, which may result in violation of the underlying assumptions of the ANOVA. By definition, proportion over time curves imply that the dependent variable is a dichotomous variable: the participant either looked at an AOI or not. Furthermore, the independent variable is time, which is a continuous predictor. This design makes an ANOVA not well suited for the analysis, since it is assumed for the ANOVA that the dependent variable is continuous, and the independent variables are nominal. Here we briefly describe an alternative method that can be used for the analysis of proportion over time curves: multilevel logistic regression.

The most promising analysis method today for analysing AOI proportions over time, is multilevel logistic regression. Logistic regression is well suited, because it is assumed that the dependent variable is a dichotomous variable, and because it allows predictors to be continuous as well as categorical. The multilevel analysis allows control of variation that is due to random factors: participant variation and item variation. Furthermore, the multilevel model allows the error to be auto-correlated (Singer & Willett, 2003), which eliminates the problem of dependent measurements. A final advantage of multilevel modelling is that the shape of the development over time may be included in the model by using time as a continuous predictor. In other words, it is possible to evaluate whether change over time is linear or curvilinear, avoiding the need divide up data into bins (except for problematic cases with much zero data). For illustrations of the use of multilevel modelling for the analysis of proportion over time curves, we refer to Barr (2008) and Mirman, Dixon, and Magnuson (2008).



**Fig. 13.7** Two operational definitions of return time using index number of fixations. The fixation sequence starts at the fixation cross in the middle. Fixations are numbered from the start of the trial, and in parentheses, from the end of the last dwell in the AOI. Return time could therefore be fixation 11 or 6 respectively, depending on the definition used. Larger circles signify longer fixation durations.

### 13.1.9 Return time

Target question	<i>How soon after onset, or after the previous dwell to an AOI does the eye return to it?</i>
Input representation	<i>Gaze samples or fixations, onset of change, and location of the AOI</i>
Output	<i>Return time (ms)</i>

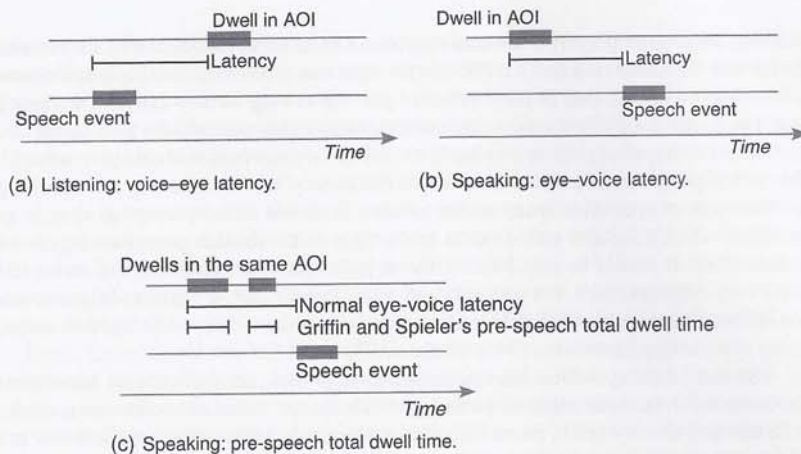
The *return time* is the time it takes until a participant returns to an AOI already looked at. Ryan and Cohen (2004) operationalize this duration as the index number of the first returning fixation to the AOI. Another possible, but apparently unused operational definition, is to count the number of fixations since the last dwell on an AOI until a return to the same AOI takes place. For both these operational definitions, time can be calculated as the number of fixations, or as real time, which would take into account the processing duration, and not only sequential order.

Independently of definition, this measure is particularly important when the AOI is manipulated or changed during a trial, but is also very revealing when returns to an AOI are a fundamental part of a diagnostic or problem solving task, where return time can serve as a measure of working memory.

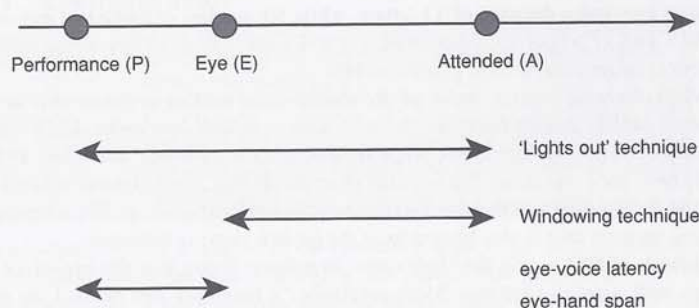
### 13.1.10 Eye-voice latencies

Target question	<i>How soon after the eye started looking at X does the participant verbalize X?</i>
Input representation	<i>Gaze samples or fixations, location of X, and onset time of verbalization</i>
Output	<i>Latency (ms), number of words, letters, or notes</i>

*Eye-voice latency* is an umbrella term for a whole family of measures, which all measure the duration between onset of speech and entry into an AOI. Figure 13.8 (a) and (b) shows the two basic measures, of which one is used to study listening to speech and the other to study verbalising pictorial material or written text, respectively. Other common terms for the same



**Fig. 13.8** In (a) and (b), the two basic eye-voice latencies. In both cases the speech event is semantically equivalent in meaning to the AOI in which the dwell is made. In (c), a variety of eye-voice latencies based on total dwell time.



**Fig. 13.9** Eye-voice latency is the time (or distance) from the eye (E) until the performance (P). The lights-out technique measures a much larger span, from performance at the time of lights-out, until the further end of what has been attended (A), or what is known as the perceptual span. For comparison, the moving window technique measures the size from the eye (E) until (A) (see page 50 for a full description of the moving window technique).

measure are 'eye-voice span' (Meyer & Lethaus, 2004) and 'first fixation time [relative to noun phrase onset]' (Brown-Schmidt & Tanenhaus, 2006).

In the majority of studies, latency is counted from onset to onset. Of course, latencies are positive when participants are speaking, and tend to be negative when they are listening to speech. There is one notable alternative operational definition, however: Griffin and Spieler (2006) only count the total dwell time on the AOI, not including the time looking at other objects. This 'pre-speech total dwell time' variety of eye-voice latency, exemplified in Figure 13.8(c), focuses on the *amount of processing* of the AOI before speech starts, while classical eye-voice latency measures the *overall time elapsed* until speech is produced.

Beware that some studies of eye-voice latencies (and eye-hand span, covered next) do not use eye-movement recordings at all, but rely on what is known as the 'lights-out' technique (Levin & Kaplan, 1970). Here, turning out the light unexpectedly while participants are



reading, singing, or playing a musical instrument from sheet music, allows the researcher to gauge how far ahead they had read before the light was turned out. Participants are instructed to continue to speak, sing or play without light, for as long as they can, and without guessing. The lights-out method returns the duration from lights-out until the participant ceases to verbalize (or stops singing or playing if the study is concerned with music reading). Thus, this technique gives an indirect measure of the latency between reading and speaking. It is another type of eye-voice span, related to what is known as the perceptual span in reading research, which measures visual intake to the right of the fixation point (see Figure 13.9 for comparison). It should be noted that different techniques for measuring eye-voice latencies yield very different results. For instance, eye-voice latency during sight-reading is seven notes for skilled readers using the lights-out technique (Sloboda, 1985), while it is four notes when using eye tracking (Jacobsen, 1941; Goolsby, 1994).

The unit of the eye-voice latency measure varies between applications: latencies in milliseconds are the default unit with pictorial stimuli for eye-voice and voice-eye studies, while with text and music stimuli, latencies are also measured in number of words, letters or notes. In a few studies, latencies are reported in pixels within the stimulus.

When participants *read texts aloud*, the main finding is that eye-voice latencies hover around 800 ms. More experienced, better readers, have *larger spans* (Buswell, 1920). Buswell found that the eye-voice latency of high-school students is approximately 0.79 s, corresponding to an eye-voice distance of 13 letters, while 5th graders exhibited an eye-voice latency of 0.91 s and 11 letters. Correspondingly, good readers have an eye-voice latency of about 4–5 letters larger compared to poorer readers.

When studying *singing*, most of the studies have looked at either *skill* or *the musical structure*. Skilled singers have an eye-voice span of around four notes, while unskilled show a latency of on average two notes, when sight-singing to melodies (Jacobsen, 1941). Goolsby (1994) finds the same result, but adds the observation that “skilled music readers look farther ahead in the notation and then back to the point of performance” (p. 77), a strategy to refresh working memory that is also known from the picture viewing literature.

Sloboda (1985), using the ‘lights-out’ technique, found that the eye-voice latency coincides with *musical phrasing*. More precisely, “a boundary just beyond the average span ‘stretches’ the span, and a boundary just before the average ‘contracts’ it” (p. 72). Musical boundary is the point at which the melody changes. Experienced musicians can cope with larger boundaries, and as this boundary is extended so to is the perceptual span. Good readers, Sloboda found, maintain a larger span size (up to seven notes) than do poor readers (up to four notes).

When singing *prima vista* to notes only, the eye-voice latency is around 500 ms, but the latency increases significantly to 750 ms when text is added to the notes (Berséus, 2002). This difference may be a linguistic effect since we are more accustomed to text, and can therefore retain more in our visual buffer, but it is also likely that it reflects the greater importance of the vertical dimension in eye movements to notes only.

The largest variation in eye-voice latencies has been found in studies of *spoken descriptions of pictures*. Griffin and Bock (2000) conclude that eye-voice latency when describing pictures is around 800–900 ms, while Brown-Schmidt and Tanenhaus (2006) found latencies as low as 600 ms. Both these studies used fairly simple utterances and stimuli consisting of distinct, but visually sparse line-drawings. When using naturalistic stimuli and free discourse, Holsanova (2008) found latencies up to 2500 ms and frequent refixations to the point of performance (speech). In a series of studies of imagery, where participants are looking into empty space, latencies were up to 5 seconds (Johansson *et al.*, 2006). This variation most likely reflects the complexity of the tasks.

A further complication with eye-voice latencies in naturalistic tasks, pointed out by

Holsanova (2008), is that the eye–voice relationships are seldom as simplistic as the upper part of Figure 13.8 suggests. For instance, it often happens that a participant looks twice or more at an AOI before making an utterance referring to it (as exemplified on page 259). The uncertainty over which of the two fixations to select for the calculation points to the need for additional requirements and more refined operational definitions of the measure.

According to Chafe (1994), conscious ideas are constantly activated and deactivated during speech. If the eye is a bit ahead of speech, pushing to activate new conscious ideas from vision, it may happen that ideas that were about to be spoken have already been deactivated (“forgotten”) before they have been spoken. This would appear in data as scanpath patterns where sudden regressions occurred back to the entity in the picture that was just about to be spoken. Such patterns were exactly what Holsanova (2008, 2001) and Goolsby (1994) found.

Bock, Irwin, Davidson, and Levelt (2003) showed analogue and digital clocks and asked participants to tell the time, finding longer latencies for analogue than for digital clocks, reflecting the difference in processing each of the two formats to language. More interestingly, they showed the clocks with two trial durations, 100 ms and 3000 ms, finding that the time and accuracy of the subsequent expression is not different between a 100 ms and a 3000 ms trial, which they argue have implications for models of language production.

### 13.1.11 Eye–hand span

Target question	<i>How soon after the eye looked at X does the hand perform the corresponding action?</i>
Input representation	<i>Gaze samples or fixations, location of hand, and onset time of action</i>
Output	<i>Latency (ms) between look at X and action, distance (pixels), or number of beats</i>

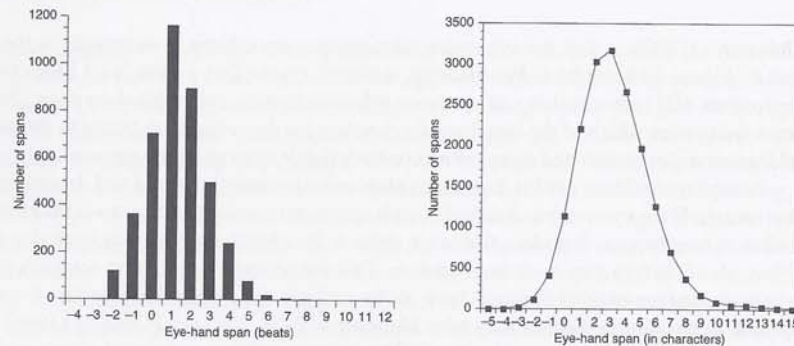
The eye–hand span is a relative of the eye–voice span. It has been frequently utilized in studies of piano players, but also in studying how participants look ahead when following a walkway. For studies of eye–hand coordination, the eye–hand span is a crucial measure.

Formally, the eye–hand span is defined as the duration from the start of a fixation on an item until the hand (or other effector) performs the action associated with the item; such as pressing the correct piano key. Other terms include ‘eye–hand latency’, ‘eye–hand lead time’ and ‘latency of the hand movement relative to the eye’. The unit of the eye–hand span measures also varies. Although time in milliseconds is the most common, eye–hand span is often expressed in number of notes or other task-specific units.

Weaver (1943) found a considerable variation in eye–hand spans during piano playing; a variation of up to eight chords. Distributions from piano players (Truitt, Clifton, Pollatsek, & Rayner, 1997) and typists (Inhoff & Gordon, 1997) show small variation and appear statistically very well-behaved (see Figure 13.10).

Recorded eye–hand spans vary from about –3 to around 1.5 seconds, but the averages depend on a variety of factors:

**Task requirements** Wilmut *et al.* (2006) had participants point at one object after another, and they report eye–hand lead times of approximately 200 ms. With the task requirement to point at sudden onset LED lights, Binsted, Chua, Helsen, and Elliott (2001) report latencies of 300 ms. Patla and Vickers (2003) required their participants to step on a series of irregularly spaced ‘footprints’ over a 10 metre walkway. They found that participants fixated the footprints on average two steps ahead, or 0.8–1 s in time. When playing the piano, eye–hand spans of 3.1 notes (0.78 s) (Weaver, 1943), a single beat



(a) A histogram of eye-hand latencies during piano playing. From Truitt *et al.* (1997), and reproduced with kind permission from Psychological Press.  
 (b) A histogram of eye-hand latencies when copy typing. From Inhoff and Gordon (1997), and reproduced with kind permission from Blackwell Publishing.

**Fig. 13.10** Distributions of eye-hand latencies (spans) in two different tasks.

or 450 ms (Truitt *et al.*, 1997), and up to 1.3 seconds (Furneaux & Land, 1999) have been reported in the literature. When playing the violin, the eye-hand span is around 1 second (3–6 notes), with a larger variance for a sonata by Telemann than a sonata by Correlli (Wurtz *et al.*, 2009), again highlighting the influence of task *even within a task*. When typing, Butsch (1932) found that typists of all skill levels have an eye-hand span of about 1 second or around five characters.

**Skill** Already Jacobsen (1941) found a difference in skill, with eye-hand spans during piano playing ranging from zero up to four notes. Later studies have found that skill in piano playing increases eye-hand span when measured in number of notes, but not in terms of time (Furneaux & Land, 1999), which would mean that although processing time is the same, throughput rate differs. This means that more information is processed in a fixed time interval—the span is increased, but the time remains the same, therefore greater efficiency.

**Repetition** Epelboim *et al.* (1995) showed a substantial reduction in eye-hand spans when participants repeated the task of tapping a sequence of lights.

**Non-repeated whole-body tasks** Land *et al.* (1999) found eye-hand latencies of about 530 ms (ranging from 430 to 680 ms) when participants were engaged in the everyday task of making a pot of tea.

**Musical tempo** A slower tempo increases the eye-hand span in piano playing to 1.3 s, while fast tempos reduce it to 0.7 s (Furneaux & Land, 1999).

**Target visibility** When the target is continuously visible, eye-hand latencies are reduced (Abrams, Meyer, & Kornblum, 1990).

**Allocation of attention** Frens and Erkelens (1991) showed that eye-hand coordination is disrupted by auditory distractors presented at the initiation of the reaching movement. This result has been interpreted as evidence for a single central attentional control of both movements.

**Look-ahead fixations** Following a look-ahead fixation, the eye departs earlier relative to the hand, increasing the eye-hand latency from on average 2.7 to 3.1 seconds (Mennie *et al.*, 2007).

A theoretical interpretation of the eye–hand span is that its size reflects a continuous ‘push–pull’ relationship between two forces: one, the need for material to be held in *working memory* long enough to be processed into musculoskeletal commands, and second, the need to limit the demand on span size and therefore the workload in the memory system. For most readers, the need to limit the workload of the memory system prevails, and this results in the very small spans that are mostly found in studies of eye movements in reading (Rayner & Pollatsek, 1997).

### 13.1.12 The eye–eye span (cross-recurrence analysis)

Target question	<i>How soon does a listener look where the speaker looks?</i>
Input representation	<i>Gaze samples or fixations from listener and speaker AOI</i>
Output	<i>Latency (ms) or distance (pixels)</i>

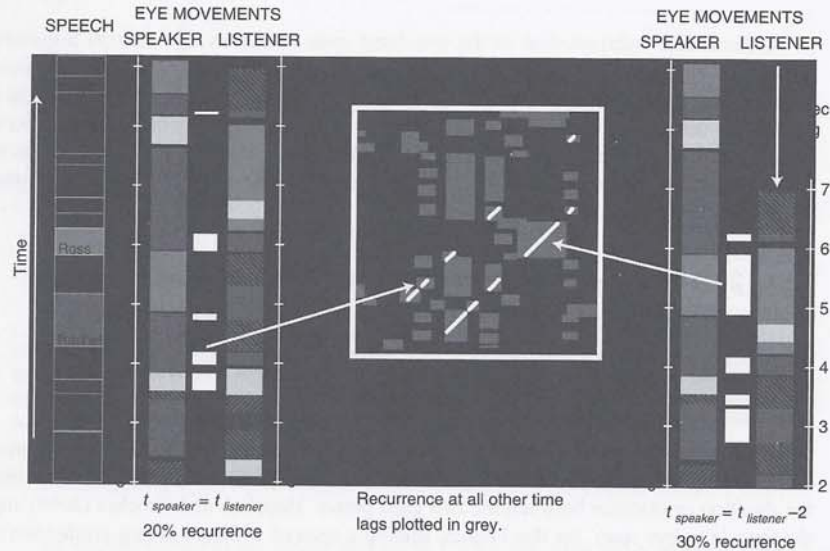
Only scattered studies have looked at collaborative tasks or face-to-face communication where both participants have been eye-tracked. In such studies, there is a need to measure the duration or distance between the two gaze points. Hadelich and Crocker (2006) suggest the term ‘eye–eye span’ for this latency, finding a span of 1400 ms during production of referring expressions. The Euclidean distance between two participants’ gaze at a frame on a video stimulus is an alternative variety of eye–eye span (p. 370). Reported values in literature vary. In a collaborative building task, Velichkovsky *et al.* (1996) note a lag of only 500 ms. In a collaborative web-based task, Cherubini, Nüssli, and Dillenbourg (2008) find that utterances not followed by a repair act (such as “I did not understand that last thing you said”) resulted in an average eye–eye distance of 85.65 pixels, while messages followed by a repair act with an average distance of 231.37 pixels.

Cross-recurrence analysis is a mathematical method to investigate the overall latency by visualising and quantifying recurrent patterns of states between two time series (essentially two eye movement data files). This type of analysis was introduced into eye-movement research by Richardson and Dale (2005), who found that a listener follows the gaze of a speaker with a latency (or lag) of around 2 seconds on average. Figure 13.11 shows the principle for calculating this latency. In short, the eye–eye latency calculated from a cross-recurrence analysis is the speaker’s eye–voice latency plus the listener’s voice–eye latency, of which there can indeed be several. The strength of this method is that it allows us to estimate the latency with the best overall fit.

Success in communication is reflected in the co-alignment of gaze between speaker and listener, Richardson and Dale argue, after showing that listener comprehension varies with the listeners ability to align with speaker gaze.

## 13.2 Distances

Distance measures compare the simultaneous spatial positions of two separate entities, for instance eye to stimulus point, left eye to right eye, or eye to mouse position. We have already seen how the same measure can have both time and space for units, for instance the eye–hand span. Under this heading, we present measures that are predominantly spatial.



**Fig. 13.11** Calculation of cross-recurrence plots and associated latencies may appear complex, but are actually fairly intuitive. To the left, speech and speaker and listener eye-movements as scarf plots in real time. We can recognize the eye-voice latencies in the speaker eye-movements and speech, for instance 'Rachel' is mentioned around 400 ms after the speaker looks at her. The listener in turn looks at the 'Rachel' AOI some 500 ms later. As the speaker dwells on the Rachel AOI for some time, it will overlap with listener dwell on the same AOI. Overlaps are marked as black in between the scarf plots. The cross-recurrence plot is produced by placing the scarf plots of speaker and listener on the x- and y-dimensions (called  $t_{speaker}$  and  $t_{listener}$ ), and each (x,y) on the surface is marked grey if both speaker and listener look at the same AOI. Now, diagonals in the plot correspond to alignment over time. One of the black-marked diagonals correspond to the overlaps between speaker and listener in real time. The other marked diagonal corresponds to a shift of the listener data in time by two seconds, corresponding to the scarf plots on the right side of the figure. The shift in time—the 'lag'—producing the longest diagonal is the eye-eye latency with the best overall fit. With kind permission from John Wiley and Sons: *Cognitive Science*, Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension, 29(6), 2005, Daniel C. Richardson and Rick Dale, pp. 1045-60.

### 13.2.1 Eye-mouse distance

Target question	What is the distance between the point of gaze and the mouse position?
Input representation	Gaze and mouse position
Output	Distance (pixels)

*Eye-mouse distance* measures Euclidean distance in number of pixels or visual degrees between the position of the mouse cursor (in a computer monitor) and the gaze position. Rodden and Fu (2007) present a histogram of eye-mouse latencies from a variety of Google search tasks (Figure 13.12). Some of the mouse-eye research has attempted to show that recording mouse movements provides comparable data to recording eye movements (e.g. Granka, Feusner, & Lorigo, 2008); however, these studies have varying results, finding only some coarse correlations.

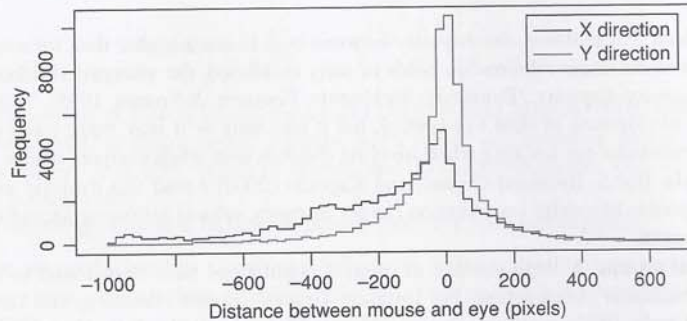


Fig. 13.12 Horizontal and vertical eye–mouse distances. Although both are centred around zero, there is a large spread, in particular in the horizontal direction. Reproduced from Rodden and Fu (2007).

Of the very few studies that report eye–mouse distance, many include alternative operational definitions. For instance, after finding that eye–mouse distance varies from 0.1 to over 1000 pixels (on average 290), Chen, Anderson, and Sohn (2001) decide to make an AOI-based analysis of dwell time for mouse and eye, and finding a fair correlation of 0.58 between the two.

Overall, it does not seem likely that recording of mouse movements can replace eye tracking. Rather, the potential with eye–mouse distance is to investigate the dynamics in the interplay between eye and mouse. Rodden and Fu (2007) argue to have found three strategies that influence the eye–mouse distance. One, keeping the mouse still while reading; two, using the mouse as a reading aid with the cursor to help guide the eye; and three, using the mouse to indicate interesting points to return to. Visually marking an area on a web page that might be interesting to return to, by placing the mouse cursor on it, is a strategy that has repeatedly been found (Ballard, Hayhoe, & Pelz, 1995; Cox & Silva, 2006), as this is the most cost-efficient use of working memory. Smith, Ho, Ark, and Zhai (2000) found three different eye–mouse behaviours when users select a target. One, eye gaze following the cursor to the target; two, eye gaze leading the cursor to the target; and three, eye gaze switching between the cursor and the target until the target is reached. Each of these give different eye–mouse distances over time.

### 13.2.2 Disparities

Target question	<i>What is the distance between the points of gaze of left and right eye?</i>
Input representation	<i>Binocular gaze position</i>
Output	<i>Distance (pixels)</i>

Binocular disparity is the distance between the gaze positions of the left and the right eye.

Note that the values calculated for this distance measure *may* be invalid if the participant's eyes are not calibrated separately (Liversedge, White, *et al.*, 2006, but see Nuthmann & Kliegl, 2009 for counter arguments). Accuracy and precision of the equipment are also of extreme importance when measuring the small distances between the gaze points of the two eyes.

Another line of research investigates variability in disparities between different participant populations:

**Children** For children, the disparity between eyes is much higher than for adults, and this inverse linear relationship holds in early childhood, the youngest children presenting greater disparity. (Fioravanti, Inchingolo, Pensiero, & Spanio, 1995). This reflects the development of dual eye control, but it also tells us it may make more sense to use binocular eye tracking when studying children than when studying adults.

**Dyslexia** Bucci, Brémond-Gignac, and Kapoula (2008) found that dyslexic children have poorer binocular coordination (larger disparity values) following saccadic eye movements.

**Clinical groups** A large number of clinical populations have been found to have poorer binocular coordination. For instance, Graves' disease (Wouters, Van Den Bosch, & Lemij, 1998), multiple sclerosis (Ventre, Vighetto, Bailly, & Prablanc, 1991), cerebellar disease (Versino, Hurko, & Zee, 1996), and deep amblyopia (passive eye) (Maxwell, Lemij, & Collewijn, 1995) can all lead to an impaired ability to direct both eyes equivalently.

Reading research has recently focused on reliably measuring and quantifying disparity; whether the eyes show alignment, crossed or uncrossed disparity. The following factors have been investigated:

**Linguistic and higher-level processing** There is little evidence that binocular coordination can be modulated by higher-level cognitive processing (Kirkby, Blythe, Benson, & Liversedge, 2009; Juhasz, Liversedge, White, & Rayner, 2006; Bucci & Kapoula, 2006).

**Gaze crossing** Some studies have found crossed disparities to be prevalent (e.g. Kliegl *et al.*, 2006 and Nuthmann & Kliegl, 2009), while others have obtained a majority of uncrossed disparities (e.g. Blythe *et al.*, 2006; Juhasz *et al.*, 2006; Liversedge, Rayner, White, Findlay, & McSorley, 2006; Liversedge, White, *et al.*, 2006).

The distance between the two eyes when reading was reported by Heller and Radach (1999) to be 1–2 characters.

### 13.2.3 Smooth pursuit gain

Target question	<i>What is the velocity ratio between point of gaze and target?</i>
Input representation	<i>Gaze and target positions/velocity</i>
Output	<i>Gain</i>

In smooth pursuit research, closed-loop *smooth pursuit gain* is almost always defined as the ratio of eye velocity to target velocity. Thus, a 1.0 value indicates a perfect match, while lower values mean that the eye falls behind, and an increasing number of catch-up saccades are made to compensate for the slowness of the smooth pursuit system. Gain values are typically slightly below 1.0, and tend to fall off at higher target velocities.

However, when using sine wave stimuli, smooth pursuit gain is often measured only at the peak of eye velocity: *Peak gain* is then defined as the ratio of peak eye velocity to target velocity. A third operational definition of smooth pursuit gain is to count the rate of catch-up saccades.

Furthermore, the root mean square (RMS) error quantifies gain as the cumulative distance between the eye and the target during smooth pursuit. Assume that at each recorded data sample  $n$ ,  $\theta_n$  is a measure of the distance between the position of gaze and the position of the target. Then RMS error is defined as:

$$\theta_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \theta_i^2} = \sqrt{\frac{\theta_1^2 + \theta_2^2 + \dots + \theta_n^2}{n}} \quad (13.1)$$

In a meta-analysis of schizophrenia research, O'Driscoll and Callahan (2008) find RMS values to be atypical in schizophrenic patients. RMS values also correlate strongly with both previous subjective quality ratings of pursuit, and with smooth pursuit gain (Gooding, Iacono, & Beiser, 1994).

The following influences on smooth pursuit gain have been investigated:

- Inattention and distraction** Conditions designed to distract attention, or to produce declining arousal and attention, produced reduced smooth pursuit gain (Brezinova & Kendell, 1977).
- Motion direction** For predictable target motions, most participants exhibit higher gain values during horizontal than during vertical smooth pursuit movements (Rottach *et al.*, 1996).
- Age** Smooth pursuit gain is on average 0.7–0.8 for 8–19 year olds and increases significantly with age (Salman, Sharpe, Lillakas, Dennis, & Steinbach, 2006).
- Schizophrenia** A large number of studies have found that participants with schizophrenia have a poorer than normal smooth pursuit gain (O'Driscoll & Callahan, 2008).
- Nicotine** Nicotine intake appears to help the smooth pursuit system of schizophrenic patients: directly after smoking, smooth pursuit gain increases and the number of catch-up saccades decreases significantly (Olincy, Ross, Young, Roath, & Freedman, 1998).
- Alcohol** Smooth-pursuit eye-movements are significantly disturbed by increasing blood alcohol levels, as measured by smooth pursuit gain (Wilkinson, Kime, & Purnell, 1974), even if the participant reports not feeling sedated (Holdstock & De Wit, 2006).
- Barbiturates and benzodiazepines** Smooth pursuit gain decreases and the number of catch-up saccades increase with increasing doses of the drug (Bittencourt, Wade, Smith, & Richens, 1983).

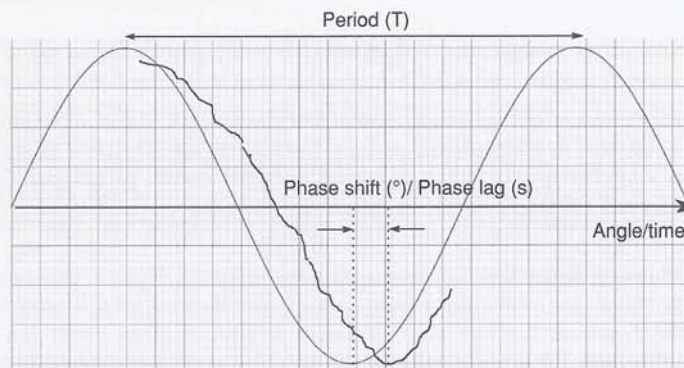
### 13.2.4 Smooth pursuit phase

Target question	<i>How far behind or ahead is the eye?</i>
Input representation	<i>Raw data samples and synchronized movement data for the stimulus.</i>
Output	<i>Phase (ms or degrees).</i>

When presenting participants with sinusoidal stimulus points (such as when tracking the motion of a pendulum), smooth pursuit performance is in addition to gain often evaluated with *phase*. Phase is a “measure of the temporal synchrony between the target and the eye” (Leigh & Zee, 2006). It can be defined either in terms of a spatial (angular) *shift* which measures the difference in phase between the target and eye traces, or in terms of a temporal *lag* which reflects the amount of time the eye is lagging behind the target. Figure 13.13 illustrates these two varieties of phase.

Phase shift and lag are often used interchangeably, and terminology and operational definitions vary across publications. The meta-analytic overview by O'Driscoll and Callahan (2008, p. 361) uses an overarching term lag, which has one spatial variety “phase lag, which quantifies the difference in phase between the eye trace and the target trace” and one temporal variety “temporal lag, which can be quantified as the interval between the moment the target reverses direction and the moment the eye reverses direction”.





**Fig. 13.13** The thin sinusoidal curve indicates the motion of a stimulus point, while the thick line superimposed represents fictitious data of a participant following this point. Phase *shift* and *lag* represent respectively the spatial (angular) and temporal difference between the position of the eye and the position of the target.

Phase has been an important measure of tracking accuracy when studying the visual processes involved in smooth pursuit. In short, phase shift has been reported to stay below  $5^\circ$  but to increase with increasing frequencies (Collewijn & Tamminga, 1984, p. 223). With predictable stimuli, phase shift may disappear although gain remains smaller than one (accounted for due to the presence of catch-up saccades). With unpredictable stimuli, phase lag can increase further (Collewijn & Tamminga, 1984; Goldreich, Krauzlis, & Lisberger, 1992).

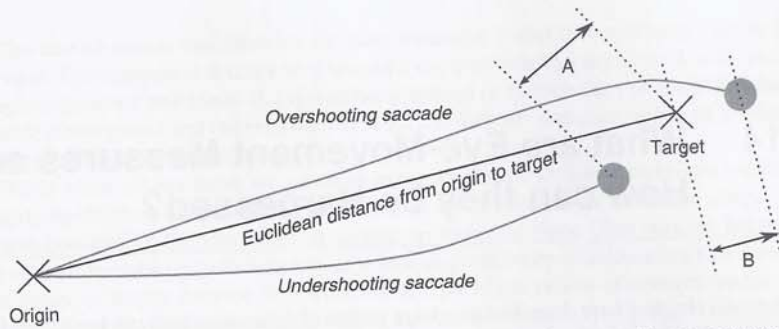
Phase measures are also employed in clinical and neurophysiological research. For instance, a significantly larger phase lag compared to controls has been observed in patients with Parkinson's disease (Bronstein & Kennard, 1985) and unilateral frontal lobe lesions (Morrow & Sharpe, 1995), which could tell us about the role of these brain areas to the predictive component in smooth pursuit.

### 13.2.5 Saccadic gain

Target question	<i>What is the distance between saccadic ending point and target?</i>
Input representation	<i>Saccadic landing position and target position</i>
Output	<i>Gain</i>

*Saccadic gain*, which is also called 'saccadic accuracy', is mostly defined as the initial saccadic amplitude divided by the target amplitude, as exemplified by Figure 13.14. Initial means that we measure only the first saccade to the target, not additional corrections, to which correction saccades count but rarely glissades. Target amplitude is the distance from the saccade starting point to the intended saccadic goal. The saccadic gain unit is %, so less than 100% is sometimes called undershoot (or hypometric), while more than 100% is overshoot (or hypermetric). In an alternative operational definition, saccadic accuracy is instead measured as the Euclidean distance between saccadic endpoint and target.

Note that if the saccade is curved, and saccadic amplitude is calculated as the length along the curve (p. 311), but the distance from origin to target as the straight Euclidean distance, then saccadic gain values will be miscalculated. A perfect hit would be counted as an overshoot, for instance.



**Fig. 13.14** The principle for saccadic gain/accuracy. An overshoot (top saccade) and an undershoot (bottom saccade). Gain is calculated as the amplitude of the saccade divided by the Euclidean distance from origin to target. A and B are the differences in landing point with respect to the target, so-called endpoint deviations.

Nummenmaa *et al.* (2008) define the very related 'endpoint deviation' as the size of the saccadic error: the Euclidean distance from the observed landing position to the intended location, A in Figure 13.14.

Saccadic gain is a measure frequently used in neurological studies of saccadic and movement control. The number of papers using the saccadic gain measure in neurology is very large, and the measure seems to be virtually unknown outside of that field. Ettinger *et al.* (2002) show that the volume of the cerebellar vermis (a brain area thought important in directing saccades) predicts saccadic gain. Saccadic accuracy appears to be unaffected by age (Yang & Kapoula, 2008). Both eye and hand movements typically undershoot the target position, requiring a second corrective movement, or in the case of the eyes, a saccade, bringing the effector to rest on the target (Binsted *et al.*, 2001). In a human factors study, Green and Farnborough (1986) found that severe sleep deprivation reduces saccadic accuracy.