10 Movement Measures

The chapters in the first two parts of this book provide detailed information about the technology and skills necessary to conduct eye-tracking research (Part I), and how to process eye-tracking data after recording (Part II). In Part III, we cover the vast range of measures which can be calculated on the basis of the events and representations described in Part I. Measures of eye *movement* are many and diverse, and this is where we begin.

Eye-*movement* measures, as defined in this chapter, refer to different properties of movement events during a finite period of time. The properties of movement are *direction*, *amplitude*, *duration*, *velocity*, and *acceleration*.

Movement measure group	Uses	Page
Movement direction measures	*In what direction did the eye move?*	301
Movement amplitude measures	*How far did the eye move?*	311
Movement duration measures	*For how long did the eye move?*	321
Movement velocity measures	*How fast did the eye move?*	326
Movement acceleration measures	*How fast did the eye accelerate?*	332
Movement shape measures	*What is the shape of the eye movement?*	336
AOI order and transition measures	*How similar are movements in AOIs?*	339
Scanpath comparison measures	*How similar are two or more scanpaths?*	346

Also, all movements have the more ill-defined property *shape*. These six general properties of movements have generated the measures that are listed in Sections 10.1–10.6.

In Sections 10.7 and 10.8, we classify measures that quantify the order of movement through space: AOI visits and transition sequences between AOIs, and methods to calculate the similarity between pairs of eye-movement sequences (i.e. scanpaths).

Many of the movement measures have a ratio value type, which makes their usage statistically straightforward. However, some measures in the later sections require the use of more advanced statistics.

10.1 Movement direction measures

Movement direction measures pertain to single instances of movement events such as saccades, glissades, drifts, microsaccades, smooth pursuits, and scanpaths. Some but definitely not all of these events move along a straight line, but by no means always and definitely not all. The movement of saccades, glissades, and smooth pursuits can be curved, i.e. altering direction along the event. When we refer to a scanpath's direction we refer to its *overall* direction, if it has any directionality at all.

The resulting values are *direction of movement* (φ) in stimulus space or towards specified AOIs. Direction should not be confused with the angular distance (measured in degrees of visual angle (°)) the eye moves during the saccade, which is referred to as the *amplitude of movement*. The difference between these two parameters is illustrated in Figure 10.1.

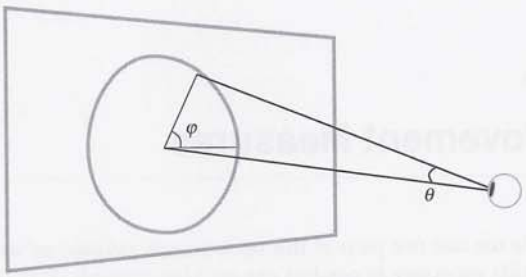
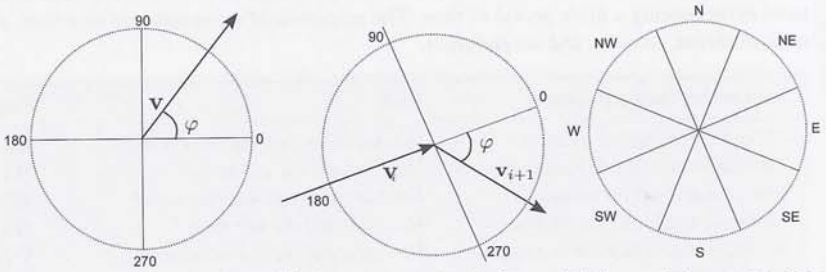


Fig. 10.1 For a saccade, the saccadic direction φ is the angle between the saccade and the horizontal axis in the coordinate system of the stimulus. The saccadic amplitude θ is the angular distance the eyes move during the saccade.



(a) Absolute saccadic direction (b) Relative saccadic direction (c) Segmented saccadic direction, as used by Ponsoda *et al.* (1995)

Fig. 10.2 Saccadic direction, sometimes also referred to as orientation.

Note that direction is a *circular*—angular—dimension (see Figure 10.2), and that this requires the use of circular statistics (see Batschelet, 1981 and Drew & Doucet, 1991 for the calculation of many sorts of descriptive and inferential statistics). While circular statistics have been widely used on eye-tracking data in neurological and biological studies (for instance, Snyder, Batista, & Andersen, 2000), they appear not to be an established analysis tool in the psychological and applied fields of eye-movement research.

10.1.1 Saccadic direction

Target question	*In what direction does the saccade take the eye?*
Input representation	*A saccade*
Output	*The direction φ (degrees)*

The saccadic direction, sometimes known as 'saccadic orientation', is the direction of any saccadic movement. Figure 10.3 describes how this calculation is done when zero is to the right and orientation order is counter-clockwise (the trigonometric version). Some studies instead use 0° for upward and a clockwise ordering of directions (the clock and compass version), which alters the calculations slightly.

There are several operational definitions of saccadic direction:

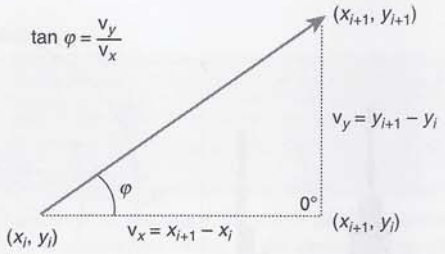


Fig. 10.3 Absolute saccadic direction φ for a saccade between fixations (x_i, y_i) and (x_{i+1}, y_{i+1}) can be calculated using basic trigonometry. Note that many saccades are curved and hence have varying directions along the trajectory. The saccadic direction only represents an ideal straight line from start to end point.

1. *Absolute saccadic direction* can be calculated based on the coordinates of the immediately preceding and consecutive fixations using simple trigonometry, if it is not already available from the event detection software (see Figure 10.2(a)).
2. *Relative saccadic direction* is calculated as the difference between the absolute saccadic directions of the current and the previous saccade (see Figure 10.2(b)).
3. *Segmented saccadic direction* means dividing—quantizing—the absolute saccadic direction into discrete bins. Ponsoda *et al.* (1995) and Lee, Badler, and Badler (2002) use eight segments as in Figure 10.2(c), while Gbadamosi (2000) uses 16 segments.
4. *Adduction*, *abduction*, *centrifugal*, and *centripetal* movement are common direction distinctions in binocular research (Collewijn *et al.*, 1988). Adduction refers to saccades towards the nose; abduction refers to saccades away from the nose; centrifugal movement refers to saccades away from the central line of fixation; and lastly, centripetal movement refers to saccades towards the central line of fixation.

Other varieties of saccadic direction measures exist, for instance segmented relative direction, but they appear not to have been used.

Histograms of saccadic direction have previously been presented linearly. In the twenty-first century, they have gradually been replaced by circular histograms known as 'rose plots' (sometimes 'polar' or 'angular' plots). Figure 10.4 shows a linear histogram, as well as a rose plot from reading data.

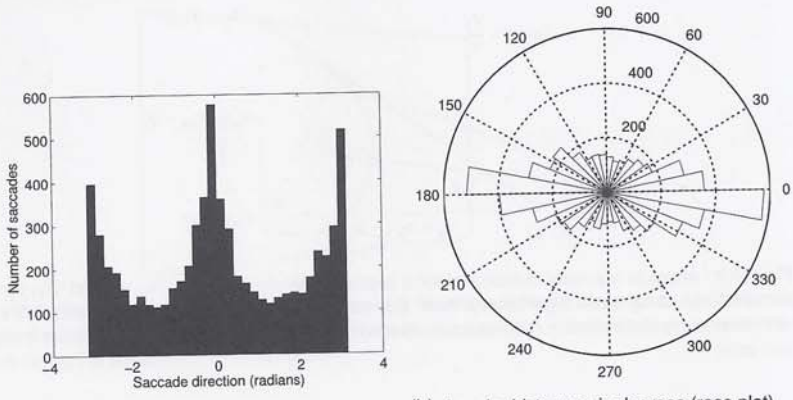
Many studies that have used saccadic direction measures have been vision experiments with a central fixation cross surrounded by peripheral targets at fixed saccade angles. A linear histogram can show how many saccades are launched in each direction, and classical non-circular statistics can be used, since in such a study there is no need to calculate the average saccade direction or to compare distributions of saccade directions between tasks or conditions.

Another large user group of saccade direction measures are neurologists. The preparation and execution of saccades in different directions can then be compared to neuronal discharge or brain activation. These researchers use both rose plots and circular statistics.

Over the past ten years, a few studies of saccade direction have emerged in general picture or scene viewing. Rose plots are then an informative way of illustrating the results. The following are general observations about saccadic direction:

Reading conventions decide the predominant saccadic direction when viewing written text.

Scene viewing Most saccades are aligned with the picture horizon (Foulsham, Kingstone, & Underwood, 2008). Pictures of natural scenes are more affected by picture rotation than



(a) Linear histogram in radians (b) Angular histogram in degrees (rose plot)

Fig. 10.4 Saccadic direction histograms for reading data. First a classical linear histogram, then an angular histogram (rose plot) for the same data. Bin size is $\pi/16$ radians (11.25 degrees).

pictures of fractals (Foulsham & Kingstone, 2010). There is a clear tendency in scene viewing to make either horizontal or vertical, but more seldom oblique saccades (Tatler & Vincent, 2008; Lee *et al.*, 2002; Mackworth & Bruner, 1970); horizontal saccades are more frequent, however.

Oblique saccades The few oblique saccades are pulled by dual muscle pairs, which are not always in perfect sync, and therefore tend to have a larger curvature than horizontal and vertical saccades (Viviani, Berthoz, & Tracey, 1977).

HV-ratio

Target question	*Are saccades predominantly horizontal or vertical?*
Input representation	*A set of saccades*
Output	*The ratio of horizontal to vertical saccades*

The horizontal to vertical (HV-) ratio is defined as the sum of all horizontal saccade amplitude components v_{hi} in a trial, divided by the sum of all vertical saccadic components v_{vi} in the trial. It is calculated as

$$\text{HV-ratio} = \frac{H}{W} \cdot \frac{\sum_i v_{hi}}{\sum_i v_{vi}} \tag{10.1}$$

where the H and W are the height and width of the stimulus, and the ratio $\frac{H}{W}$ normalizes for stimulus proportions.

Lau, Goonetilleke, and Shih (2001) used the HV-measure to study the difference in search direction between mainland and Hong Kong Chinese participants, finding that the search pattern differed significantly. They interpreted this result as reflecting the differences in reading direction between the two areas. This has implications for design where populations have differing reading directions.

Overall fixation vector

Target question	*Where are the fixations relative to the central gaze line?*
Input representation	*A set of fixations and a central gaze point*
Output	*Direction (degrees) and length (mm or pixels) of the overall fixation vector*

The overall fixation vector is an average direction of fixation positions from a central gaze point, which for monitors is simply the centre of the monitor. The fixation vector takes the number, position, and duration of fixations into account and indicates, on average, which part of a stimulus has received most attention. Given N fixations, the measure is calculated as

$$\text{Overall fixation vector} = \sum_{i=1}^{N} t_i v_i \tag{10.2}$$

where $v_i = (x_i - x_c, y_i - y_c)$ represents a vector from the centre point (x_c, y_c) to the fixation position (x_i, y_i), and t_i is the duration of fixation i.

The measure was introduced by Chi and Lin (1997), who recorded data from participants viewing pre-recorded videos of car-driving situations. They argue that the overall fixation vector indicates visual workload. On urban roads, for instance, the overall fixation vector had an average length of 499 inches and a direction of 163° during daylight, while at night, the length was 751 inches and the direction 190°. The latter case would indirectly imply a greater workload.

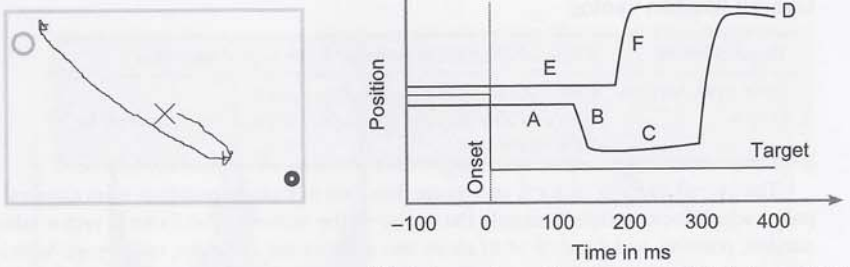
The antisaccadic metrics

Target question	*To what degree can the participant inhibit reflexive saccades to suddenly appearing targets?*
Input representation	*Raw data samples recorded during antisaccade trials*
Output	*Proportion of errors in saccade direction and saccade latency (ms)*

Antisaccade metrics are a whole family of measures related to a specific task, and the core measure is saccade direction. The task is simple: have your participants looked at a central fixation cross, flash a peripheral target on one side, at which the participant must immediately look in the *opposite* direction. Antisaccade tasks can cause fatigue, so it is especially important that trial orders are randomized when the experiment is long.

The antisaccade metric is comprised of a collection of six measures illustrated in Figure 10.5(b). They all relate temporally to the onset of the target in the periphery which corresponds to time 0, marked in the figure with a dashed line. The *error latency* (A) is the time in ms it takes to launch a saccade towards the target. The *error amplitude* (B) is the saccadic amplitude towards the target. The *time to correct* (C) is the fixation duration (or dwell time if there are more fixations) at or near the target before the participant corrects himself by a saccade in the opposite direction. The *final eye position* (D) is the closest the eye reaches to the direct opposite mirror position of the target during the trial. *Correct antisaccade latency* (E) and *amplitude* (F) are the time from onset of the target until the correct saccade is launched, and the amplitude of that saccade, respectively. An additional metric is the *number of correct trials*.

The gap condition variety of the antisaccade task is when the the central fixation cross offsets some time before the appearance of the peripheral target (so as to release attention from the centre, p. 430). As Hutton and Ettinger (2006) point out, in the gap condition, antisaccade errors are more common and correct antisaccade latencies are shorter.



(a) Hypothetical raw data in an antisaccade trial. The black dot is the target, and the grey ring (never visible on the stimulus) shows the goal position directly opposite to the target.

(b) Antisaccade metrics. One correct trial (upper black line) and one incorrect trial (lower black line). A: Error latency. B: Error amplitude. C: Time to correct. D: Final eye position. E: Correct antisaccade latency. F: Correct antisaccade amplitude. Modified from Hutton and Ettinger (2006).

Fig. 10.5 Antisaccade trial and measures.

Ettinger *et al.* (2003) found a between- but not within-session practice effect on the antisaccade task, and point out that this has to be considered with repeated testing of the same participants on different occasions. There are trial-by-trial effects also, as found by Tatler and Hutton (2007). For instance, after a successful trial, if the target appeared in the same hemifield on the next trial, the error rate is low and the primary saccade latency (E) is slightly reduced. Trial-by-trial effects call for a careful design of the experiment, or the introduction of more advanced statistical methods such as multilevel modelling.

Data validity in the antisaccade task also depends on how well participants' states are known. For instance, a large number of drugs and medications affect antisaccade behaviour. A precise and accurate diagnosis of participants is also crucial to the validity of antisaccade studies.

Poor participant performance may indicate that (prefrontal) voluntary control, and hence inhibition of reflexive saccades to the periphery, is poor, possibly due to a neurological impairment. The antisaccade paradigm is widely used in the field of neuropsychiatry (next to smooth pursuit tests), but is also used in several other areas of psychological research concerning eye-movement control.

Psychiatric disorders There is considerable evidence that genetic predisposition to schizophrenia can be diagnosed using antisaccadic metrics. Similarly, affective disorders, polar disorders, psychosis, obsessive compulsive disorder, Tourette's syndrom, attention deficit hyperactivity disorder, Alzheimer's desease, Huntington's disease, Lewy body dementia, palsy and Parkinson's disease show atypical effects in antisaccadic metrics (list from Hutton & Ettinger, 2006, p. 309). Prosaccades seem to be more impaired in neurological disorders than in the psychiatric ones. The precise neural and cognitive mechanisms underlying antisaccadic metrics are under debate, and the interpretation of data—as always—relates to how they were recorded.

Drugs and medication Many drugs have an effect on antisaccade performance, but not always in the way one would naturally assume. For instance, low levels of ethanol increase the proportion of correct antisaccades (Khan, Ford, Timney, & Everling, 2003), and nicotine reduces antisaccadic errors in participants who suffer from schizophrenia (Kumari & Postma, 2005).

Minor neurological damage The observation that greater distraction resulting from whiplash—experienced during a car crash most commonly, where the head is violently flung backwards then forwards—may benefit from specific rehabilitation using the antisaccade task was tested by Mosimann, Muri, Felblinger, and Radanov (2000). Moreover, very prematurely born children, although having largely normal control over saccades and smooth pursuit as adults, make significantly more antisaccadic direction errors (Newsham, Knox, & Cooke, 2007).

Age Children perform more poorly than adults on the antisaccade task, but error rates and latencies will decrease as their prefrontal control develops. Older adults appear to make more direction errors, and/or exhibit longer saccadic latencies for correct antisaccades (Eenshuistra, Ridderinkhof, & Molen, 2004).

Working memory Several studies have shown that antisaccade measures are sensitive to variations in working memory capacity (e.g. Kane, Bleckley, Conway, & Engle, 2001).

Direction of regressions, backtracks, look-aheads, and leading saccades

Target question	*Did the saccade bring the eye to a task relevant AOI?*
Input representation	*A saccade and an upcoming action*
Output	*Binary*

Regressions by definition are directed *in the opposite direction of the text*, and backtracks in the opposite direction *of the previous saccade* (pp. 262–265). Their directions are not used as measures but only as part of the event definition. Look-aheads are instead movements *towards upcoming activities*, such as looking at the soap some time before reaching towards it in order to make sure that it is there for later hand-washing use. The direction of a look-ahead saccade is given by the task and not by established (reading) conventions or absolute angular directions. Although researchers talk about 'look-ahead fixations', it is the saccade that moves the eye to the look-ahead position.

The look-ahead concept assumes that look-ahead fixations are limited in time and number, and that tasks preceding the use of the object which receives a look-ahead fixation require visual attention. If using a greasy tool like a spanner, for instance, this task requires attention up until completed, even though the next planned action is to wash your hands, indicated by a look-ahead fixation to some soap near the the sink. The following factors have been observed regarding the direction of look-aheads:

Speech planning Participants describing a picture often look ahead to objects soon to be described (Holsanova, 2008).

Physical task requirements Within a hand-washing task, participants looked ahead at a soap dispenser, towel dispenser, and waste bin significantly more often than with a cup-filling task (Pelz & Canosa, 2001). Similar task-dependency has been observed with tea- and sandwich-making actions (Land *et al.*, 1999; Hayhoe, 2000).

Accuracy in guidance Mennie *et al.* (2007) found that look-ahead saccades may facilitate the accuracy of the future eye movement that guides the upcoming action.

Leading saccades are another type of eye movement linked to an upcoming action. These are saccades with an amplitude over 1° that occur during smooth pursuit. A leading saccade is directed *in the direction that the participant expects of the target motion*, but the leading saccade skips over the target, and is followed by a period of slow pursuit so that the target can catch up with the eye. In other words, the eye's movement is slower than the target's so

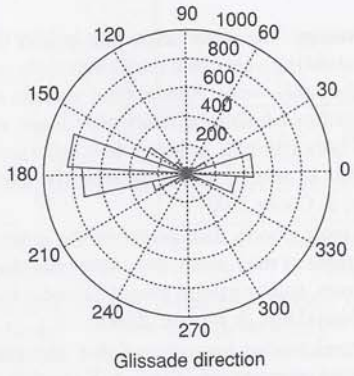


Fig. 10.6 Direction of glissades, in circular degrees, during reading. Glissades were calculated using the algorithm described in Nyström & Holmqvist, 2010.

that the two can become aligned. Like many smooth pursuit measures, 'leading saccades' are used in research on schizophrenia (e.g. Ross *et al.*, 1999).

10.1.2 Glissadic direction

Target question	*In what direction does the glissade take the eye?*
Input representation	*A glissade*
Output	*Direction φ (degrees)*

Glissadic direction is a measure of the orientation of the small movement appended to many saccades that we call glissades (p. 182). Just like saccadic direction, it can be calculated from the start and end points, although in reality glissades are very curved, even more so than saccades. Figure 10.6 shows a rose plot of glissadic direction during reading of English (study 1 on page 5). It can been seen that the majority of glissades go in the opposite direction of forward reading saccades; this is what should be expected since a glissade is an event that often swings the eye backwards compared to saccade direction.

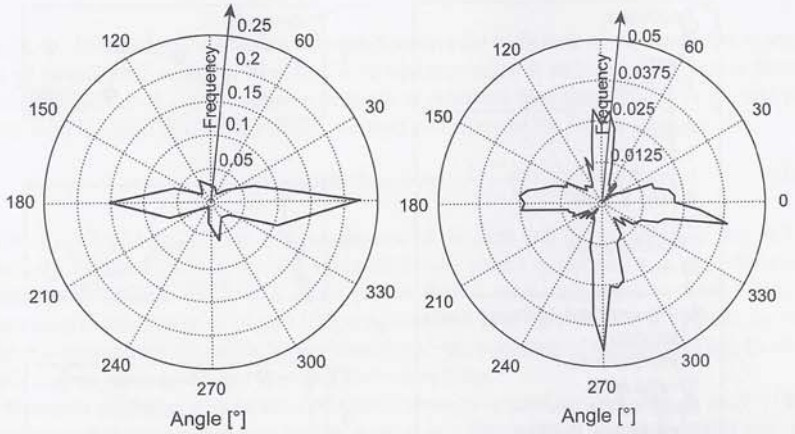
10.1.3 Microsaccadic direction

Target question	*In what direction does the microsaccade take the eye?*
Input representation	*A microsaccade*
Output	*Direction φ (degrees)*

Microsaccade direction is a measure of the orientation of transient, fixational micro-movements of the eye. Once the start and end point of the movement have been found, the direction can be calculated the same way as for saccades and glissades.

The distribution of microsaccade orientation is very similar to ordinary saccades, with a strong dominance for horizontal and vertical directions, as data in Figure 10.7 from Engbert (2006) illustrate. Interestingly, vertical microsaccades seem to be largely monocular.

Studies making use of the microsaccade direction measure are common in neurology, and scarce elsewhere. Rolfs, Engbert, and Kliegl (2005) found that when playing sounds on one side of the participant, microsaccades moved in the opposite direction of the sound within 100–200 ms. But when additionally showing visual stimuli on the same side, a second



(a) Binocular microsaccade direction. (b) Monocular microsaccade direction.

Fig. 10.7 Distributions of microsaccade directions for binocular (a) and monocular (b) microsaccades. Reprinted from *Progress in Brain Research, 154*(1), Ralf Engbert, pp. 177–192, Copyright (2006), with permission from Elsevier.

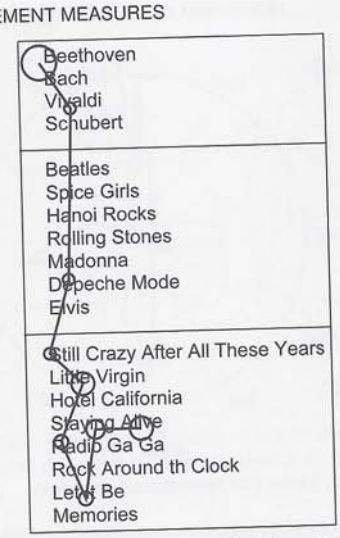
microsaccade moves the eye back towards the visual stimulus. Rolfs *et al.* conclude that auditory and visual information are integrated even at the microsaccadic level. Engbert and Kliegl (2003) found that microsaccade direction reflects the locus of visual attention, a view that is not shared by Horowitz, Fine, Fencsik, Yurgenson, and Wolfe (2007). In fact, there is an ongoing debate concerning how biases in microsaccade direction are related to covert attention shifts and motor programming (Martinez-Conde *et al.*, 2009).

10.1.4 Smooth pursuit direction

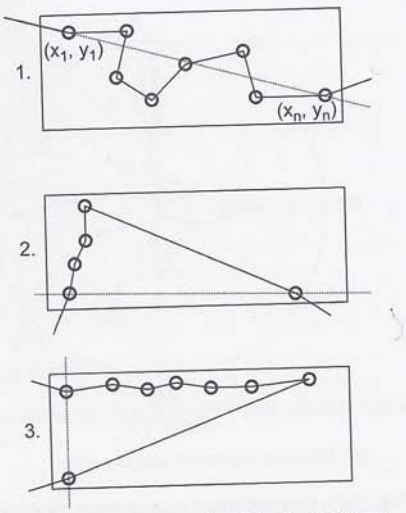
Target question	*In what direction does smooth pursuit take the eye?*
Input representation	*A period of smooth pursuit*
Output	*Average or individual smooth pursuit direction (degrees)*

Smooth pursuit direction is a measure that may appear to attribute a single direction to a characteristic of eye movements which is considerably varied—smooth pursuits can track a stimulus as it moves between a multitude of locations, thus giving a range of directions within one smooth pursuit. It is only possible, therefore, to attribute an objective measure of smooth pursuit direction if we allow the direction value to change along the smooth pursuit path. From the point of view of smooth pursuit detection, this is often achieved by dividing long periods of pursuit into smaller pursuits, interleaved with catch-up saccades. As an alternative, a momentous direction value between any two points $((x_i, y_i)$ and $(x_{i+1}, y_{i+1}))$ along the path, can be calculated as in Figure 10.3 by replacing fixations with raw data samples. High precision data is required or the measure will have very noisy values.

Smooth pursuit has directional preference: most humans and primates tend to be better at horizontal than vertical smooth pursuit. This is defined by the ability to initiate smooth pursuit quickly and to pursue smoothly without making catch-up saccades, which is worse when the required pursuit target moves along the vertical axis. For these vertical pursuits, moreover, most humans are better at downward than upward pursuit (Rottach *et al.*, 1996). Horizontal



(a) Free scanpath 'sweep' with vertical direction, adapted from Aaltonen *et al.* (1998).



(b) Scanpath direction in AOIs. Dashed lines indicate the operational definition by Renshaw *et al.* (2003).

Fig. 10.8 Scanpath directions. An example of a free vertical downward scanpath (a). Illustration of the operational definition of scanpath direction as the dotted line through the first and the last fixation in the AOI (b; 1). Examples of when the definition is not intuitive (b; 2 and 3).

superiority of smooth pursuit in the visual system is often related to the extensive experience of following objects in motion along the horizontal meridian (Collewijn & Tamminga, 1984).

The ability to elicit smooth pursuit in many directions is commonly studied in the assessment of neurological disorders. The neurological sites that subserve smooth pursuit mechanisms, or the effect of medication on smooth pursuit in neurological conditions, can therefore be examined. However, very few studies have made use of smooth pursuit direction as a measure in its own right.

With the purpose of developing a detection algorithm, Agustin (2010) used relative sample-to-sample directions within a moving window to, after saccade detection, separate pursuit samples from fixation samples. When the variance of relative sample-to-sample directions was low, pursuit was more likely than fixation and vice versa.

10.1.5 Scanpath direction

Target question	*In what direction did the scanpath take the eye?*
Input representation	*A scanpath (and sometimes an AOI)*
Output	*The direction of the scanpath (degrees)*

Scanpath direction is a measure of the *general* direction of a sequence of fixations and saccades while scanning a stimulus. Like smooth pursuit, the whole scanpath can change its direction many times over time. Therefore, it is more common to measure direction on smaller, local parts of the scanpath. If these smaller parts consist of individual saccades, the scanpath direction measure simplifies to a standard saccade direction measure.

There exist at least two different ways of quantifying scanpath direction. First, detecting

sweeps (p. 267) and calculating an average direction for it. In lack of sweep events, a second way of quantifying scanpath direction is to operationalize it in terms of direction through AOIs. Renshaw *et al.* (2003) define a measure of scanpath direction through AOIs, and refer to this with the term 'gaze orientation', defined according to the following equation:

$$\text{Scanpath direction} = \frac{y_n - y_1}{x_n - x_1} \tag{10.3}$$

where (x_1, y_1) and (x_n, y_n) are the coordinates of the first and last fixations in the AOI respectively. Figure 10.8(b) shows the principle. Scans with a ratio of up to 1 were designated as horizontal whereas those with ratios greater than 1 were designated vertical. Using the trigonometric calculation in Figure 10.3 would instead give the direction in degrees. In either case, this operational definition is very sensitive to the positions of the first and last fixations, which may be misleading, as Figure 10.8(b) exemplifies.

Scanpath direction is indicative of search strategy according to Aaltonen *et al.* (1998), an interpretation which is shared by Renshaw *et al.* (2003), Poole and Ball (2005) and other usability researchers. The scanpath direction measure appears to be used sparingly, however.

10.2 Movement amplitude measures

Movement amplitudes concern events such as *saccades*, *scanpaths*, *smooth pursuit*, *glissades*, and *microsaccades*. *Amplitude* has two common operational definitions (Figure 10.9). The first refers to the shortest distance between the start and end point of a movement, i.e. the displacement, and the second is the total distance along the trajectory of the movement between the start and end points. The latter can be approximated by multiplying the average velocity of the movement by its duration. Since many types of eye movements do not follow a straight path, these operational definitions are often different.

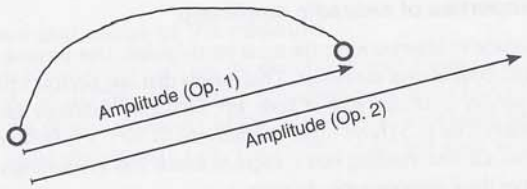


Fig. 10.9 The two definitions of saccade amplitude. Amplitude according to the first operational definition (Op. 1) is the shortest path between the end points, known as the Euclidian distance, whereas the second (Op. 2) refers to the total distance covered by the saccade along its trajectory. Amplitude according to Op. 2 can be seen as taking the end points of a curved saccade, and stretching the saccade until it forms a straight line.

Both operational definitions are used in journal articles and research reports, but only occasionally is it explicitly stated which is meant. For example, Smeets and Hooge (2003) use the Euclidian definition 1 in Figure 10.9, as do the earlier SR Research manuals from around year 2000, while later EyeLink manuals (SR Research, 2007) and SMI Technical Notes (SMI, 2007) propose a calculation according to definition 2, namely:

$$\theta = \frac{1}{1000}(t_{\text{offset}} - t_{\text{onset}})\dot{\theta} \tag{10.4}$$

where $\dot{\theta}$ [°/s] is the average velocity of the saccade, and $(t_{\text{offset}} - t_{\text{onset}})$ the saccade duration in ms. If you have not written your own software for calculation of amplitude, beware of whether

the program you use calculates amplitude according to the first or the second operational definition.

10.2.1 Saccadic amplitude

Target question	*Across what distance did the saccade take the eye?*
Input representation	*A saccade*
Output	*The amplitude of the saccade (pixels or degrees)*

The saccadic amplitude (earlier 'magnitude', and in reading research, 'size') is the distance travelled by a saccade from its onset to the offset. The unit is typically given in visual degrees (°) or pixels, but has also been segmented into 16 discrete categories by Gbadamosi (2000). Saccadic amplitude is much used in itself, as well as in combinations to form complex measures.

If glissades are appended to preceding saccades in your event detection algorithms, the saccadic amplitude will be somewhat shorter, while adding it to the following fixation will make the amplitudes longer. Event-detection algorithms with too high a threshold also leave out raw data samples that could be argued to belong to the saccade and hence underestimate saccadic amplitude.

'Interfixation distances' are a coarse approximation of saccadic amplitudes. Interfixation distances are calculated as the Euclidean distance between fixation points (Megaw & Richardson, 1979), and are typically used on low-speed, low-precision eye-trackers, or when for other reasons, saccade measurements cannot be made. There is always a danger that some event other than saccades, such as a blink or lost data samples, is hidden between what the detection algorithm considers to be fixations, and then interfixation distances do not equate to saccadic amplitudes.

Important properties of saccadic amplitude

Saccadic amplitude is *idiosyncratic* for most participants, like fixation durations and several of the other basic oculomotor measures. This means that one person's parameters are different from another person's, irrespective of task. For instance, although data from the reading to music study (study 2 on p. 5) have highly significant ($0.90 < r < 0.98$) correlations *within the individual* across all four reading tasks, there is much less consistency *between individuals*. Participants have their idiosyncratic defaults.

Saccades larger than about 15° are often inaccurate, failing to reach the intended target; in such cases they are commonly followed by another, corrective saccade (Becker & Fuchs, 1969). This difficulty in making longer saccades is a motor property of the saccadic system, which has been studied using the amplitude-based measures saccadic gain (p. 452), and the main sequence (p. 316).

Large saccades are often followed by small corrective saccades, but otherwise there appears to be no systematic relationship between the amplitudes of successive saccades (Motter & Belky, 1998). It is worth noting, however, that this conclusion is based on primate eye-movement research, and may not generalize. Tatler and Vincent (2008) found some evidence for sequences of short amplitude saccades in scene viewing (see p. 266), indicative of focused inspection of details. This was not the case for long saccades, which showed no consistent patterns in terms of sequences.

Saccadic amplitudes can be made shorter by flashing a light at the time of the onset of the saccade. This is known as 'saccadic compression' (Ross *et al.*, 1997), and reveals the mechanisms by which we update our perceptual experience of what we are viewing so that it

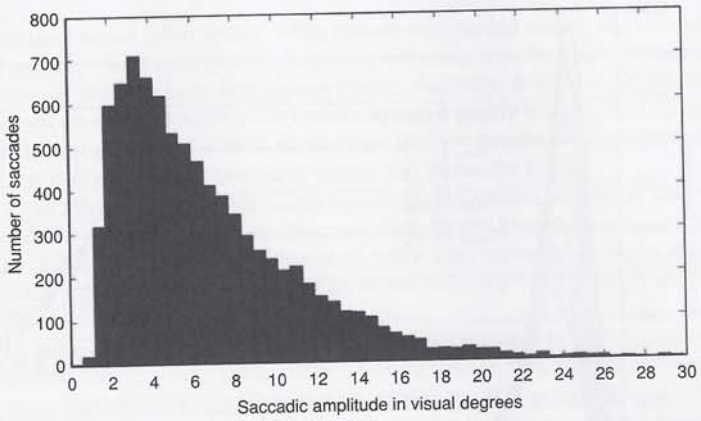


Fig. 10.10 Histogram over saccadic amplitudes (bin size 0.6°) during scene viewing (natural photographs shown on a computer screen). Recorded with a tower-mounted system at 1250 Hz, and saccades derived using the velocity-based algorithm in BeGaze 2.1 with a peak velocity setting of 40°/s.

remains consistent across eye movements.

The size of the stimulus image influences the mean and median saccadic amplitudes (Von Wartburg *et al.*, 2007). When there is no stimulus, as in mental imagery studies, Zangemeister and Liman (2007) found that saccadic amplitudes are shorter than during real picture viewing. Humphrey and Underwood (2008), however, found the opposite trend when comparing amplitudes from imagery to picture encoding.

When recording binocular eye movements, the saccades recorded from each eye often have slightly different amplitudes (Kenyon, Ciuffreda, & Stark, 1980).

Amplitude values and usage of the measure

Saccadic amplitudes from scene viewing are distributed with a very clear rightward skew, as illustrated in Figure 10.10. This means that participants are more likely to saccade towards a location within the proximal vicinity of the last fixation, with large exploratory saccades to new locations being increasingly rare. For reading data, the skew is even larger (Figure 10.11).

Saccadic amplitude is one of the most used eye-movement measures. During reading, for instance, saccadic amplitude is known to adapt to combined physical, physiological, and cognitive factors. Reading saccades are limited in length by the visual spanwidth, which is around 7–8 letters (2°) in the average reading situation (Rayner, 1998). It would be detrimental to understanding to make such long saccades that you fail to see some parts of the words, and therefore reading saccades must be on average 7–8 letters long.

When using other stimuli than text, saccadic amplitudes also adapt to task demands, workload, the stimulus, and the needs of the current cognitive process. *Decreased saccadic amplitudes* have been found in relation to:

Search task difficulty Saccadic amplitudes are shorter in more difficult search tasks (e.g. Zelinsky & Sheinberg, 1997). Using a search task and studying performance improvements, Phillips and Edelman (2008) found that saccade metrics (primarily saccade amplitude) accounted for much more of the variability and improvement in performance than did fixation duration.

Increased cognitive load With increasingly demanding counting complexity, May, Kennedy,

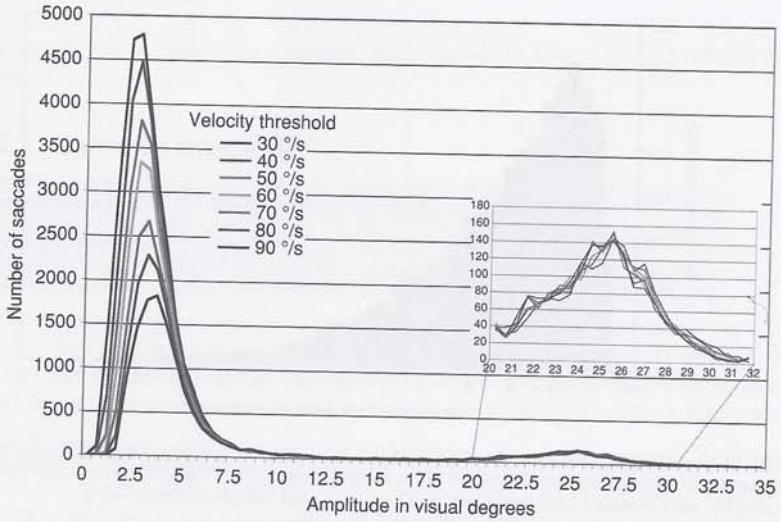


Fig. 10.11 The distribution of saccadic amplitudes with 10 participants while reading (study 1 on page 5). Bin size 0.5°. For larger velocity threshold settings, the velocity-based event detection algorithm outputs much fewer saccades in the lower amplitudes, in particular the many short reading saccades below 5°. Longer saccades, here exemplified by the return sweeps at around 25°, are not affected at all, since their velocity peaks are far above any of the thresholds.

Williams, Dunlap, and Brannan (1990) found a reduction in saccadic amplitude. Ceder (1977) and Troy, Chen, and Stern (1972) also found that when task time increases, the number of saccades with amplitudes above 9.5° decreases. Car drivers who perform cognitive operations unrelated to driving have a smaller 'saccadic extent' (Recarte & Nunes, 2003). Reduced saccadic extent has implications for mobile phone use when driving because this may encourage engagement in mental activity about things which are not physically present, or task relevant. These findings have been interpreted as an effect of cognitive load, resulting in 'tunnel vision' (Williams, 1988), which restricts the amplitude of saccades.

Careful inspection versus meaningfulness When a participant inspects an object carefully, saccades are shorter than during overview scans (Buswell, 1935). This observation has developed into the scanpath event ambient/focal (p. 266), but it is also part in the focal/global measure (p. 338). Interestingly, in usability research, larger saccadic amplitudes have been interpreted as a sign of *more* meaningful visual clues than shorter ones (Goldberg *et al.*, 2002).

Adaptation to memory load When stipulated to do so, large saccadic amplitudes can be avoided by participants. Moreover, visual memory usage increases to compensate for the less often updated access to visual information (Phillips & Edelman, 2008; Inamdar & Pomplun, 2003). The avoidance of making long saccades is the effect of a cost, which the authors do not specify clearly, but which could relate to the main sequence measure described on page 316.

The onset of brain injury If saccadic amplitudes become inaccurate, due to neurological injury or growth, the brain adapts and corrects the saccades back to correct amplitude sizes (McLaughlin, 1967). This has given rise to neurobiological research on 'saccadic adaptation' that is focused on understanding the neural site for saccadic plasticity.

High-frequency visual information When free-viewing natural scenes, shorter saccades (< 8°) appear to be driven by high-frequency information at the landing position, while longer saccades are scale independent (Tatler, Baddeley, & Vincent, 2006). Basically, our eyes are drawn towards areas containing detail near to where we are already looking. Longer saccades are more exploratory, because peripheral vision cannot disambiguate coarse and fine visual detail further away from the fovea.

A lower error tolerance in pursuit Catch-up saccades in smooth pursuit are made to compensate for participant tracking error, and many short saccades have been taken as an index of low tolerance for position error, while long saccades are made after a large tracking error has occurred, according to the overview in O'Driscoll and Callahan (2008).

Beginning, poor, and dyslexic readers When children first learn to read, saccades have an amplitude corresponding to the distance between letters. Poor readers and readers with dyslexia also exhibit shorter than average amplitudes (Rayner & Pollatsek, 1989).

Oral reading In oral reading, average saccadic amplitude falls to around 6 letters (1.5°), while during music reading and typing, saccades are a mere 1° on average (Rayner & Pollatsek, 1989).

Decreased musical tempo For participants reading musical scores, Kinsler and Carpenter (1995) found that the mean saccadic amplitude increased as the tempo of the music increased.

Lower participant age Young children exhibited shorter amplitudes than adults when examining pictures of familiar scenes (Mackworth & Bruner, 1970).

When the distances are very small, it may be detrimental to make saccades (Kowler & Steinman, 1977). In this study participants were asked to count thin bars 7–14 arcmin apart. The results showed that participants performed better if they made fewer saccades (Kowler & Steinman, 1977). The authors hypothesize that at this scale, attention can move without accompanying saccades. In other words, in some circumstances it may be favourable to preserve the image on the retina and tune attentional resolution than to make saccades.

The skewness of saccadic amplitude distribution

Target question	*Do shorter or longer saccadic amplitudes dominate?*
Input representation	*A set of saccades from a trial or whole recording*
Output	*A skewness value for the frequency distribution*

Skewness refers to the degree of asymmetry in the distribution of saccadic amplitudes. It should not be confused with the skew of the saccadic velocity profile, the 'saccadic skewness' (p. 333). As Figures 10.10 and 10.11 show, saccadic amplitude distributions are heavily skewed, and the degree of skewness is task dependent. Characteristic for a skewed distribution is that the mode, the median, and the mean do not coincide. For saccadic amplitudes, fixation durations, and many other eye-movement measures, the distribution tends to be skewed to the right (positive skew), indicating that the portion under the curve is larger on the right side of the mode than on the left side. There are several ways of calculating skewness, and, unfortunately, it is not always clear from journal papers which mathematical definition was used. One definition is based on the distance between the sample mean and the sample median. Another definition is based on the difference between the sample mean and the sample mode. In both cases, the larger this distance, the more skewed is the distribution. A third common definition is (Crawley, 2005)

$$\text{skewness} = \frac{\Sigma(X-\overline{X})^3/N}{s^3} \tag{10.5}$$

where $X, \overline{X}$ and N represent an amplitude value, the average amplitude, and the number of amplitude values, respectively. Often this value is divided by the approximate standard error $\sqrt{\frac{6}{N}}$, yielding a standard score. A non-significant value (roughly between -2 and $+2$) indicates that the distribution is normal. A value that falls outside that range is an indication that the distribution is skewed. The standard error may also be used to estimate the confidence interval of a distribution's skewness, and, consequently, also be used as an indication of whether one distribution is significantly more skewed than another.

As a hypothetical example, consider distribution A with a skewness of 0, with a standard error equal to 1. Distribution B has a skewness of 1.5 and a standard error of 2. Since there is considerable overlap in the two confidence intervals[27] $[-1.96, 1.96]$ for A and $[-2.42, 5.42]$ for B, we can not conclude that the distribution in B is significantly more skewed than that in A.

Reported values for the skewness of the saccadic amplitude distribution range from 0.5 to 1.2. Many of the effects on saccadic amplitude are likely to also affect this skewness measure, but few of them have been examined. Two exceptions are:

The task may change the skewness (Welchman & Harris, 2003).

The monitor size does effect the skewness value (Von Wartburg *et al.*, 2007).

Inflection point of main sequence

Target question	*At which amplitude are motor neurons bursting at full throttle?*
Input representation	*A set of saccades*
Output	*An inflection point*

The *main sequence* is a systematic relationship between specific pairs of saccadic parameters, in particular the saccadic amplitude and the peak velocity (Bahill *et al.*, 1975b). The shorter saccades form a dense set with an almost linear relation to peak velocity. Above a certain velocity—called the *inflection point*—longer saccades have another slope, and in naturalistic data also often a more variable velocity. Such data are illustrated in Figure 10.12(a).

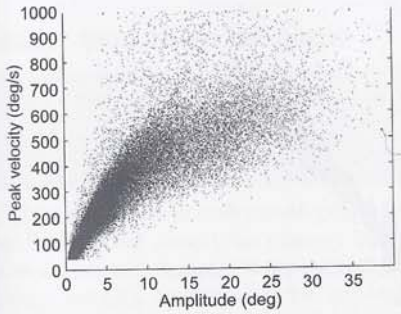
Some authors argue that if another event (such as a microsaccade) adheres to the same main sequence as the saccade, then that other event is also of a saccadic nature. Figure 10.12(b) shows that microsaccades obey the main sequence, and Figure 10.13 that glissades do also.

The *inflection point* is the point on the main sequence curve where this systematic relationship ends. It is typically reached for saccades with amplitudes between 15° and 20°. This inflection point has been interpreted as the point at which motor neurons are bursting at full throttle, after which further increase will require greater effort, and saccadic gain will be lower (Bahill *et al.*, 1975b).

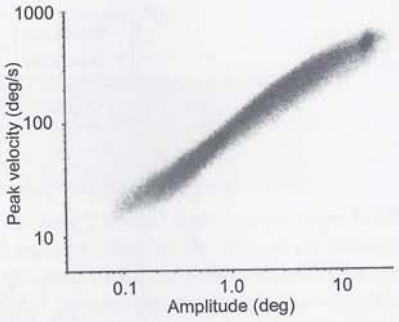
Van Opstal and Van Gisbergen (1987) and Smit, Van Gisbergen, and Cools (1987) argue that the data used to study the main sequence need to take into account both the saccadic skew and the slow-moving saccades, which had previously been ignored and removed from analysis. Note that when recording with search coils, the saccadic velocity is reduced for longer amplitudes. In the worst case scenario, this could undermine a correct main sequence calculation (p. 326 and Träisk *et al.*, 2005).

The main sequence has been used to build and test neurological models of saccade generation, and appears virtually unused in other research using eye movements. The measure

[27] As the z value for a 95% confidence value equals 1.96·standard error.



(a) Illustration of the main sequence effect using 57668 saccades. An inflection point can be seen at around 10°. The saccades are generated with the velocity algorithm and a threshold of 40°/s. Data from the mathematical problem solving project, number 4 on page 5.



(b) Microsaccadic main sequence to left (below circa 0.8°) softly overlap with the lower end of the saccadic main sequence. Reprinted from *Trends in Neurosciences, 32*(9), Susana Martinez-Conde, Stephen L. Macnick, Xoana G. Troncoso, and David H. Hubel, Microsaccades: a neurophysiological analysis, pp. 463–475, Copyright (2009), with permission from Elsevier.

Fig. 10.12 The main sequence.

is often used in combination with the saccadic gain/accuracy measure (Ciuffreda & Tannen, 1995).

Epelboim *et al.* (1997) showed that participants who tapped sequences on 3D targets located on a table in front of them exhibited a different main sequence than when they were just looking at the objects.

Main sequences for humans and monkeys are significantly different during free viewing of dynamic stimuli (Berg *et al.*, 2009). For all amplitudes in this study, monkeys had faster velocities than humans.

10.2.2 Glissadic amplitude

Target question	*How far did the glissade bring the eye?*
Input representation	*A glissade*
Output	*An amplitude (degrees)*

The glissade amplitude is the length of the small movement that is often appended to saccades. Amplitude calculation is identical to that of saccades (p. 312). As Figure 10.13 shows, glissades have similar dynamics to saccades, and extend the saccadic main sequence. Glissadic amplitudes are typically less than 1°, and glissades "appear to serve no useful purpose" (Kapoula *et al.*, 1986, p. 386). Neither does there seem to be a correlation between the amplitude of a saccade and its subsequent glissade (Nyström & Holmqvist, 2010).

10.2.3 Microsaccadic amplitude

Target question	*How far did the microsaccade bring the eye?*
Input representation	*A drift or microsaccade*
Output	*An amplitude (degrees)*

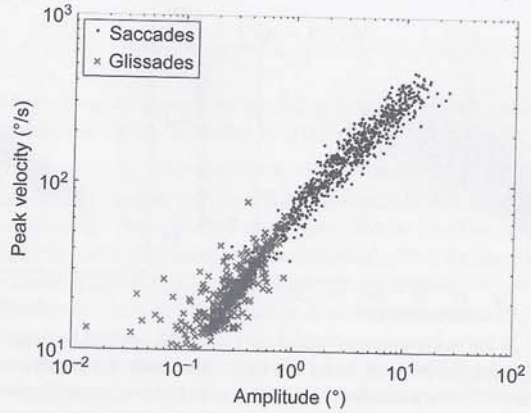


Fig. 10.13 Glissades extend the main sequence of saccades. Reading data at 1250 Hz. Saccades and glissades detected with the algorithm developed by Nyström and Holmqvist (2010).

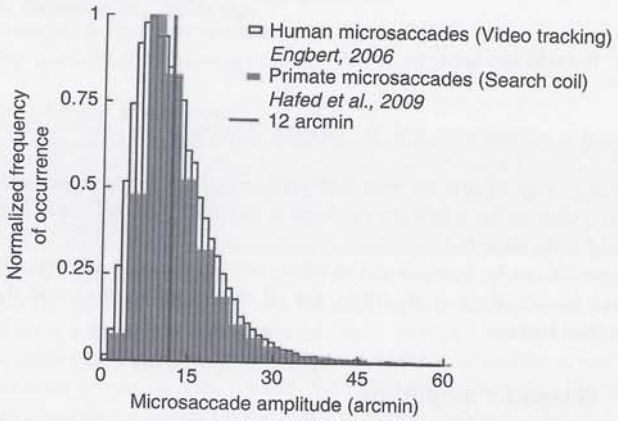


Fig. 10.14 Microsaccade amplitudes have a similar skewed distribution to ordinary saccades. 60 arcmin = 1°. Reprinted from *Trends in Neurosciences, 32*(9), Susana Martinez-Conde, Stephen L. Macnick, Xoana G. Troncoso, and David H. Hubel, Microsaccades: a neurophysiological analysis, pp. 463–475, Copyright (2009), with permission from Elsevier.

Microsaccadic amplitude calculation is identical to that of saccades (p. 312). The distribution of microsaccadic amplitude very much resembles that of saccades, only smaller in scale, as illustrated by Figure 10.14.

In the dark, both drift and microsaccades are 2–3 times larger than in light, while torsional drift remains the same as in light. Morisita and Yagi (2001) interpret this difference as showing that foveal displacement is necessary to detect and correct drift. Torsional drift alone implies peripheral displacement but not foveal displacement, which would not be enough.

Rolfs, Kliegl, and Engbert (2008) found a significant decrease in microsaccadic amplitude owing to the inhibition induced shortly after the onset of a visual stimulus, and argue that this could reflect activity in the central motor map.

10.2.4 Smooth pursuit length

Target question	*How long is the smooth pursuit event?*
Input representation	*A smooth pursuit event*
Output	*Length (degrees or pixels)*

The length of a smooth pursuit period is the distance travelled from pursuit onset to offset. With natural stimuli, the measure is operationalized either as the sum of all distances between raw data samples along the path or as the product of the pursuit duration with its average velocity. As with saccades, the amplitude of smooth pursuit can also be calculated as the shortest distance between the points of on- and offset. However, this assumes that the smooth pursuit direction remains relatively constant. Square-wave jerks, catch-up, back-up, and leading saccades may all occur during smooth pursuit, and may cause the detection algorithm to miscalculate the length.

'Smooth pursuit amplitude' sometimes refers to the distance between the extreme points of the pursuit path with sinusoidal stimuli.

When the recording is made with a controlled stimulus target, the length of perfect pursuit should equal the distance that the target travels. However, poor tracking is usually measured with the smooth pursuit gain measure (p. 450). In studies with natural stimuli, smooth pursuit can go on and off many times, resulting in a multitude of events of varying lengths.

Since general smooth pursuit detection algorithms have been difficult to develop, the smooth pursuit length measure is very little used.

10.2.5 Scanpath length

Target question	*How long is the scanpath?*
Input representation	*A scanpath*
Output	*Length (degrees or pixels)*

Scanpath length is often defined as the sum of all saccadic amplitudes in a scanpath. As many studies using this measure record data that do not allow for saccade curvature measurements, the Euclidean distance (Op. 1 in Figure 10.9) of saccadic amplitude is used in scanpath length calculations. In order to account for saccadic curvature, and to circumvent peculiarities in event detection algorithms, scanpath length can also be calculated for raw sample scanpaths as the sum of distances between samples. Note however, that when using raw data samples for scanpath length calculation, the result is dependent on the sampling frequency of the eye-tracker, much in the same way that the length of a coastline depends on the length of the yardstick used to measure it. Moreover, a poorer precision in the data increases the scanpath length. For these reasons, reported absolute length values cannot always be trusted.

As an example of this, in studies of the face viewing behaviour of schizophrenic, autistic, and socially phobic patients, some researchers present both versions of scanpath length, one based on raw samples, and the other based on fixation/saccade data (Loughland, Williams, & Harris, 2004; Green, 2006; Green *et al.*, 2008). These authors all report large differences between the scanpath length calculated from raw data samples compared to when calculated from fixations. Horley *et al.* (2004) write that "raw scanpath length [was] included to ensure that group differences in fixation scanpath length were not simply due to group differences in the number of fixations." The authors detected fixations using the I-DT algorithm using a minimum duration criterion of 200 ms with a dispersion criterion of 1°, and report that fixation based scanpaths are between 25–40% of the raw scanpath length. Much of this dif-

ference can probably be explained by the very high duration threshold, which undermines the detection of many fixations and hence makes the scanpath shorter (p. 148).

If trials—or separate tasks—have different durations, it may be necessary to divide the scanpath length by the duration of the trial. In principle, this results in an eye-movement speed measure (in the unit °/s), closely related to average saccadic velocity (p. 330). If scanpath length is not calculated over trials in full, a major challenge to an accurate calculation of scanpath lengths is in choosing the starting and ending points of the scanpath portion measured. In longer or self-paced tasks, the choice of scanpath start and end fixation is not obvious, and care must be taken so you can properly justify your cut off points.

Scanpath length as a measure has been used to:

Evaluate interfaces The optimal scanpath for a given task should be a straight line to a desired target. An unduly long scanpath could indicate non-meaningful representations or poor layout. With these assumptions, Goldberg and Kotval (1999) did observe a significant difference between scanpath lengths in good and poor user interfaces for a writing and drawing task. Renshaw *et al.* (2003) likewise also observed a strong significant difference in scanpath lengths between a poor and a good visual design of a graph. In both cases, the better design gave a smaller total scanpath length. Simonin, Kieffer, and Carbonell (2005) tested four different designs of digital photo albums, presenting photos either radially (star-formed), elliptically, as a square matrix, or just randomly positioned. The shortest scanpath length was found for elliptical layouts, which (together with a shortest search time) made the authors interpret it as being the most visually comfortable.

Approximate more complex concepts The term 'restricted scanpath' has been operationalized using scanpath length (p. 259). Generally, results show that participants with schizophrenia have a shorter—hence restricted—scanpath, and those with social phobia a significantly longer raw scanpath when looking at faces compared to controls.

10.2.6 Blink amplitude

Target question	*What distance does the eyelid travel during the blink?*
Input representation	*A blink*
Output	*An amplitude (mm)*

Blink amplitude is defined as the distance travelled by the eyelid, as illustrated in Figure 10.17 on page 325. This distance is not accurately measured with pupil and corneal reflection eye-trackers, but needs to be measured electro-oculographically or using dedicated eyelid trackers. Note that if the eye is slightly closed when the blink starts, the blink will have a smaller amplitude than if the blink starts from a fully open position. This fact is the foundation for using blink amplitude as a measure of workload, fatigue, and drowsiness.

According to Carney and Hill (1982) and Wolkoff, Nøjgaard, Troiano, and Piccoli (2005), for 80% of blinks, the descending upper eyelid covers more than two thirds of the cornea. For about 18%, the descending eyelid covers less than that; and around 2% of blinks are twitch blinking (flutter).

When blink amplitude decreases, performance errors in flight simulators increase for sleep deprived pilots (Morris & Miller, 1996). Blink amplitude was found to be a better predictor of performance than either blink rate or blink closure.

10.3 Movement duration measures

Movement duration measures all pertain to the time the movement took. Given the onset and offset, duration is trivial to calculate. The major challenge instead lays in detecting proper start and end points of the movement event.

10.3.1 Saccadic duration

Target question	*How long does the saccade take?*
Input representation	*A saccade*
Output	*Duration (ms)*

Saccadic duration ('transition time'; not the same as transitions between AOIs) is defined as the time the saccade takes to move between two fixations or instances of smooth pursuit. Onset of saccades is easy to calculate, but offset is ambiguous due to glissades, which can increase or decrease the duration of single saccades by up to 50% (Nyström & Holmqvist, 2010). Blinks also interfere with saccadic durations: when 20° saccades co-occur with a blink, their durations increase on average by 36% (Rottach, Das, Wohlgemuth, Zivotofsky, & Leigh, 1998) (horizontal only) and (Rambold, Sprenger, & Helmchen, 2002) (both horizontal and vertical).

Saccadic duration (in milliseconds, ms) is closely related to saccadic amplitude (in degrees visual angle, °). This relation has been found by Carpenter (1988) to be

$$\text{duration} = 2.2 \cdot \text{amplitude} + 21 \tag{10.6}$$

The measure is sensitive to the same weaknesses as saccadic amplitude of event detection algorithms. Collewijn *et al.* (1988) report that centrifugal saccades larger than 15° have significantly longer durations than corresponding centripetal saccades, and provide an adjustment to Carpenter's relationship:

$$\text{centripetal duration} = 2.5 \cdot \text{amplitude} + 27 \tag{10.7}$$

$$\text{centrifugal duration} = 3.9 \cdot \text{amplitude} + 13 \tag{10.8}$$

Saccadic duration is bimodally distributed during reading. Figure 10.15 shows a sharp peak for the short saccades between words and a softer hill for the return sweeps. There are no saccades with durations below 21 ms, corresponding to the constant in Equation (10.6). Also note how sensitive the measure is to the threshold setting of the velocity algorithm, not only for small saccades but across all durations. At a 30°/s setting there are twice as many saccades in the 50–100 ms interval as there are for the 90°/s setting. High peak velocity settings lead the algorithm to omit a considerable number of saccades.

The duration of a saccade has largely been thought of as a period with no visual intake (p. 379 and Dodge, 1900; Volkmann, 1986). During a saccade, stimuli can be manipulated without the participant noticing. Somewhat in opposition to the complete blindness view on saccades, a number of studies have been conducted to find out whether some cognitive processes can nevertheless take place during the progress of saccades. Irwin and Brockmole (2004) review these studies, the main findings of which we summarize below:

- Stimulus encoding is blocked during saccades
- Mental rotation appears to be suppressed
- It is unclear whether memory comparisons are suppressed
- Response selection is not suppressed

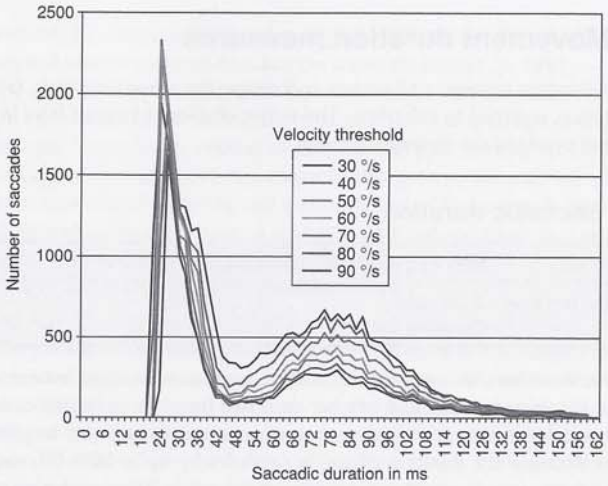


Fig. 10.15 Distribution of saccadic duration in reading at different peak velocity settings for the velocity algorithm. Bin size 1 ms. Study 1 on page 5. Recorded with a tower-mounted system at 1250 Hz with a high data quality. Saccade durations were calculated with BeGaze 2.1.

- Letter priming is not suppressed
- Lexical processing is not suppressed

In practice, saccadic duration is a measure frequently used in neurology and pharmacological papers, and only occasionally in human factors. The following factors have been found to increase saccade duration:

More difficult tasks For instance, saccades to remembered targets as well as antisaccades have strongly reduced peak velocities coupled with markedly increased durations compared to stimulus elicited saccades (Smit *et al.*, 1987). Saccade duration increases with increasing blur in an image, and can be used for evaluation of image quality, according to Vuori, Olkkonen, Pölönen, Siren, and Häkkinen (2004).

A decreased processing capacity Independent of stimuli, participants with schizophrenia and bi-polar disorder have longer saccade durations (Bestelmeyer *et al.*, 2006). 60 hours of sleep deprivation resulted in a 24% prolongation of saccade durations in a study conducted by Green and Farnborough (1986). McGregor and Stern (1996) report time on task effects on saccade duration. Saccadic duration is also increased by alcohol (Lehtinen, Lang, Jäntti, & Keskinen, 1979), but the changes in velocities and durations correlated more closely with feelings of intoxication than with blood alcohol concentrations. A large group of pharmacological and neurological journals use the saccadic duration measure to investigate the effect of drugs, psychological illness, and neurological injuries.

Cumulative transition time between fixations

Target question	*What is the cumulative duration of all saccades, as a percentage of the trial?*
Input representation	*A collection of saccades*
Output	*Percentage of time occupied by saccades*

In absolute terms, the cumulative transition time is defined as the sum of all saccade durations in a trial. The pragmatic idea behind the measure is that since vision is effectively shut down during saccades, a user interface or a pilot cockpit should be designed so that the cumulative transition time is minimized. Since saccade duration and amplitude correlate closely, this equates to having a user-interface design which results in producing few long saccades. For the same reason, cumulative transition time and scanpath length measures can be used interchangeably.

Cumulative transition time is a rare measure, used in human factors and ergonomics research. Abbott, Nataupsky, and Steinmetz (1987), who introduce the measure, found significant differences in the measure between two different designs (averages 18.6 versus 12.6% of trial time).

10.3.2 Scanpath duration

Target question	*How long did the scanpath take, from onset to offset?*
Input representation	*A scanpath*
Output	*Scanpath duration (s)*

Scanpath duration is defined as the time from onset of the scanpath until offset. The major challenge is to decide the exact points at which the scanpath starts and ends.

In usability research, scanpath duration measurements, including on- and offsets, are defined as "the sum of all the fixation duration times whilst completing a task" (Cowen, Ball, & Delin, 2002). The originators of the measure, Goldberg and Kotval (1999), do not specify the on- and offset conditions in sufficient detail. In practice, therefore, this means that scanpath duration more or less equals task completion time—but then the question becomes, how can we be sure when engagement in the task has commenced and ceased?

In reading research, several varieties of the scanpath duration measure have been developed for the study of sentence and global text processing. The following two examples are taken from Hyönä *et al.* (2003) and Liversedge, Paterson, and Pickering (1998), which both provide a methodological discussion on scanpath duration measures in reading. Note how elaborate the conditions for scanpath on- and offset are.

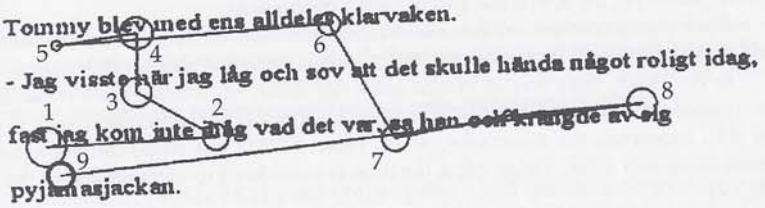


Fig. 10.16 Piece of data from a participant reading a longer text. Fixation number 2 on the word (ihåg) is followed by a regression scanpath two lines up. The regression scanpath ends at the 7th fixation, when the eye reaches a word beyond the last new one looked at. In this case go-past duration is the duration of the regression scanpath from after the second until the seventh fixation. Recorded 2000 with an SMI RED II, 50 Hz. Fixations detected with the I-DT algorithm using a 60 ms duration threshold.

Go-past duration is defined as the duration from first entering an AOI—in this case a word—until moving *forward* beyond the AOI, in the direction of the language read. Figure 10.16 shows a go-past scanpath with four fixations, where the difficulty was in resolving the

Swedish pronoun 'jag' ('I' in English), the second word on the third line, and overcoming it required looking back at the proper name 'Tommy' (this interpretation of the scanpath was verified by a retrospective interview with the participant immediately after recording). The closely related measure 'regression path duration' is part of the go-past duration; this is because go-past duration also includes eye movements which return to parts of the sentence already read. The go-past duration is considered to reflect difficulties in integrating a word within a context. This may be an early effect of inadequate initial processing, but it may also reflect overcoming comprehension difficulty despite sufficient processing initially, in which case it should be seen as a late effect (Clifton, Staub, & Rayner, 2007).

The *Extended first-pass fixation time* is the dwell time on a sentence or word AOI including the duration of regression scanpaths. The idea is that inconsistencies in the AOI can be better verified in eye-tracking data if the regression time—time to resolve the inconsistency—spills over to that AOI. In our example, the dwell time of the word 'ihåg' ('remember'), underneath the second fixation, would be increased by the duration of the entire regression scanpath (up until fixation 7). Be careful that the measure may be misinterpreted as 'the participant needed to read all this to understand the text they left', while in fact he has only lost a term from working memory.

10.3.3 Blink duration

Target question	*For how long is the eye closed during the blink?*
Input representation	*A blink*
Output	*Blink duration (ms)*

Blink duration is the complete time from when the eyelid starts moving down until it is fully up again. Real eyelid trackers record data as shown in Figure 10.17. Blink duration is only one of a number of duration measures that can be defined with data from an eyelid tracker. Data and software from pupil and corneal reflection eye-trackers operationalize blinks differently (p. 176), and typically underestimate blink duration. In addition, the commercial algorithms calculating blink duration tend to interpret noise in the data as blinks. High-speed optic artefacts in particular may cause an abundance of very short blink durations, as can be seen in Figure 10.18. Summary statistics such as average calculations should not be made before removing the artefactual part of such a distribution.

Blink duration values exhibit considerable individual variation across all drowsiness levels (Ingre, Åkerstedt, Peters, Anund, & Kecklund, 2006) and may be idiosyncratic.

In the 1940s, there was an intense academic debate on whether blink duration could be associated with anything like visual fatigue. Luckiesh (1947) and others were at the centre of this, supporting the association, while Tinker (1945) and others did not agree that this association was valid. Today, blink duration is considered to increase due to the following factors:

Drowsiness Blink duration and its close relative 'eye cleft' are among the most reliable camera readable signs of drowsiness. Re-opening time, in particular, changes reliably with increasing drowsiness. As an example, bus drivers with the obstructive sleep apnea syndrome have significantly longer blink durations compared to controls (Häkkänen, Summala, Partinen, Tiihonen, & Silvo, 1999).

Loss of vigilance There is an increase in blink duration during sustained attention tasks that correlates with decreased performance (Morris & Miller, 1996).

Mental workload Blink duration is highly correlated with object tracking errors of participants (Van Orden *et al.*, 2000), in agreement with research conducted by Stern, Boyer,

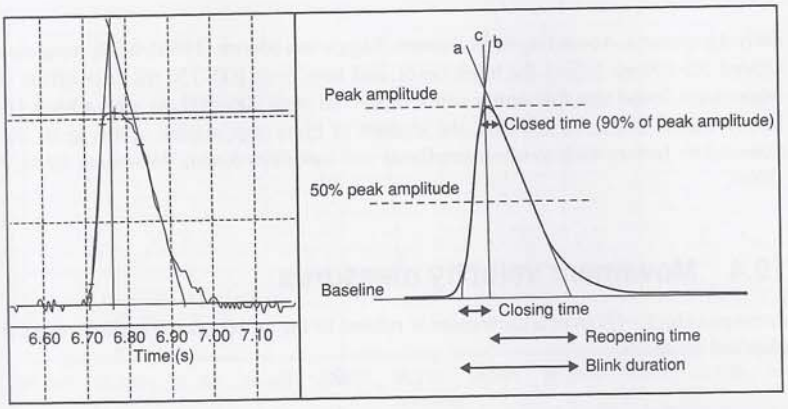


Fig. 10.17 To the left a plot of a blink recording, with the lid closure as the vertical amplitude dimension. To the right, blink duration and other blink measures defined on the background of a blink amplitude curve. With kind permission from Springer Science+Business Media: *European Journal of Applied Physiology*, Experimental evaluation of eye-blink parameters as a drowsiness measure, *89*(3), 2003, Phillipp P. Caffier.

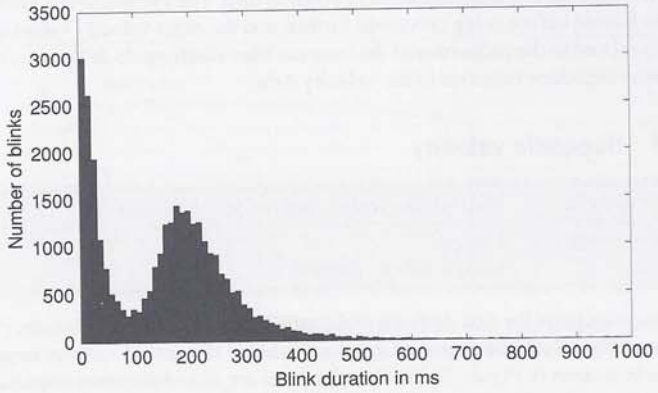


Fig. 10.18 The histogram of blink durations during mathematical problem solving, as measured with a tower-mounted 1250 Hz system, and analysed in BeGaze 2.3. Bin size 10 ms. Total number of blinks in this histogram is 32700, of which 203 are above 1000 ms in duration. The peak below 80 ms shows that this particular blink detection algorithm accepts various measurement noise as blinks.

and Schroeder (1994). Morris and Miller (1996) found that blink duration increases as a function of time on task. However, Veltman and Gaillard (1996), in a study of pilots, conclude that blink duration is affected by the visual demands of the task rather than by the cognitive workload in general.

Alcohol and anaesthetic sedation Blink duration appears to be sensitive also to low levels of sedation (Jandziol, Prabhu, Carpenter, & Jones, 2006) and alcohol (Biederman *et al.*, 1974).

Like saccadic suppression, where visual intake is reduced before the physiological movement of the eye begins, there is a similar phenomena for blinks known as *blink suppression*. Ridder III and Tomlinson (1995) propose that these mechanisms are produced by similar un-

derlying systems. According to Volkmann, Riggs, and Moore (1980), blink suppression starts around 50–100 ms before the blink onset, and lasts until 100–150 ms after offset. However, others have found that full acuity is not recovered until 200–500 ms after a blink (Ehrmann, Ho, & Papas, 2005). In addition, the amount of blink suppression seems to be directly influenced by factors such as blink amplitude and task (Stevenson, Volkmann, Kelly, & Riggs, 1986).

10.4 Movement velocity measures

Average velocity ($\bar{v}$) over a movement is related to the amplitude (θ) and duration (t) by the classical equation

$$\bar{v} = \frac{\theta}{t} \tag{10.9}$$

Velocity levels can change along a movement, and that is why average velocity for one eye movement event (e.g. a saccade) is only of marginal use. Instead the instantaneous, *tangential velocity* ($\dot{\theta}$) is approximated by the distance between consecutive raw data samples (θ) multiplied by the sampling frequency of the eye-tracker $f_s = \frac{1}{\Delta t}$. Mathematically, this calculation of ($\dot{\theta}$) corresponds to a differentiation of position data. The instantaneous velocity is typically lowpass filtered before being processed further, and the exact velocity values are therefore intimately related to the properties of the lowpass filter (Inchingolo & Spanio, 1985). Note that filters *may* introduce latencies in the velocity data.

10.4.1 Saccadic velocity

Target question	*What was the peak/average velocity of the saccade?*
Input representation	*A saccade*
Output	*Saccade velocity (degrees/s)*

Saccadic velocity is the first derivative of position data with respect to time. Saccadic velocity is typically calculated as part of event calculation (Chapter 5). An example of a velocity plot can been seen in Figure 10.20. Velocity plots are one of the most important data inspection tools in eye-tracking research. Precision problems, optic artefacts, and the successes and failures of event detection algorithms can be readily inspected in them.

Träisk *et al.* (2005) showed that search coils reduce the velocity of saccades, in particular for longer amplitudes. Saccades are therefore probably most accurately recorded with video-based pupil and corneal reflection eye-trackers. Furthermore, McGregor and Stern (1996) found that saccades occurring during a blink were significantly slower than those occurring independently of a blink. An example of a blink-accompanying saccade is given in Figure 10.19.

In Figure 10.20, we can see the velocity curves for three saccades that develop very differently in time. The first saccade is a fairly common type. The velocity peak appears slightly before the middle of the saccade, so that the acceleration phase is faster than the deceleration phase. Two small glissadic movements are appended to this saccade. The middle saccade in Figure 10.20 is what has often been considered the ideal velocity curve. It has a clear velocity peak with only a slightly faster acceleration than deceleration. The glissade is minimal. The third saccade has multiple velocity peaks, of which the first two are saccadic and the last one possibly a glissade. The eye moves in spurts, but is never completely still between each burst. Zivotofsky, Siman-Tov, Gadoth, and Gordon (2006) report very similar data from

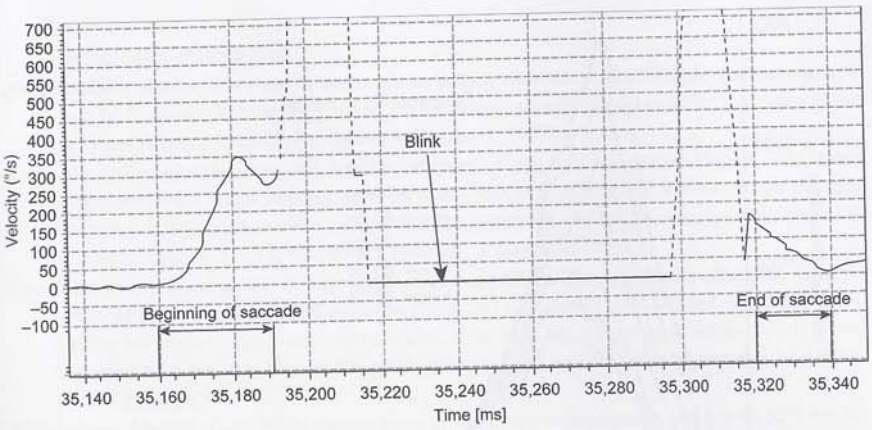


Fig. 10.19 180 ms saccade (black velocity) coinciding with a blink (dashed velocity). The saccadic velocity peaks at 35,180 ms, and the blink takes over at 35,190 ms (dashed artefactual velocity line). At 35,340 ms, note how the now prolonged saccade ends, around 180 ms after starting. Recorded during the mathematical problem solving task, number 4 on page 5, using a tower-mounted system at 1250 Hz. The data quality was high.

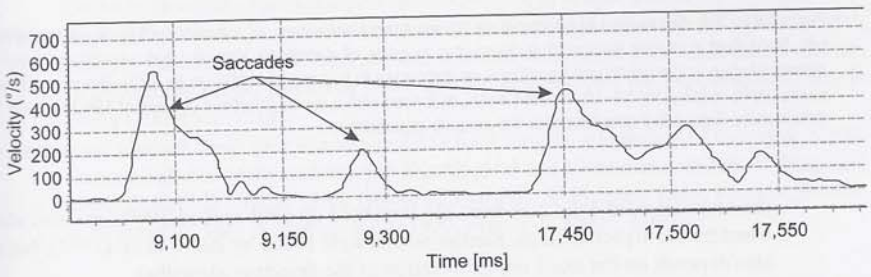


Fig. 10.20 Three saccades by one participant shown in a velocity over time diagram. Data taken from the mathematical problem solving task, number 4 on page 5, recorded with a tower-mounted 1250 Hz eye-tracker. The data quality was high.

a participant with Stiff-Person Syndrome, where the multiple peaks are most likely a result of incomplete muscle control. Rucker *et al.* (2004) link these multi-step saccades to Tay-Sachs disease. The data in Figure 10.20 are from a presumably healthy participant—a young student—but he was the only person from over 300 participants in this reading study who exhibited this pattern. The saccade is curved, and different velocities occur along the curved trajectory, possibly as the three pairs of eye muscles alternate controlling the eye.

Three common values are derived from velocity data such as in Figure 10.20:

1. Average saccadic velocity is an average of velocities over the entire duration of a saccade. Averages are however poor representations of the Gaussian-shaped saccadic forms.
2. Peak saccadic velocity is the highest velocity reached during the saccade. Average peak saccadic velocity reported by your software may be heavily affected by the peak velocity threshold in the event detection algorithm, as Figure 10.21 illustrates. Not only are saccades with a peak velocity lower than the threshold excluded—a large number of

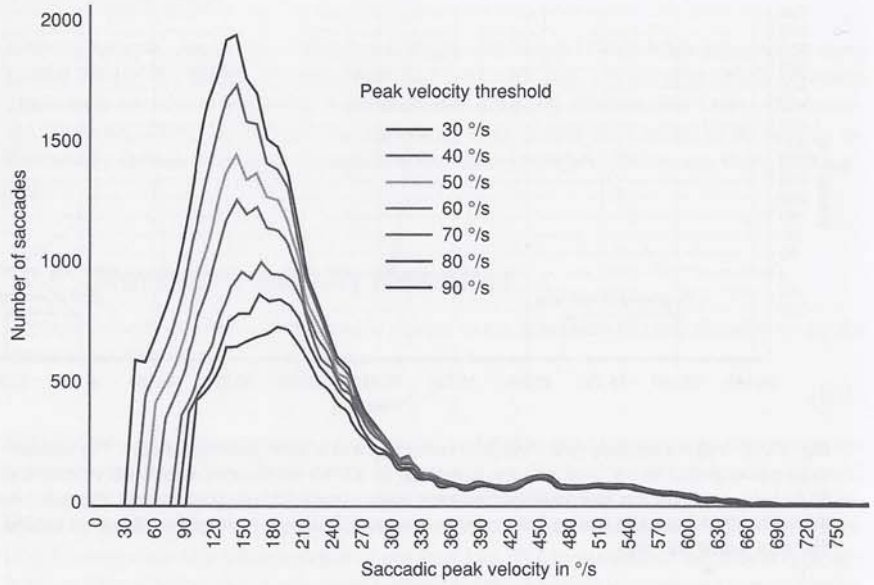


Fig. 10.21 The distribution of saccades at different peak velocities, as a function of peak velocity threshold. Note that a higher threshold reduces the number of saccades also at peak velocities *above* the threshold, due to the additional conditions on saccades in this implementation (p. 173). Bin size 10°/s. 10 participants reading for 10–15 minutes each, recorded with a tower-mounted system at 1250 Hz with high data quality. Saccades detected with the velocity algorithm.

the saccades with velocities from the threshold up until around 300°/s are also eliminated by the higher settings. Similar results were found by Bahill *et al.* (1981), but this also depends on the exact implementation of the detection algorithm.

3. Time to peak is the duration from the onset of a saccade until the peak velocity is reached.

The highest recorded peak saccadic velocity tends to be around 1000°/s, as Figure 10.12 on page 317 shows for a large saccade population during mathematical problem solving. Such fast saccades are very rare, however. In the peak velocity histogram in Figure 10.21, even the fastest reading return sweeps across the entire monitor have velocities no larger than approximately 700°/s.

The lower end of peak saccadic velocities is less investigated, although of great importance in event detection algorithms. On page 171–172, we concluded that a proper setting is about 30–40 °/s for a high-speed system with good precision. If there are long fixations in the data, there is a certain risk that occasional microsaccades will be detected, as a few microsaccades have peak velocities above 50°/s (Engbert, 2006).

Saccadic velocity has been used as a measure of cognitive activation level, or what is often called arousal level. Circumstances and factors that influence saccadic velocity include:

Arousal levels and sleepiness Low vigilance decreases saccadic velocity (Galley, 1989), and so does tiredness (McGregor & Stern, 1996; Becker & Fuchs, 1969) and sleep deprivation (Russo *et al.*, 2003; Bocca & Denise, 2006). However, McGregor and Stern (1996) present results that suggest caution in interpreting saccadic velocity change as an index of 'fatigue', since the reduction in average saccadic velocity may be secondary

to increases in blink rate.

Anticipation Anticipatory saccades, made to targets that are so predictable that the saccade can be launched before target onset, have lower velocities than reactive saccades (Smit & Van Gisbergen, 1989; Bronstein & Kennard, 1987).

Task Saccadic velocity increases as the difficulty of the task increases (Galley, 1993) and decreases with an increasing time on task (McGregor & Stern, 1996). When the task requires a higher saccadic rate (greater frequency of saccades), the saccadic peak velocity increases (Lueck, Crawford, Hansen, & Kennard, 1991).

Age Saccadic velocities are of the same size in children as with adults (Salman, Sharpe, Eizenman, *et al.*, 2006), and Abrams, Pratt, and Chasteen (1998) found velocities not to differ between younger and older adults. However, Moschner and Baloh (1994) found velocities to be 20% slower for participants older than 75 years compared to participants younger than 43.

REM sleep During sleep, REM saccades—i.e. rapid eye movements—are about half the velocity of equal amplitude saccades made when awake (Aserinsky, Joan, Mack, Tzankoff, & Hurn, 1985). However, a more recent study could not confirm this finding (Sprenger *et al.*, 2010).

Melancholia This is a disorder of low mood and lack of enthusiasm, and is associated with a difficulty in increasing peak velocities as target amplitudes increase (Winograd-Gurvich, Georgiou-Karistianis, Fitzgerald, Millist, & White, 2006).

Neurological disorders Slow saccades can be an indication of lesions in the pons, the midbrain, or the basal ganglia. Low saccade velocities also occur with Alzheimer's disease, AIDS, certain drugs, and a few other specific diseases. See the excellent summary in Wong (2008) for details.

Drugs and alcohol Peak saccadic velocity has for a long time been one of the prime oculomotor measures when studying the neurological and behavioural effects of drugs and alcohol (Abel & Hertle, 1988; Griffiths, Marshall, & Richens, 1984; Jürgens, Becker, & Kornhuber, 1981; Lehtinen *et al.*, 1979; Franck & Kuhlo, 1970).

10.4.2 Smooth pursuit velocity

Target question	*What was the peak or average velocity of the smooth pursuit event?*
Input representation	*A smooth pursuit event*
Output	*Velocity (degrees/s)*

Smooth pursuit velocity is calculated just like saccadic velocity. Velocity plots on pages 169 and 179 show smooth pursuit velocity from a participant watching a pendular movement. We differentiate between average and peak smooth pursuit velocity.

It is generally thought that smooth pursuit velocity, peaking at 25–40°/s, is much slower than saccadic velocity (Boff & Lincoln, 1988; Young, 1971). However, Meyer *et al.* (1985) recorded 100°/s smooth pursuit on ordinary participants. Measuring professional baseball players who simulated hits on a baseball, Bahill and LaRitz (1984) recorded smooth pursuit velocities of up to 130°/s. According to Bahill and LaRitz, p. 235, "The success of good players is due to faster smooth pursuit eye movements, a good ability to suppress the vestibulo-ocular reflex, and the occasional use of an anticipatory saccade". Due to their large overlap in velocity range, smooth pursuit and saccades may be hard to separate by considering velocity alone. When the smooth pursuit follows targets moving with velocities of greater than about 30°/s, there tend to appear catch-up saccades in the data.

Smooth pursuit velocity has been argued to reflect the following:

Path curvature When the target moves in curved paths, smooth pursuit velocity decreases with decreasing radii of the curves (De'Sperati & Viviani, 1997).

Age Smooth pursuit velocity is slower for older participants (mean 67 years) than for younger (mean 42) (Sharpe & Sylvester, 1978). Newborn children have an undeveloped smooth pursuit, and have difficulty tracking targets even at low speeds such as 25°/s (Kremenitzer, Vaughan Jr, Kurtzberg, & Dowling, 1979).

Drugs Few studies have been conducted on the effects of drugs, and often with negative results. For instance, Tedeschi, Bittencourt, Smith, and Richens (1983), contrary to expectations, found no effect of amphetamine on smooth pursuit velocity.

Disorders Children with autism (Takarae, Minshew, Luna, Krisky, & Sweeney, 2004), adults with a childhood history of physical and emotional abuse (Irwin, Green, & Marsh, 1999), as well as patients with schizophrenia and post traumatic stress disorder (Cerbone *et al.*, 2003) all show a decreased ability to smoothly track targets at higher velocities.

10.4.3 Scanpath velocity and reading speed

Target question	*What was the velocity of the scanpath?*
Input representation	*A scanpath, consisting of a sequence of saccades*
Output	*Velocity (degrees/s)*

Scanpath velocity (which you may also see referred to as 'average saccadic velocity', 'eye movement speed', or 'eye velocity') is defined as the product of average saccadic amplitude (in °) of saccades of which the scanpath is comprised, and saccadic rate (in 1/s). Apart from being a measure of scanpath velocity, this measure allows for a crude approximation of average saccadic velocity for data collected with such low sampling frequency and poor precision that actual saccadic velocities cannot be calculated.

The measure was defined by Saito (1992), who compared monitor work (23°/s) with similar work without a monitor (9°/s).

In reading studies, an alternative calculation of the same scanpath velocity has been made as the product of reading speed (in characters per second) and letter size (in visual degrees per character). The originators of this calculation, Krischer and Zangemeister (2007), investigated optimal conditions for reading, and conclude that the best skill- and acuity-matched letter size gives an eye movement speed of 8°/s during reading. Small letters, in particular, slow down the eye movement speed.

Beymer, Russell, and Orton (2005) instead divided text distances read by the time it took to cover them. This velocity measure was used to compare paragraph widths, and the authors found that a 4.5 inch monitor text is read only slightly faster than a 9.0 inch text.

An ambitious operational definition of reading speed (RS) was provided by Bullimore and Bailey (1995), who defined it as

$$\mathrm{RS} = \mathrm{FR}\frac{\#\text{forward saccades}}{\#\text{total saccades}}(\text{average saccadic amplitude in letters}) \tag{10.10}$$

where FR is fixation rate (p. 416). The measure was used to study participants with macular degeneration and the effect of luminance, which both have effect on reading speed.

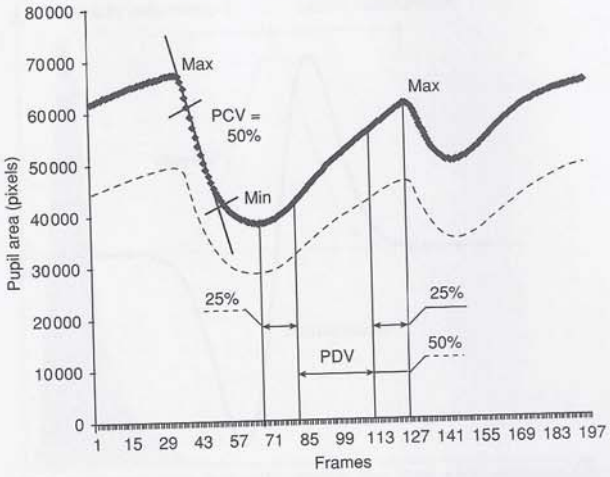


Fig. 10.22 Broken line, left eye; unbroken line, right eye. Max, maximum pupil area; Min, minimum pupil area. PCV, pupil constriction velocity, is calculated from the middle portion of the curve between Max and Min; PDV, pupil dilation velocity, is calculated from the middle portion of the curve between Min and Max. Data from a healthy participant about 60 years of age watching a variable light source. Recorded using a custom-built 60 Hz pupillometer. With kind permission from John Wiley and Sons: *Acta Ophthalmologica*, *85*(5), 2007, Lada Kalaboukhova, Vanja Fridhammar, and Bertil Lindblom pp. 519–525.

10.4.4 Pupil constriction and dilation velocity

Target question	*What was the velocity of the pupil closure or opening movement?*
Input representation	*Raw sample data*
Output	*Velocity (mm/s or mm^2/s)*

In pupillometry, pupil velocity is an established dependent variable. Pupil constriction velocity is approximately three times faster than dilation velocity (Ellis, 1981), so these should be measured separately. Figure 10.22 shows a pupillometric recording, and demonstrates what portions of data should be selected for pupil velocity calculations.

Average pupil velocity is calculated by selecting a constriction or dilation period with constant velocity and dividing the change in pupil diameter or area by the duration of the period (Figure 10.22). *Tangential pupil velocity* is calculated either by differentiating the horizontal diameter by time (Bitsios, Prettyman, & Szabadi, 1996), or by differentiating the pupil area by time (Figure 10.22). After tangential velocity has been calculated, *maximum pupil velocity* can easily be calculated.

Pupil velocity measurements have mainly been used clinically, for instance to assess pupil parasympathetic function, effects of medication, or the effect of glaucomas on pupil dilation. Age differences are known. For instance, Bitsios *et al.* (1996) found a smaller maximum dilatation velocity in an elderly compared to a younger group.

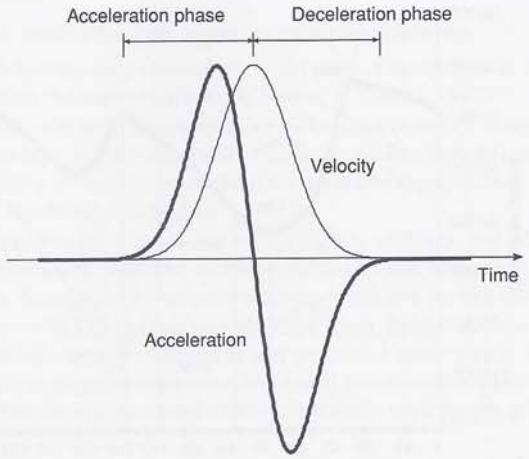


Fig. 10.23 Acceleration (thick line) and velocity (thin line) over time for an ideal saccade. Note that acceleration rises before velocity; and that acceleration is 0 when velocity peaks. Also note that during the deceleration phase, acceleration is negative.

10.5 Movement acceleration measures

The instantaneous, tangential acceleration $\ddot{\theta}$ is calculated by differentiating velocity $\dot{\theta}$ with respect to time:

$$\ddot{\theta} = \frac{d}{dt}\dot{\theta} \tag{10.11}$$

Mathematically, acceleration is therefore the second derivative of position data. Peak and average acceleration can be retrieved from the continuous acceleration data. Deceleration—slowing down—can be seen as negative acceleration (Figure 10.23). Jerk ($\dddot{\theta}$) is the third derivative, measuring the change in acceleration over time.

Movement acceleration is necessary to start a movement and to increase the velocity of it. Acceleration and jerk drive velocity, and therefore a change in acceleration are accompanied by a change in velocity. Unfortunately, numerical differentiation magnifies noise in the signal, and may require additional filtering.

10.5.1 Saccadic acceleration/deceleration

Target question	*What was the peak/average acceleration of the saccade?*
Input representation	*A saccade*
Output	*Acceleration (degrees/s^2)*

Saccadic acceleration is the derivative of saccadic velocity with respect to time. Acceleration thresholds are used in some of the saccade detection algorithms (most prominently in the EyeLink parser from SR Research), and are valuable to separate smooth pursuit movements (low acceleration due to small variation in velocity) from saccades (high acceleration/deceleration at onset/offset).

Saccadic acceleration rises very fast, and reaches a maximum peak of up to 100000°/s^2, with typical peak values ranging between 6000–12000°/s^2. Extremely high acceleration values may be an indication of optic artefacts in the data. As Figure 10.24 shows, the distribution has a rightward skew.

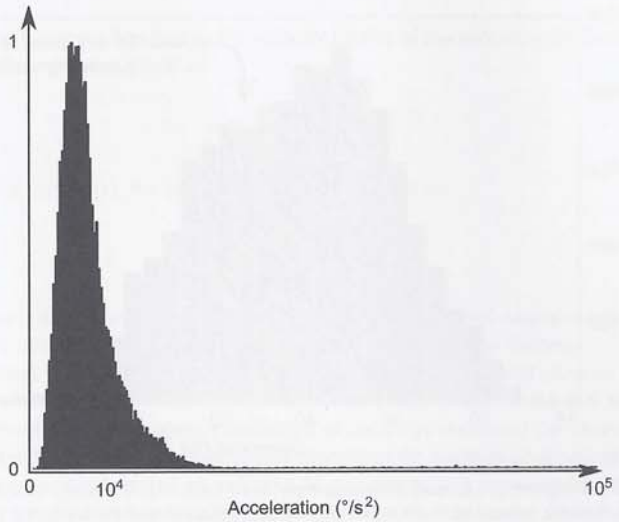


Fig. 10.24 A normalized distribution of saccades at different peak accelerations. Acceleration is calculated using the algorithm by Nyström and Holmqvist (2010). Ten participants reading for 10–15 minutes each, recorded with a tower-mounted 1250 Hz system. Data quality is high.

Acceleration and deceleration were found by Collewijn *et al.* (1988) to increase as a function of saccade amplitude; however, data were recorded with coils, not the type of commercial video-based eye-trackers in common use in research today.

Saccadic acceleration is a very uncommon measure that appears to have attracted mainly neurologists with an interest in the cerebellum and superior colliculus (important saccade programming areas of the brain). All participants in Straube and Deubel (1995) showed an idiosyncratic pattern of saccadic acceleration and deceleration.

The existence of glissades and saccades with multiple velocity peaks means that some saccades have more than one acceleration phase. Whether this also means that the eye muscles are actually pulling the eye during all acceleration phases, or whether some acceleration phases are the consequence of eye lens inertia during deceleration still remains an open question.

10.5.2 Skewness of the saccadic velocity profile

Target question	*How much of the saccadic duration is taken up by acceleration and deceleration phases, respectively?*
Input representation	*A saccade*
Output	*Skewness*

The skewness of the saccadic velocity ('saccadic skewness') is defined as the degree of skewness of the velocity plots of saccades. It attempts to measure the duration of the acceleration versus deceleration phases in saccades, as shown in Figure 10.23.

The measure has been operationalized in at least three different ways: First, Collewijn *et al.* (1988) define the skewness value as the acceleration phase (time to peak velocity) divided by the total saccade duration. A symmetrical saccade therefore has a skewness of 50%. Figure

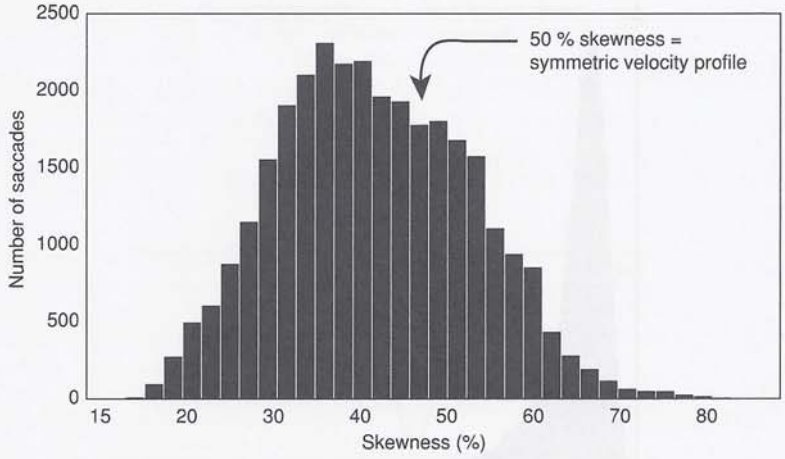


Fig. 10.25 Histogram of saccadic skewness from more than 30,000 reading saccades from project 1 on page 5. Skew is defined as the duration of the acceleration phase divided by the total duration of the saccade. Bin size 2%. Recorded with a tower-mounted eye-tracker at 1250 Hz. Saccades detected using the velocity algorithm at a threshold setting of 30°/s. Glissades are largely included in saccades.

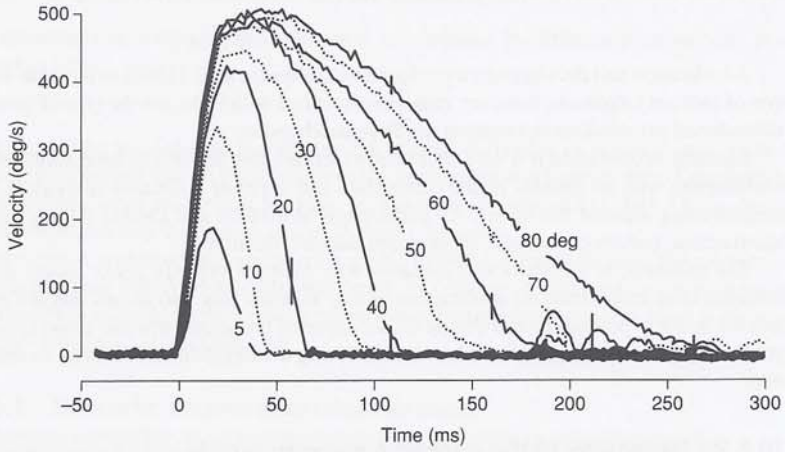


Fig. 10.26 The skewness of a saccade has previously been shown to be dependent on the saccadic amplitude. Data recorded with a coil system. Reprinted with kind permission from John Wiley and Sons: *Journal of Physiology*, Binocular co-ordination of human horizontal saccadic eye movements, *404*(1), 1998, Collewijn, H., Erkelens, C.J., & Steinman, R.M., pp. 157–182.

10.25 shows a histogram of saccadic skewness using this calculation. A similar operational definition used by Straube and Deubel (1995) and others is to calculate the acceleration phase divided by the deceleration phase. Symmetrical saccades in this case have the value 1; right-skewed are lower, and left-skewed (the few that exist) have a value above 1. Using this, the distribution of skewness values in Figure 10.25 will be even more skewed than as depicted. A third operational definition was provided by Van Opstal and Van Gisbergen (1987) who

approximated a gamma function to the velocity profile of the saccades. By finding values for α, β, and γ that optimize the fit of

$$v(t) = \alpha \cdot (\frac{t}{\beta})^{\gamma-1} e^{-(\frac{t}{\beta})} \tag{10.12}$$

to the velocity curve $v(t)$, the skewness can be calculated as

$$\text{Skew} = \frac{2}{\sqrt{\gamma}} \tag{10.13}$$

Skewness is a little used measure. Anticipatory (or predictive) saccades were found to be slightly more skewed than visually guided saccades (Smit & Van Gisbergen, 1989). Liao *et al.* (2006) found that head saccades (turning the head) have a constant skew of 0.5, while the skew of eye saccades varies. Soetedjo, Kaneko, and Fuchs (2002) found that injection of the substance muscimol in the superior colliculus of monkeys increased the skewness since the duration of the deceleration phase increased more than the duration of the acceleration phase.

In previous research based on scleral coil recordings, the acceleration phase has been shown to be of approximately the same duration across all amplitudes, while the deceleration phase increases rapidly with increasing amplitudes up to 90° (Figure 10.26). However, using the data described on page 5, from video-based eye-trackers, we have not been able to find any correlation between amplitude and the skewness of the saccadic velocity profile for saccades up to 40°.

The many glissades that exist in video-based eye-trackers (but are suppressed with coil systems) are categorized with the saccades by some of the detection algorithms. In effect, this means that glissadic saccades of whatever amplitude will be heavily skewed. In fact, the skewness measure will itself have a skewed distribution, as shown in Figure 10.25. The skew of the skewness distribution *may* indirectly reflect the amount of glissadic saccades, depending on your particular event detection algorithm.

10.5.3 Smooth pursuit acceleration

Target question	*What was the peak/average acceleration of the smooth pursuit event?*
Input representation	*A smooth pursuit event*
Output	*Acceleration (degrees/s^2)*

Using targets moving at 40°/s, Kao and Morrow (1994) found smooth pursuit acceleration values of up to 350°/s^2. Acceleration values were significantly higher when the target motion was predictable, yet more evidence of the anticipatory nature of smooth pursuit. Moschner *et al.* (1999) reported smooth pursuit accelerations between 43–128°/s^2 across individuals over the first 60 ms of "visually guided smooth pursuit". During smooth pursuit with constant velocity, the acceleration is zero.

10.5.4 Saccadic jerk

Target question	*What was the peak/average jerk of the saccade?*
Input representation	*A saccade*
Output	*Jerk (degrees/s^3)*

The saccadic jerk ($\dddot{\theta}$) is the derivative of saccadic acceleration, and thus the third derivative of gaze position. For this reason, expect jerk data to be noisy. Alternative terms are 'jolt', 'surge', and 'lurch'. Wyatt (1998) proposes a saccade detection algorithm based on jerk instead of velocity or acceleration. Since saccades mostly start off very sharply, jerk is particularly useful for determination of precise saccadic onset; Wyatt suggests the simple jerk threshold of $200,000°/s^3$. Jerk is not very helpful in determining saccadic ending, because of the slow glissades that are often appended to saccades.

10.6 Movement shape measures

Shape refers to the form and figure of movements, such as the *curvature* of saccades and glissades, the *smoothness* of pursuit, the *selfcrossing* of scanpaths, and shapes associated with specific cognitive processes such as *reading*, or *local versus global* scanning. Although shapes are easily discernible for human viewers, 'shape' is a vague concept, and can therefore be hard to define precisely. No attempts have been made to categorize or quantify all possible movement shapes. There are undoubtedly countless possible shapes of movements, but so far only occasional shape measures have been developed.

10.6.1 Saccadic curvature

Target question	*What is the curvature of the saccade?*
Input representation	*A saccade*
Output	*Curvature (degrees or more complex units)*

Saccadic curvature is a measure of the overall spatial shape of the saccade. Few saccades are perfectly linear, and some really large saccadic curves can be elicited by having a participant circle their gaze quickly between four dots in a square, which is how the saccades in Figure 10.27 were generated.

There are several different operational definitions of saccadic curvature. Port and Wurtz (2003) use standard circular statistics, and define the saccadic curvature as the angular standard deviation of the tangential direction of the saccade. The curvature unit will then be in orientation degrees (like φ, p. 302). Attempts have also been made to use polynomial curve fitting, and Ludwig and Gilchrist (2002) found a quadratic curvature metric to provide a robust fit, where the polynomial coefficients describe the curvature. Harwood, Mezey, and Harris (1999, p. 9098) define a measure called the *spectral main sequence*, which is claimed to be "exquisitely sensitive to the saccade trajectory and should be used to test objectively all present and future models of saccades."

It is clear however, that curved saccades often have multiple velocity peaks, since changes in saccadic direction co-occur with local velocity minima.

The properties and influences on saccadic curvature as known from research can be summarized as follows:

Oblique saccades These are more curved than horizontal and vertical saccades. The curvature of oblique saccades compensates for the divergence of the initial direction from the direction called for by the target (Becker & Jürgens, 1990). Note that in oblique saccades two pairs of muscles are involved, and may change their relative influence on the movement over time.

Curve direction This is dependent on the *location of the target* and the *type of saccade* (Smit & Van Gisbergen, 1990; Viviani *et al.*, 1977).

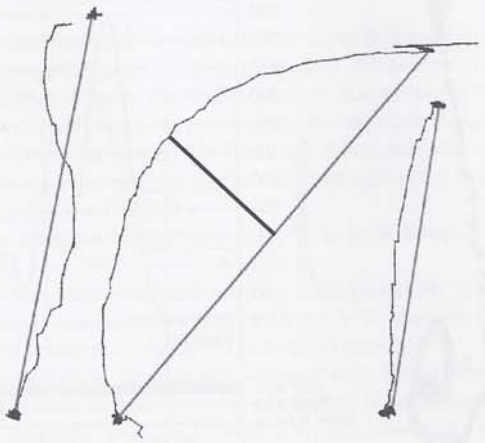


Fig. 10.27 Three saccades with varying degrees of curvature. Grey lines show the shortest distances from saccadic origin to endpoint. The thick black line shows the operational definition of saccadic curvature by Nummenmaa, Hyönä, and Calvo (2008). Recorded with a tower-mounted eye-tracker at 500 Hz; data quality is high.

Irrelevant distractors When near the saccade target, visual distractors also cause curved saccades. Curvature is always directed *away from* the distractor, and this finding also holds for verbal distraction (Doyle & Walker, 2001). Even emotional content may curve a saccade (Nummenmaa *et al.*, 2008).

Instructions When told to look at something other than the saccade target, instructions also affect saccadic curvature (Sheliga, Riggio, Craighero, & Rizzolatti, 1995).

Pharmacological effects For example, injection of muscimol into the superior colliculus was found by Aizawa and Wurtz (1998) to increase the curvature of saccades in monkeys. The authors argue that this can been taken as evidence that the superior colliculus is critically involved in regulating the trajectory of saccades during their execution.

10.6.2 Glissadic curvature

Target question	*What is the curvature of the glissade?*
Input representation	*A glissade*
Output	*Curvature in ° or more complex units*

Glissadic curvature refers to the path of a glissadic eye movement, as exemplified in Figure 10.28. Glissades are typically more curved than saccades, perhaps since they may lack controlled and prototypical programming, and hence can take on a variety of shapes.

Calculation of glissadic curvature is in principle the same as for saccadic curvature. However, the greater degree to which glissades curve must be taken into account. Average circular variance certainly works, using polynomial curves is a possible option, but one cannot use the maximum distance from the saccade to the line spanning the Euclidean distance between glissadic onset to offset, as with saccades (see Figure 10.27, middle curve).

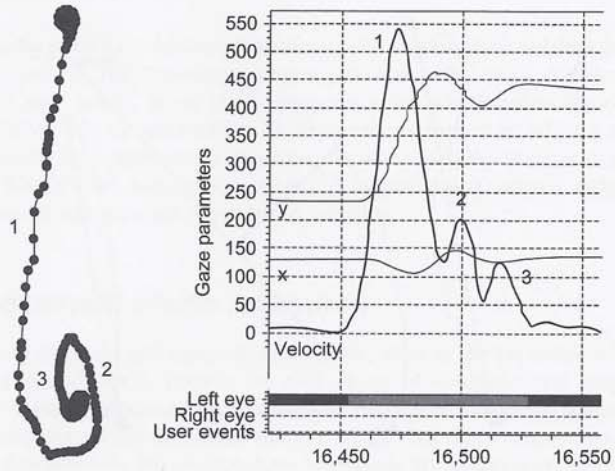


Fig. 10.28 Glissadic curvature: A saccade (1) followed by a double glissade (2 and 3), shown as raw data sample plot over stimulus (left), and velocity plot (right), recorded with a tower-mounted system at 1250 Hz; data quality is high.

10.6.3 Smooth pursuit: degree of smoothness

Target question	*How 'smooth' is the pursuit event?*
Input representation	*A smooth pursuit event*
Output	*Smoothness value*

The degree of smoothness of a smooth pursuit refers to whether the pursuit movement is smooth, or jagged (e.g. as in square-wave jerks). Smoothness ratings are typically performed by experts, and inter-rater reliability is then reported. They normally do not take into account catch-up saccades, but if catch-up saccades are included, the frequency of these can also be used as a measure of smoothness. This strategy was used by Mather and Putchat (1983) to evaluate pursuit in patients with schizophrenia.

Diefendorf and Dodge (1908), and Holzman, Proctor, and Hughes (1973) provide evidence that schizophrenic patients and their relatives demonstrate disordered smooth pursuit eye movements. Later studies have shown that schizophrenia-related smooth pursuit deficits are more likely due to an inability to correctly *execute* eye movements which track a stimulus, rather than an inability to correctly *encode* velocity information (Avila, Hong, Moates, Turano, & Thaker, 2006). Studies using the smoothness of pursuit have developed into a paradigm of eye-movement research.

10.6.4 Global to local scanpath ratio

Target question	*How much overview scanning and detailed inspection?*
Input representation	*All saccades in a scanpath*
Output	*Ratio*

First introduced by Groner *et al.* (1984), the ratio of global to local saccades (g/l) of which a scanpath is comprised is calculated as the number of saccades *longer* than a threshold

amplitude, divided by the number of saccades *shorter* than a threshold amplitude. The former threshold is set to determine 'global' scanning saccades, while the latter is set to identify 'local' saccades indicative of more detailed inspection. Zangemeister, Sherman, and Stark (1995) suggest 1.6°, 4.6°, 7.9°, or 11° as possible thresholds, while Zangemeister and Liman (2007) use 1.1° to separate the two types of saccade. Since saccadic amplitude correlates strongly with saccadic velocity, this type of threshold operates much like the velocity-based saccade detection algorithms in Chapter 5.

A larger g/l ratio indicates a longer scanpath, fewer local fixations, and faster average eye movement speed.

The downsides of this measure are the arbitrary settings, and the fact that saccades scale with the size of objects in the stimulus scene; making a local scanpath on a large object not comparable to a global scanpath covering lots of small objects.

Small saccades, nevertheless, are generally assumed to be the result of detailed inspection of the stimulus, while longer saccades are eye movements *between* areas inspected—this is usually referred to as scanning. Thus the g/l ratio is synonymous with a comparison between scanning and detailled inspection, and has been used for:

Finding expert versus novice differences In a study of art perception, Zangemeister, Sherman, and Stark (1995) found that professional art viewers had a g/l ratio of 2.8, while non-professionals had a ratio of 1.3. Hence the authors conclude that the measure is capable of distinguishing between levels of cognitive skill.

Comparing imagery versus real stimulus perception The g/l ratio was found to be higher during stimulus viewing by Zangemeister and Liman (2007), who also found effects of *task* and *stimulus type*.

A similar definition was given by Krischer and Zangemeister (2007), who defined a 'global scanpath length' (sum of all long saccades) versus a 'local scanpath length' (sum of all small saccades), and found that the global scanpath length in picture viewing makes up around 95% of the total length.

10.7 AOI order and transition measures

AOI order and transition measures take the sequential order of movement into account. A 2D transition matrix is used exclusively for single transitions between two AOIs, and does not in any way represent what happened before or after those transitions. Each limited sequence in a transition matrix therefore, only has a 'history' of two dwells in two AOIs—this is its trace. Similarly, a first-order Markov model predicts the occurrence of a state (fixation, saccade, or AOI hit) on the basis of the preceding state, and does not include earlier states in the prediction. There are few measures that take longer intervals of a sequence into account, but the practical length of intervals seems to be around 2–10 dwells.

10.7.1 Order of first AOI entries

Target question	*Do participants look earlier at some AOIs than at others?*
Input representation	*AOI strings with first entries*
Output	*p-value for order effects*

The order of first AOI entries allows us to have significance tests for the difference in entry order between AOIs visited early and AOIs visited late. We form the same sort of table (such as Table 10.1) as in the calculation of average scanpaths (p. 282).

Table 10.1 Order of first entries for five AOIs seen by five participants (fictitious data). In this example, participant 2 first enters AOI 2, then AOI 1, then AOI 3 and so on.

Participant	AOI 1	AOI 2	AOI 3	AOI 4	AOI 5
1	1	3	2	4	5
2	2	1	3	5	4
3	1	2	3	4	5
4	3	1	2	5	4
5	1	3	2	4	5
Average	1.60	2.00	2.40	4.40	4.60

The average rankings, given in the bottom row of the table, indicate the average order in which the participants visited the five AOIs. On average, they visited AOI 1 first, AOI 2 second, and so on. Note that the amount of variation between the average rankings reflects whether the participants tended to visit the AOIs in the same order or not. If there had been no systematic ordering of the rankings across the five AOIs, the average rankings would be approximately the same. Since this is not the case in Table 10.1, the variation between the average rankings is relatively large.

The amount of variation between the average rankings is used to calculate W, Kendall's coefficient of concordance (Siegel & Castellan, 1988). If the AOI $i = 1,2,\ldots,n$ is first visited at order $r_{i,j}$ by participant $j = 1,2,\ldots,m$, then

$$W = \frac{12S}{m^2(n^3 - n)} \tag{10.14}$$

where

$$S = \sum_{i=1}^{n} (K_i - \bar{K})^2 \tag{10.15}$$

$$\bar{K} = \frac{1}{2} m(n+1) \tag{10.16}$$

$$K_i = \sum_{j=1}^{m} r_{i,j} \tag{10.17}$$

The coefficient W was originally used as a measure to determine whether a number of objects (here: AOIs) receive similar ratings (here: are visited in the same order) by a number of judges (here: participants). The value of W may vary between 0, indicating absence of agreement, and 1, indicating perfect agreement. In the example, W is equal to 0.784, suggesting that the AOIs were visited in more or less the same order by the participants.

If there is a high degree of agreement between the participants, then it is the case that one or more AOIs were visited before the other AOIs. The researcher may ask, then, whether the difference in average ranking between two AOIs is significant or not. A test that can be used for answering that question is Friedman's chi-square for related samples, which is the non-parametric alternative to a one-way within-subjects ANOVA. The resulting chi-square calculated on the data in Table 10.1 is 15.680, which, with four degrees of freedom, is significant. A procedure, Nemenyi's procedure, for doing pairwise comparisons is available (Heiman, 2006).

Since there is a direct relation between the amount of agreement and the differences between the average ratings, it is not surprising that the values of Kendall's W are also related

to these. Friedman's chi-square may be used to determine whether W is significant or not. Computationally, the relation between W and Friedman's chi-square is:

$$\chi^2 = n(m-1)W \tag{10.18}$$

in which n is equal to the number of AOIs, and m is equal to the number of participants.

In spite of its simplicity, this method has been surprisingly little used. One exception is Le, Raufaste, and Demonet (2003), who found that a participant with prosopagnosia (a neuropsychological disorder in which patients are unable to recognize faces) still looked at the parts (AOIs) of normal and scrambled faces in an order that could not be differentiated from the control group.

10.7.2 Transition matrix density

Target question	*Is the scanpath across AOIs directed or randomly distributed?*
Input representation	*A transition matrix*
Output	*A sparseness value*

Transition matrix density was defined by Goldberg and Kotval (1999) as the number of non-zero entries in the transition matrix divided by the total number of cells. It is better to divide the number of non-zero entries by the number of real cells in the matrix (see p. 195), in particular for dimensions (string lengths) higher than two.

A low transition density is an indication of efficient and directed search, while a dense matrix would imply random search, Goldberg and Kotval say. They apply gridded AOIs to two very different computer interface designs, and study the extent of search behaviour. The significant difference in the measure is argued to show that the two interfaces are differently searched. Note that this is true for two-dimensional transition matrices. For longer strings, a low transition density means that the participants have preferred paths through the AOIs of the stimulus and that many other possible paths are avoided.

10.7.3 Transition matrix entropy

Target question	*Is the scanpath across AOIs directed or randomly distributed?*
Input representation	*A transition matrix*
Output	*Entropy (bits)*

Entropy, originally defined by Shannon (1948), is a measure that calculates the uncertainty in a random variable, in our case a transition matrix. To explain entropy, we are going to run through an example illustrated in Figure 10.29 and Table 10.2.

Assume that we have a normalized transition matrix according to the figure. This is a fairly non-random transition matrix, since half of the transitions are of the type $B \rightarrow C$. The entropy H is then calculated as the negation of the sum of the last column, or in our example, 2 bits. More generally, entropy is defined as

$$H(R) = -\sum_{r_i \in R} p(r_i) \log_2 p(r_i), r_i > 0$$

where R is a normalized transition matrix and r_i are the cell values of that matrix with probabilities $p(r_i)$.

Entropy can be calculated for any dwell map or transition matrix. The lowest possible value is zero (0), which is only reached if there is just one cell in the matrix, i.e. when there

	A	B	C
A		0.0625	0.0625
B	0.0625		0.5
C	0.25	0.0625	

Fig. 10.29 Fictitious transition matrix with normalized values.

Table 10.2 Step-by-step entropy calculations for the transition matrix in Figure 10.29.

Transition	$p(r_i)$	$\log_2 p(r_i)$	$p(r_i)\log_2 p(r_i)$
$A \rightarrow B$	0.0625	-4.0	-0.25
$A \rightarrow C$	0.0625	-4.0	-0.25
$B \rightarrow A$	0.0625	-4.0	-0.25
$B \rightarrow C$	0.5000	-1.0	-0.50
$C \rightarrow A$	0.2500	-2.0	-0.50
$C \rightarrow B$	0.0625	-4.0	-0.25

is no uncertainty about what type of transition will occur. The maximum value for entropy is when all cells have the same value. In our example with six cells, the maximum value would be $-6\frac{1}{6}\log_2(\frac{1}{6})$, which equals 2.59 bits. 'Bits' is not a very intuitive unit, however, but by dividing the $H(R)$ value with the theoretically maximal value for the system (in our example 2.59), we arrive at a *normalized entropy* that allows for comparisons of results across groups and stimuli.

Calculating the entropy of a transition matrix, Shic, Chawarska, and Scassellati (2008) argued that a high resulting value is aligned with a preference for exploration, while low values indicate data with transitions mainly between a few of the AOIs. Jordan and Slater (2009) interpreted a drop in scanpath entropy over time as an indication that the virtual environment they tested "had cohered into a meaningful perception" (p. 185). They calculated entropy from transition matrices created from 1-second intervals, and collapsed data from all participants to estimate the total change in entropy over time.

10.7.4 Number and proportion of specific subscans

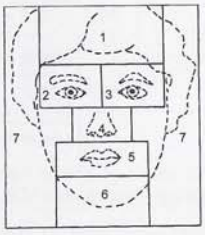
Target question	*How common is a specific subscan?*
Input representation	*AOI-processed data*
Output	*The number or proportion of each subscan, commonly determined in a histogram*

This measure counts the number of unique subscans in string representations of scanpaths. Groner *et al.* (1984) decided "in a compromise between theoretical considerations and statistical arguments" (p. 529) to analyse subscans with a length of three AOIs. As their face stimuli had seven AOIs (Figure 10.30(b)), they found the total number of possible subscans of length three to be 210 (see Equation 6.1 on page 195). Interestingly, the 20 most common subscans subsumed 57% of the total number of subscans. The two most common subscans moved between the eyes, but there was considerable individual variation (Figure 10.30(a)). Each subscan corresponds to a cell in a 3D transition matrix as the one on page 194.

TRIPLETS	SUBJECTS						
	1. RU	2. ME	3. BR	4. BK	5. WI	6. TR	Total
2-3-2	7	9	13	36	0	68	133
3-2-3	20	9	10	25	1	46	111
7-2-3	13	2	5	4	0	31	55
2-3-4	8	6	3	19	0	17	53
4-2-3	10	9	3	10	3	17	52
5-2-3	14	13	3	6	3	6	45
2-3-5	21	1	2	8	2	9	43
3-2-4	5	7	0	12	3	13	40
5-4-3	3	2	0	0	15	8	28
3-2-5	5	4	3	7	3	6	28
4-3-2	6	1	0	5	7	8	27
7-5-4	9	2	0	0	8	8	27
4-5-4	1	3	0	4	11	2	21
2-4-3	4	3	1	4	3	4	19

Table 1: The 14 most frequent triplets. Numbers indicate frequencies of observation. Under linings symbolize significant higher frequency than expected by independence (Chi-square test, $p < .01$)

(a) Subscan 'triplet' frequency table for six participants looking at faces.



1. Forehead
2. Left eye
3. Right eye
4. Nose
5. Mouth
6. Chin
7. Ears and sides

(b) Stimulus and the seven AOIs.

Fig. 10.30 Subscan frequency analysis. Reprinted from *Advances in Psychology*, Volume 22, Rudolf Groner, Franziska Walder, and Marina Groner, Looking at Faces: Local and Global Aspects of Scanpaths, pp. 523–533., Copyright (1984), with permission from Elsevier.

Again using subscans of length three, Koga and Groner (1989) compared non-native learners of Japanese Kanji sign sequencies before and after training using two different presentation modes. Based on the 20 most common subscans, they conclude that subscan frequency is not influenced by the presentation mode.

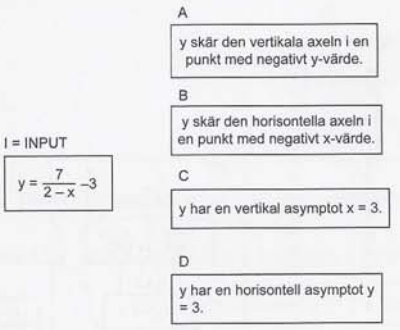
When analysing eye-movement data from participants picking up and dropping block objects, Ballard *et al.* (1992) found that the most common subscan sequence consisted of four AOIs. Tasks and participant strategies definitely influence the prevalence of longer subscans.

10.7.5 Unique AOIs

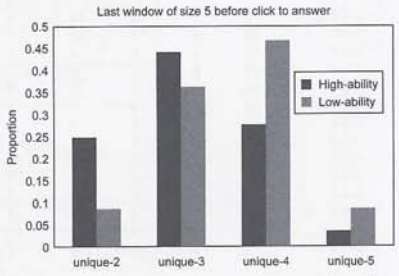
Target question	*Does the scanpath over AOIs represent a focused or an overview scan?*
Input representation	*AOI strings*
Output	*Proportion of scanpaths in each category*

The unique AOI measure counts how many unique AOIs there are in a substring, but is not concerned with the exact order of the AOI visits. For example, the strings `ABB` and `BBA` of length three both have two unique AOIs (A and B) but different AOI order. Since the unique AOI measure does not have to consider all possible subscans in a transition matrix, substantially fewer data need to be processed, and longer subscans become easier to investigate. However, the price is that sequential information within the subscan is completely disregarded. The measure resembles the local versus global and ambient versus focal of pages 265–267, but differs by using AOIs rather than amplitudes.

The unique AOI measure is calculated by letting a window of length ℓ slide over the recorded AOI sequence, and counting the number of unique AOIs in the window. Suppose for instance that for a hypothetical participant and trial, we record an AOI sequence `IAIBACIAIABCDIAIAI`. Repetitions are not allowed, and are collapsed to include only a single letter (as in compressed string edit representation). Unique-AOI sequences are then defined by letting a window of size 5 travel along the AOI sequence. First the window will



(a) Typical stimulus with AOIs.



(b) Histogram of uniqueness values, all participants and the last five AOIs before clicking to answer.

Fig. 10.31 Unique-AOI data. Sequence length $\ell = 5$, number of AOIs $n = 5$. High-ability participants clearly made more pairwise AOI inspections compared to low-ability participants.

encounter `IAIBA`, and see that there are *three unique AOIs* (i.e. I, A, B). Next, we move the window one step, and find `AIBAC`, which has four unique AOIs. We continue like this, until we have reached `IAIAI` at the end, which has only two unique AOIs. In total we will have 14 AOI-uniqueness numbers from the recorded sequence of 18 AOIs, namely 3–4–4–4–3–3–3–4–5–5–5–4–3–2. We now count how many unique-2, unique-3, unique-4 and unique-5 there are in this sequence and we find: 1 unique-2, 5 unique-3, 5 unique-4 and 3 unique-5.

Figure 10.31 shows a mathematical problem and the unique-AOIs values resulting from the last five AOIs looked at before participants clicked on the alternative they thought was correct (experiment described on p. 5). During this decision phase, high-ability students (defined as those solving a larger proportion of problems compared to the whole student group) made 25% pairwise comparisons—that is cases with five consecutive dwells, in which only two unique AOIs were visited—and another 43% unique-3 sequences. Low-ability students have a higher tendency to make overview scans.

When the uniqueness values calculated from each window are not collapsed over time, they can be used to study the relation between focus and overview looking over time. Calculations of the chance levels for each uniqueness group give chance probabilities of 0.0156 for unique-2, 0.328 for unique-3, 0.562 for unique-4, and 0.0938 that participants will sequence all five AOIs consecutively. Holmqvist, Andrà, *et al.* (2011) present calculations of chance probabilities for a whole variety of numbers of AOIs and window sizes, and describe how significance calculations should be carried out.

10.7.6 Statistical analysis of a transition matrix

Target question	*Are some transitions significantly more common than others?*
Input representation	*A transition matrix*
Output	*Significant transitions and p-values*

In this section, we illustrate the use of log-linear statistical analysis for the analysis of the transition matrix given in Table 6.1 on page 194. The principles behind log-linear analysis are briefly explained on page 90.

Table 10.3 Adjusted residuals for the data in Table 6.1 on page 194.

	LS	LF	RF	RS	I	E	O
Left Side (LS)	.	**8.58**	-1.93	**−2.04**	-1.41	-1.90	**−4.45**
Left Front (LF)	**5.50**	.	0.42	-1.93	**3.35**	-1.57	**−4.39**
Right Front (RF)	-1.34	1.67	.	**5.92**	-1.58	-1.44	-1.88
Right Side (RS)	-1.84	**−5.21**	**9.58**	.	**−2.07**	**5.90**	1.65
Instruments (I)	-0.60	0.58	**−2.53**	-1.63	.	-0.28	1.74
Engine (E)	-1.84	**−4.89**	**−2.82**	-0.29	0.85	.	**8.21**
Other (O)	-1.06	1.57	**−2.49**	0.31	-0.22	-0.25	.

In a two-dimensional table, as in Table 6.1, the first step in the analysis is to exclude the two-way interaction from the model, and to calculate the expected number of transitions based on only the two main effects. Note that the calculation of expected number of transitions in tables that contain structural zeros is realized by means of an iterative procedure, and should be left to a computer. The resulting value of chi-square indicates whether or not the expected numbers of the more parsimonious model are still reasonably close to the observed frequencies. In the case of Table 6.1, they are clearly not, since the chi-square value of 473.411 with 29 degrees of freedom[28] is highly significant.

The fact that chi-square is significant indicates that the two factors are dependent on one another. Translated to transition matrices, this means that transitions between certain AOIs were either significantly more or less likely than expected.

The natural question that follows is, of course, which transition numbers in the matrix deviate significantly from expected. Cells for which the expected number deviates significantly from the observed number may be identified through examination of the adjusted residuals, given in Table 10.3. Adjusted residuals are indications of the distance between an observed and an expected value, similar to standard scores. Positive values indicate that the observed number of transitions is higher than the expected number, whereas negative values indicate that the observed number is lower than the expected number. Absolute values that are larger than 1.96 are significant with $p < 0.05$. These are marked in bold in Table 10.3. The table shows, for instance, that the number of transitions from left side to left front was significantly higher than expected, whereas the transitions from left side to other were significantly fewer than expected.

If you have an a-priori interest in one or more transitions, the analysis of adjusted residuals may be pursued further (Bakeman & Gottman, 1997). Suppose that the transitions from left side to left front in cell (1,2) were of particular interest for the study. The effect of these transitions on the overall chi-square can be evaluated by declaring cell (1,2) structurally zero, and then recalculating the expected numbers. In the above example, the exclusion of this cell results in a new value of chi-square of 395.666 with 28 degrees of freedom. This is a reduction compared to the earlier value, but obviously, the new value is still highly significant.

In the case where transitions are being examined within the context of an experimental design, the situation becomes more complicated. A direct comparison of transition numbers across groups is not allowed. Instead, however, the researcher may choose to use a measure of the strength of association within the table as the dependent variable within the design. One such measure that is suitable for the analysis is the log odds ratio (Bakeman & Gottman, 1997). A restriction to this measure is that it is based on a two-by-two table. Therefore, the researcher either needs to incorporate this in the design of the experiment, or collapse across

[28](Number of columns–1) * (Number of rows–1)–Number of structural zeros = Degrees of freedom for a two-dimensional transition matrix.

multiple rows and columns in a larger table. If the cells of a 2×2 table are designated as F_{11}, F_{12}, F_{21}, and F_{22} respectively, the log odds ratio can be calculated as,

$$\ln\left(\frac{F_{11}/F_{12}}{F_{21}/F_{22}}\right) \tag{10.19}$$

Log odds ratios may be calculated for the individual items and participants in the experiment, and the resulting values are used as the input to a statistical test.

Two other methods have been used for the analysis of transition matrices: Markov modelling and correspondence analysis. Markov models were introduced on page 196. This method has been used within eye-tracking research by Liu (1998) and others. Correspondence analysis has not been applied often for the analysis of eye-tracking data. An exception is found in Loslever, Popieul, and Simon (2007). Correspondence analysis is an exploratory technique that may be used to see which rows or columns in the table are similar. Usually, the degree of similarity is visualized in a so-called biplot that shows which rows or columns are at close distance within a two-dimensional space. For an introduction to correspondence analysis, we refer the reader to Greenacre (2007).

10.8 Scanpath comparison measures

Scanpath comparison measures are used to estimate the similarity between two or more scanpaths. They draw on the methods for scanpath representation, simplification, and sequence alignment that we introduced in Chapter 8. In this section only measures that take into account the *ordinal* aspect of scanpaths are listed, and therefore position-based measures like attention map difference, Kullback-Leibler distance, and Mannan distance can be found in Chapter 11, devoted explicitly to gaze position.

Many of the measures described here perform mainly pairwise comparisons, but there are measures that allow for groupwise comparisons as well. For example Feusner and Lukoff (2008) describe a method for calculating statistical significance between groups of scanpath comparisons, even if the basic comparison measure is only pairwise. Moreover, several groupwise similarity measures have been devised using attention map sequences.

10.8.1 Correlation between sequences

Target question	*Do participants visit the AOIs of your stimulus in the same way, or in the particular order that you predict?*
Input representation	*AOI strings*
Output	*Correlation value [-1,1]*

The correlation between sequences takes two numerical representations of AOI strings and correlates them. The simplest is to correlate the strings of two participants or conditions. An alternative is to have a predicted order and correlate each participant's string to the string representing the predicted order.

Table 10.4 shows fictitious data from a participant looking at two different designs, one after the other, each having nine AOIs (named 1 to 9). Data in the table only include the *first* entries into an AOI. Calculating the correlation between the observed and predicted order of the AOI entries then gives us an estimate of the accuracy of the prediction. In the example, Design 1 has a correlation value of 0.95 and Design 2 the value -0.017. This indicates that Design 1 better triggered the predicted eye-movement behaviour from this participant.

Table 10.4 Order of predicted AOIs above, and below the fictitious AOI sequences of one participant looking at two different designs.

Predicted	1	2	3	4	5	6	7	8	9
Design 1	2	1	3	4	6	5	8	7	9
Design 2	8	7	2	3	6	1	5	4	9

Suppes (1990) used this measure to compare a normative model of arithmetic problem solving against eye-tracking data, finding average correlation values ranging from 0.583 to 0.874 for the different participant groups. Holsanova *et al.* (2008) used the measure to compare a serial versus a radial design of information graphics, finding an average correlation of 0.95 between predicted and actual reading behaviour for the serial information graphics, and no correlation for the radial design.

10.8.2 Attention map sequence similarity

Target question	*Do individual participants, or groups of participants, sequence the AOIs of your stimulus in a particular order?*
Input representation	*Raw data samples*
Output	*Similarity value*

Attention maps are typically static entities. However, attention maps can evolve over time and become *map sequences* (Wooding, 2002a). Such sequences can be seen as a collection of scanpaths from one or many participants that preserve ordinal information. As such, an attention map sequence is a three-dimensional scanpath function $f(S)$, as defined on page 253.

If we have a still image as a stimulus, we can generate different attention maps, each for a separate time interval along the trial duration. The resulting map sequence gives insight into how the attention map evolves over time, information which cannot be extracted from the final attention map collapsed over the entire trial. In fact, initial attention and overall attention can differ significantly in distribution, as Figure 10.32 shows (from Nyström, 2008).

Since map sequences can be generated from single and multiple scanpaths, both pairwise and groupwise scanpath comparisons are possible. Nyström *et al.* (2004) used the harmonic Kullback-Leibler distance (p. 376) to compare map sequences generated from two videos: an original video and a foveated, compressed version of it. Comparing the eye movements predicted by their well-known computational model of saliency to the real eye movements of human observers, Itti (2004) devised a metric that evaluated each recorded data sample on the saliency map, and then calculated the resulting average value over all data samples and frames. Itti found the saliency model to predict gaze locations better than chance. Grindinger, Duchowski, and Sawyer (2010) proposed a similar groupwise measure where fixation samples from all viewers in one group were compared to an attention map built from all viewers in another group. To validate their approach, a classifier was used to test whether experts' scanpaths were more similar to other experts' scanpaths than they were to novices', and vice versa. Grindinger *et al.* (2010) used still images as stimuli, and reported that their method achieved higher classification-accuracy than did the string edit distance (adapted for groupwise comparison).

Given two map sequences, any of the attention map based measures in Chapter 11 can be used to compare the maps at a certain time instant (or ordinal position). Then all similarity values can be added to find the overall scanpath similarity. This approach can be generalized to all point-based measures that compare two or more sets of data samples over time.

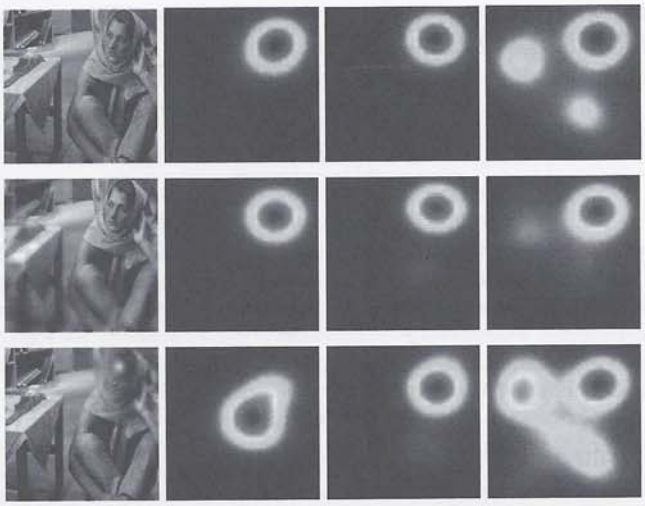
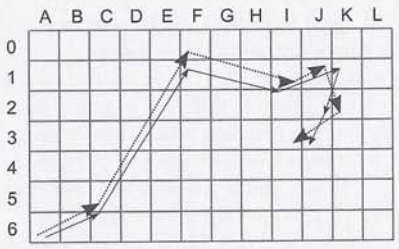
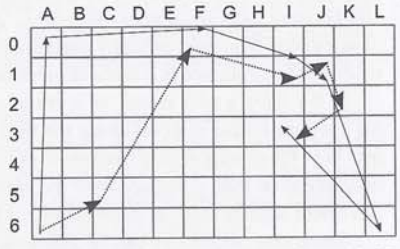


Fig. 10.32 The heat maps visualize how recorded data sample positions from seven viewers are distributed over different time intervals when viewing three different versions of an image. The second column shows where attention is located between 300–350 ms, the third column where attention is located between 600–650 ms, and the fourth column shows the overall attention. The photo is reproduced here with permission from the Signal Compression Lab, University of California Santa Barbara.



(a) Scanpaths represented by strings `A6 C5 F0 I1 J1 K2 I3` and `A6 A0 F0 I1 J1 L6 I3` have a string edit distance of 2.



(b) Scanpaths represented by strings `A6 C5 F0 I1 J1 K2 I3` and `A6 C6 F1 I2 K1 J2 J3` have a string edit distance of 6.

Fig. 10.33 Examples showing a major weakness of the string edit measure: the sensitivity to AOI borders.

10.8.3 The string edit distance

Target question	*What is the AOI sequence similarity between two scanpaths?*
Input representation	*AOI strings of dwells*
Output	*String edit distance (string symbols)*

Comparing two scanpaths and giving a distance value for them, the string edit measure (known as the 'Levenshtein distance' in computer science after originator Levenshtein, 1966) assumes that both scanpaths are represented with an AOI string of dwells. Both gridded and semantic AOIs are being used with the string edit distance measure (see p. 206 for the distinction between these AOI types).

Distance is calculated as the minimum number of *insert*, *delete*, and *substitute* operations needed to transform one string into the other. A smaller distance means that fewer transformations have to be made, and therefore the scanpaths should be more similar. For instance, if we have two scanpaths

$$S_1 : \texttt{A6 C5 F0 I1 J1 K2 I3}$$
$$S_2 : \texttt{A6 A0 F0 I1 J1 L6 I3}$$

we need to substitute A0 with C5 and L6 with K2. This gives us a total editing distance of two (2) for the comparison of S_1 against S_2. If instead we compare S_1 to a very different string S_3,

$$S_1 : \texttt{A6 C5 F0 I1 J1 K2 I3}$$
$$S_3 : \texttt{A6 C6 F1 I2 K1 J2 J3}$$

we need to substitute the last six AOIs in one string, giving an edit distance of six (6). This pair of scanpaths is thus three times more dissimilar than the first pair.

In order to compare two scanpaths of lengths m and n, the string edit distance d is often normalized against the maximum string length, which equals the largest possible edit distance.

$$\hat{d} = 1 - \frac{d}{\max(m,n)} \tag{10.20}$$

Thus, $\hat{d}$ varies from 0 to 1, where 1 signifies two identical strings. In our example, we would have a $\hat{d}$ of 0.71 for the more similar pair, and 0.14 for the less similar pair.

The string edit measure is undoubtedly the most employed measure for pairwise scanpath similarity. In the literature, it has been used to study:

Scene perception versus subsequent imagery For instance, Brandt and Stark (1997) and Gbadamosi and Zangemeister (2001) both compare imagery theories.

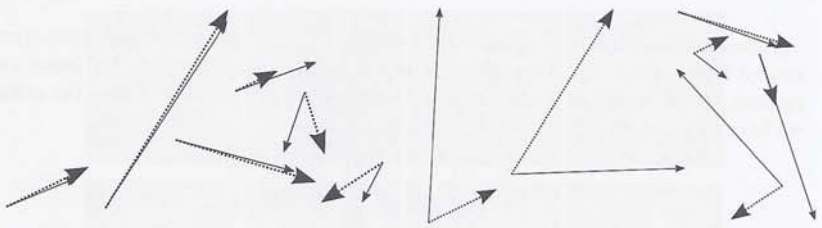
First viewing versus a second viewing Holm (2007) studied the role of expectations on the perception of an upcoming stimulus.

Website scanning Josephson and Holmes (2002) tested the scanpath theory with repeated viewings of the same web content. Also see Pan *et al.* (2004) who investigated behaviour when viewing webpages using the string edit method.

The validity of different theories about scanpaths Privitera and Stark (2000) and Foulsham and Underwood (2008) compared the fixation sequences of human viewers of a scene versus the scanpaths predicted by saliency-based models of eye movements. Hacisalihzade, Stark, and Allen (2002) investigated the relative role of deterministic and probabilistic influences on scanpaths, and the originators of this research, Choi *et al.* (1995), directly questioned the "usefulness of using string editing algorithms as a tool to quantify the similarity of fixation sequences for human participants searching a quasi-natural stereoscopic three-dimensional environment".

Moreover, the string edit algorithm is implemented in the publicly available software programmes of West *et al.* (2006) and Myers and Schoelles (2005).

Privitera and Stark (2000) denoted the edit distance 'sequential similarity' and complemented it with a 'locus similarity' measuring the proportion of letters in one string that were present in the other string, regardless of order. To represent the large number of similarity scores within and between participants and stimuli, they proposed what they call parsing diagrams, where only the average score for each combination was given. Comparing computer generated and human fixations, Privitera and Stark (2000) conclude that the algorithms they test do a poor job at predicting human fixation sequences.



(a) Pairs of vectors from scanpath (a) in Figure 10.33.

(b) Pairs of vectors from scanpath (b) in Figure 10.33.

Fig. 10.34 The length of the differential vector in each pair equals the distance between the two arrow tips.

String editing is not limited to data representations based on two-dimensional AOIs, but can be extended to handle vectors. The *vector string* is calculated based on the amplitude and direction of the scanpath saccades, as described on page 269. In short, each vector is represented by two values: one for amplitude and one for direction. Using the same example as in Figure 10.33, the three vector strings will be:

- Left side thin scanpath: `23 1B 46 34 83 92`
- Dashed thick scanpath: `24 1B 47 22 83 A3`
- Right side thin scanpath: `0D 3A 46 52 7A E9`

Since the example uses hexadecimal values, only 16 possible amplitudes/directions can be represented.

As in standard AOI-string editing, two strings are compared using editing operations, and the result is an editing distance. Since the strings represent real vectors, Gbadamosi (2000) added geometrical costs to the comparisons:

- *Insertion* or *deletion* equalled the amplitude of the inserted or deleted vector.
- *Substitution of vector u by vector v* added a cost equal to the length of the differential vector $||u - v||$.

In our example, there are no insertions or deletions but only substitution in the minimal edit distance, so the cost consists of a sum of the differential vectors. Figure 10.34 shows the pairs of vectors. For each pair, the differential vector equals the Euclidean distance between the tips of the arrows. For the pairs in (a), the tip distances are 7.95, 5.83, 12.14, 20.37, 20.6, 21.09, which gives a total editing cost (i.e. similarity) of 87.98. In the (b) pairs, the tip distances are 116.05, 106.16, 11.09, 26.08, 74.10, 97.33, with a total editing cost of 430.81.

Gbadamosi (2000) suggests normalizing all distances (d) against the maximum $d_{\max}$ (in our case 430.81) to yield a value span of between -1 and 1, giving the normalized edit cost

$$\hat{d} = 1 - 2\frac{d}{d_{\max}} \tag{10.21}$$

Vector string editing handles scanpath *shape* much better than the classical AOI string editing. The major weakness of the vector string editing method is that it completely ignores spatial position. For example, two scanpaths which produce the same vectors—or angles—when connecting fixations, will give identical similarity scores when using vector string editing, even if the overall scanpath is focused in a different area, if it is spatially scaled, or if the amplitudes between saccades in the sequence are completely different (see Figure 10.35). Vector string editing per se therefore needs to be complemented with a position comparison method.

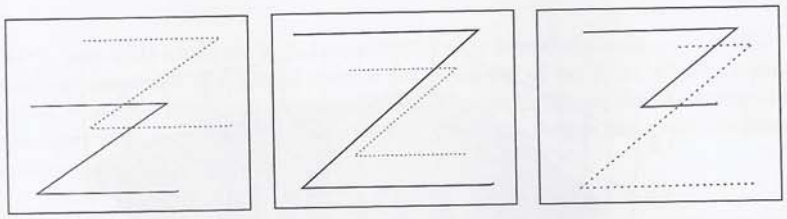


Fig. 10.35 Pairs of scanpaths where the first fixation in each begins in the top left. Vector-based string editing omits any comparison of position: each scanpath (thick line) and its pair (dashed line) are identical in terms of vectors, but completely different in terms of overall spatial position (left panel), scaling (middle panel), and amplitude between fixations (right panel).

Design issues with the string edit measure

Originally, the string edit distance algorithm by Levenshtein (1966) was developed to be a string comparison method that is extensively used on one-dimensional strings in computer communication theory and genetic research. The method was imported into the eye-tracking world around 1990 by Lawrence Stark and collegues (compare Choi *et al.*, 1995). Was the import successful? Does the string edit measure correspond to our subjective feeling of similarity? Figure 10.33 shows the scanpaths represented by the strings we used earlier. As we can see, the pair of scanpaths found to have distance 2 appear much less similar than the pair with distance 6.

It is easy to create other examples showing non-intuitive results of this measure. One fundamental weakness of the measure is that it was originally designed for single-dimensional strings and not for a two-dimensional space with its built-in Euclidean distance. Therefore two AOIs at a far distance are considered equally similar to two AOIs in immediate proximity. That is the reason why we constructed the example in Figure 10.33(a) with the A0 and L6 fixations in it. In Figure 10.33(b), we placed fixations close to each other, but on either side of a border, to make two very similar scanpaths appear very dissimilar in the string edit measure. In Figure 10.33(a), we placed fixations as far away as possible from each other, yet within the same AOI, to make two dissimilar scanpaths appear similar to the measure.

The weaknesses illustrated in Figure 10.33 are not uncommon, as the probability for two fixations to be located on either side of a border is not very low. Noise and poor precision increase the danger that two fixations that should be at the same position are in different AOIs. However, over many scanpath comparisons, the effect can be expected to be somewhat milder. Also, some users of the string edit measure, for instance Foulsham (2008), ran the same similarity calculations over several different grid sizes (Figure 10.36), to reduce the arbitrariness of the division of space into gridded AOIs.

If semantic AOIs with very varying sizes are used, the intuitiveness of the string edit measure is even lower. Figure 10.37(a) shows two widely different scanpaths that would be classified as identical, even though this is clearly not the case.

Apart from the size of AOIs, the semantic content of the AOI matters. If each AOI is semantically homogeneous, as in Figure 10.37(b)—where every AOI has only one semantic object and it does not matter where we look—then semantic AOIs give more intuitive results with the string edit measure, than if AOIs contain many different semantic elements. Also, the distance between the AOIs should not be of the same kind as the distance within the AOIs. In Figure 10.37(b), there is no meaningful semantic or spatial distance in the stimulus between the bottle and the glasses that can be expressed in terms of the height of the bottle or the diameter of the glasses. The lack of such between-AOI distances eliminates the possibility that AOIs which seem distant to us are treated by the string edit measure as equally close

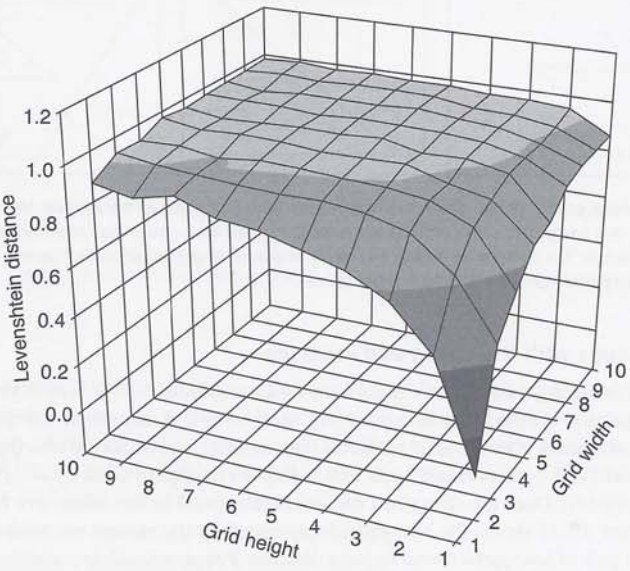
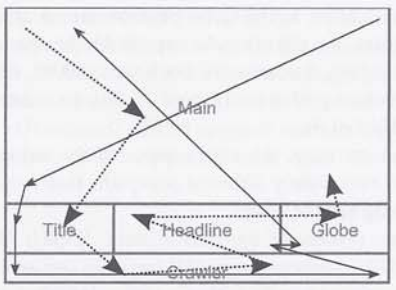
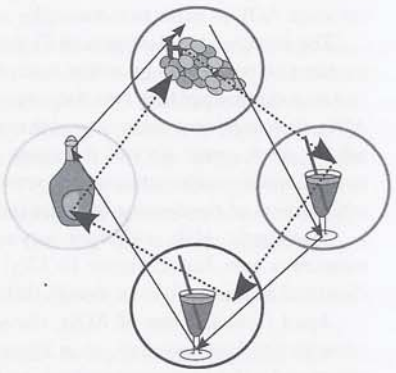


Fig. 10.36 The normalized Levenshtein string-edit distance for two random scanpaths as a function of size of the gridded AOIs. Grid dimensions vary from 1 x 1 (the stimulus is a single region where all fixations are evaluated as equal) to 10 x 10 (100 regions). The distance expected by chance increases as a finer grid is used. With kind permission from Springer Science+Business Media: HCI and usability for education and work, Knowledge-Based Patterns of Remembering: Eye Movement Scanpaths Reflect Domain Experience, *Lecture Notes in Computer Science* 5298, 2008, Andreas Holzinger, Figure 3.



(a) Semantic AOIs and two scanpaths (big arrowheads and broken line, versus small arrowheads and full line) that are considered identical by the string edit measure.



(b) The string edit measure best suits clearly delimited AOIs with a good distance between them, a generous margin and a single semantic unit inside each AOI.

Fig. 10.37 Poorer and better AOIs when using the string edit measure for scanpath similarity.

together as two AOIs which are in fact separated by only a small distance.

The third important AOI design issue is the proximity of the AOIs. Ideally, when the string edit measure is employed, AOIs should be clearly spatially separated, and have large enough safety margins, as in Figure 10.37(b), to reduce the danger that small variations in gaze position cause large differences in string edit distance.

10.8.4 Refined AOI sequence alignment measures

Target question	*How similarly have two scanpaths sequenced the AOIs?*
Input representation	*AOI strings*
Output	*Similarity score*

The limitations of the string edit algorithm call for more flexibility in adapting semantic and spatial relationships between AOIs, as well as cost parameters to match the specific question under investigation. To some extent the string edit algorithm can be extended by assigning different costs to operations. For example, Hacisalihzade *et al.* (2002) empirically found that 1 for substitution, 2 for insertion, and 3 for deletion were relevant costs with regard to their experimental questions.

In recent years, more advanced implementations of sequence alignment algorithms have been used for scanpath analysis; in particular the Needleman-Wunsch algorithm (Needleman & Wunsch, 1970) and the Clustal family of algorithms (Chenna *et al.*, 2003). These algorithms use a comparison matrix combined with gap penalties to dynamically find the optimal alignment between two sequences (p. 274).

The Needleman-Wunsch algorithm is at the heart of the publicly available software packages ScanMatch (Cristino *et al.*, 2010) and eyePatterns (West *et al.*, 2006), and has been used to compare scanpaths when studying decision strategies (Day, 2010). Two implementations of the Clustal software, ClustalG (Wilson, Harvey, & Thompson, 1999) and ClustalW (Thompson, Higgins, & Gibson, 1994), have been used by Fabrikant, Rebich-Hespanha, Andrienko, Andrienko, and Montello (2008) "to systematically compare and summarise individual inference making histories collected through eye-movement analysis", and by Turano, Geruschat, and Baker (2002) and Turano, Geruschat, and Baker (2003) to quantify scanpath similarity during walking.

'ScanMatch' by Cristino *et al.* (2010) compares AOI strings in which duration has been taken into account by repeating letters in proportion to their fixation durations. It also allows AOIs to be labelled with two letters, allowing more AOIs than the 26 of the English alphabet. AOI strings are aligned with the Needleman-Wunsch algorithm, which appears earlier to have been used for scanpath alignment by West *et al.* (2006) in their 'eyePattern' software. Given a comparison matrix M and a gap penalty, the ScanMatch algorithm computes a similarity score (d) by summing values along the optimal path through the matrix. Each element in the comparison matrix represents a relationship between two AOIs, for example how far apart they are in space, or their semantic relationship.

To normalize for sequence length (ℓ_S), the final similarity score is computed as

$$\hat{d} = \frac{d}{\max(M)\max(\ell_{S_1}, \ell_{S_2})} \tag{10.22}$$

giving the best possible match the value 1.

Evaluating the algorithm in three experiments, Cristino *et al.* (2010) found ScanMatch to be more robust than the string edit distance in correctly classifying similar scanpaths, as well as identifying scanpaths recorded during a specific task.

Also using the Needleman-Wunsch algorithm, Day (2010) offers a slightly different way to compute similarity by dividing the number of identical letters N_I in the aligned sequences by the sequence length ℓ_S

$$d' = \frac{N_I}{\ell_S} \tag{10.23}$$

10.8.5 Vector sequence alignment

Target question	*How similarly have two scanpaths sequenced through space?*
Input representation	*Saccade vectors, fixations, and input parameters*
Output	*A suite of similarity scores* $[0, 1]$

Regardless of the method used to calculate similarity between AOI strings, string edit principles are inevitably limited by the fact that stimulus space is carved up, which only crudely represents the fullness of scanpaths. An alternative to AOI strings is to represent scanpaths with Euclidean vectors.

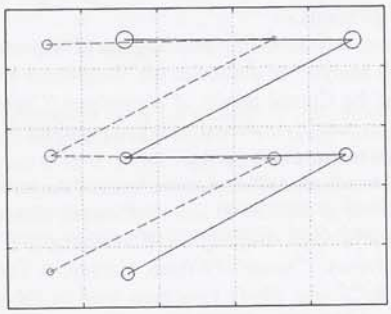


Fig. 10.38 Two example scanpaths where one is spatially shifted in relation to the other. Circle diameter represents fixation duration.

A scanpath can be seen as a collection of vectors, and hence comprise a vector space. This way of representing scanpaths was adopted by Jarodzka, Nyström, and Holmqvist (2010). Prior to comparison, the authors simplify scanpaths using thresholds for saccade direction and amplitude such that subsequent saccade vectors that go in a similar direction or have very short amplitudes are merged into larger vectors.

On the simplified representations, they propose the following sequential steps to compare two scanpaths:

1. Create a comparison matrix where each value corresponds to the similarity between two vectors in the scanpaths.
2. Create a graph (as in graph theory) from a set of rules defining how the matrix elements are connected. Assign each node in the graph with a weight proportional to the similarity value in the comparison matrix.
3. Calculate the shortest path from the top left element in the matrix to the bottom right element using Dijkstra's algorithm.
4. Align scanpaths along the shortest path such that each vector in one scanpath is matched with a vector in the other scanpath.

Then similarity is calculated on the aligned scanpaths with respect to five different aspects of the scanpaths: shape (length of vector difference), difference in amplitude between vectors, distance between fixation positions associated with saccades (equals the starting position for the saccade vector), difference in direction between vectors, difference in duration between fixations. Each measure is normalized with its largest possible value, to obtain values on the interval $[0, 1]$; 0 represents the best possible match and 1 the opposite.

Calculating similarity in several ways, the vector based algorithm has the potential to capture aspects of similarity that would otherwise be disregarded or hidden in the overall score. Figure 10.38 shows an example of two otherwise identical scanpaths, where one has been shifted in relation to the other. Given a 5×5 gridded AOI division of space as indicated in the figure, the string edit distance is 1 since no fixations do share the same AOI. However, calculating similarity according to the vector-based algorithm gives no difference in amplitude, direction, and shape, but clearly points out the difference in fixation position (0.15) and duration (0.32) (Jarodzka, Nyström, & Holmqvist, 2010).