

11 Position Measures

The previous chapter presented measures of the movements of the eye. In this chapter, focus is on measures of stillness of gaze in one or many positions. The position measures pertain to where participants look, if they look at the same place, and to the properties, such as duration of fixations and dwells at that position. We differentiate between five groups of position measures.

Position measure group	Uses	Page
Basic position measures	<i>Where did the participant look?</i>	357
Position dispersion measures	<i>How focused versus distributed is the gaze data?</i>	359
Position similarity measures	<i>How similar are the positions of two groups of gaze data?</i>	370
Position duration measures	<i>For how long did gaze stay in the position?</i>	376
Position dilation measures	<i>What is the pupil dilation at the position?</i>	391

We begin this chapter by spelling out the basic properties of position, that is, simply where a participant looks in terms of raw data samples, fixations, and dwells. These *basic position measures* report the very (x, y) - and area of interest (AOI)-values of data, and where in a specific AOI the eye lands.

Next we go on to define measures which describe variability in basic position, that is, measures of *position dispersion*. Position dispersion measures calculate how focused versus distributed a collection of position data (which we will call \mathcal{A}) are. Note that the many dispersion measures are in fact not measures, but different mathematical definitions of the vague concept of dispersion. The reason that we list these definitions as measures is that the research community writes about them as measures, not as definitions. As a consequence, the target question and use of the measures are close to identical, and the measure descriptions are more mathematical, while less emphasis is placed on functional interpretations. Although they all define dispersion, each such 'measure' behaves differently, as the comparison on page 359 demonstrates.

From dispersion, questions arise as to how comparable different position data are, and therefore measures of *position similarity* come next. Position similarity measures compare one collection \mathcal{A} of positions—say the fixations of one group of people—and give a value describing how similar the positions in \mathcal{A} are to the positions in the collection \mathcal{B} . Again, these measures are rather definitions which attempt to capture the vague concept of position similarity. Also note that these measures only compare the similarity in position, not taking sequence information into account. A particular statistical quirk with the similarity measures is that they output one single value for the comparison.

In the next two sections of this chapter we consider the *properties* of position data. Measures of *position duration* focus on the temporal characteristics of eye movement events at specific positions in space. Fixation duration and dwell time are the foundation for all

the other position duration measures (with the exception of the inter-microsaccadic interval, which we will come to).

Pupil dilation is another property of position data, giving the pupil size for the current position of the eye. These properties of position data are very important because they can reflect information processing with respect to the location a participant is looking.

As a whole, the position measures have very different units, from (x,y) -coordinates and percentages of AOI widths, to whole attention maps, and in the case of the Kullback-Leibler measure, a value without a unit. Consequently, statistical methods vary between the position measures. Note that many of the ratio measures in this chapter have values that are restricted by 0, 1, or by trial durations, and that this may give invalid outcomes from some statistical tests, for instance if used as outcome variables in a regression analysis, or in some cases, when performing ANOVAs on dwell times recorded with fixed trial durations.

There is much research on factors that makes us look at positions, which—if not taken into account when designing an experiment—may turn up as confounds, but which can also be actively manipulated in your experimental design. Therefore the final section of this chapter summarizes *potentially confounding factors* which may influence position data in unwanted or unanticipated ways (p. 394).

11.1 Basic position measures

The basic position measures address questions such as where a participant looks, and what areas of interest (AOIs) are not looked at.

11.1.1 Position

Target question	<i>Where did the participant look?</i>
Input representation	<i>Raw data, fixation data, or AOI-based data</i>
Output	<i>The position (pixels) or an AOI name (symbol)</i>

In eye-tracking data, *position* is given as Cartesian (x,y) -coordinates in a two (or three)-dimensional space; this is either the stimulus for remote and tower-mounted eye-trackers, or the scene video recording for head-mounted recordings. Typically, origo is located at the top-left corner (p. 61–64).

After we have run the raw position data through fixation analysis and related this to our AOIs, we are left with three types of data, each containing different position information:

Raw data samples These (x,y) -positions are the most reliable and detailed position data, their quality endangered only by low precision, inaccuracy, latencies, and real-time recording filters.

Fixations When transformed into events, the (x,y) -positions of the raw data samples are replaced by an average (x,y) -position for samples belonging to the fixation. Fixation position data are more commonly used than the raw data samples from which they are deduced. Fixation positions are additionally subject to the peculiarities of filters, the selected fixation detection algorithm and its settings.

Dwells The dwell event does not have (x,y) -positions. The value of a dwell position is its AOI, so dwells have a whole area as their position value. When AOIs are large, recording imprecision typically does not matter much, while inaccuracy can be problematic. As a consequence, the variable type is categorical.

When recording binocular data, position can be reported as (x, y, z) with z as a third dimension inferred from the relative distance between the eyes. Typically, z contains considerable noise since it is derived from a subtraction of two already noisy measurements. Therefore, it is used only rarely.

When recording data on gaze-overlaid scene videos, position is immediately visualized by the gaze cursor.

Quantifying basic eye movement position can be interesting to researchers for many specific reasons. For instance, Land *et al.* (1999) classified fixation position into different functions when participants completed the everyday task of making tea. Four categories could be identified: *locating*—where in space are the objects needed to complete the task, the kettle for instance; *directing*—just before contact the target direction must be relayed to the hand before the object is grasped; *guiding*—when several objects need to be manipulated in an action sequence, supervisory fixations facilitate this process; and *checking*—verifying whether the outcome is achieved or not, whether the kettle is full for instance. The researcher should keep in mind, however, that looking at a position does not necessarily mean full understanding of the information available there, as anyone who has tried to read a foreign language will know. In a study of squash players, Abernethy (1990, p. 63) conclude the same: “Not finding any differences between experts and novices provided further support for the conclusion that the limiting factor in the perceptual performance of the novices is not an inappropriate search strategy but rather an inability to make full use of the information available from fixated display features”.

11.1.2 Landing position in AOI

Target question	<i>How far into an AOI A does the fixation land?</i>
Input representation	<i>A fixation in the AOI A</i>
Output	<i>Percentage of horizontal extension of A, or the letter position (pixels) in the word in A</i>

Landing position in AOI is mostly used in reading research. There, it not only assumes a conventionalized AOI order (word order), but also that the reading direction is manifest inside the AOIs. Landing position is typically reported as number of characters and sometimes also as a percentage of the AOI size. Early studies in the 1970s and 80s showed that reader gaze does not land randomly on words, at least not in single sentence reading. There are two ‘positions’ in the words that have attracted the interest of reading researchers:

1. The *optimal viewing position*, located slightly to the left of the centre of the word. This is the position that gives the shortest fixation durations and naming times, and therefore presumably the most effortless lexical activation. Naming time increases on average with 20 ms per character offset from the optimal viewing position. Also, the larger the distance is between a reader’s landing position and the optimal viewing position, the more likely he is to make a refixation on the word (McConkie, Kerr, Reddix, Zola, & Jacobs, 1989).
2. The *preferred viewing location* is the position in a word where most readers land. Its average is located a bit further to the left from the optimal viewing position. McConkie, Kerr, Reddix, and Zola (1988) showed that the preferred viewing location in a word depends on the launching position in the previous word.

Vishwanath and Kowler (2004) used the measure outside of reading to investigate where saccades land inside AOIs outlining objects. They showed that saccades land either near the

centre of gravity in a three-dimensional (3D) object, or near the centre of gravity in its 2D projection on the retina.

11.2 Position dispersion measures

Dispersion ('variability') refers to the extent to which data are spread out or scattered. Dispersion measures appear in the literature in different contexts and under several different names. For example, 'saccadic extent' is a measure of the extension of scanpaths in the vertical and horizontal dimensions during a task. 'Fixation density' is probably the most common term when fixations are the events under investigation, although that term more often refers to the counting measure number of fixations in an AOI (e.g. Henderson & Hollingworth, 1999). When measuring the extent of raw data samples contributing to fixations, the term 'fixation dispersion' is often used. 'Distribution of gaze intensity', 'spread of search' and 'scanpath area' are other common terms for the same variability.

Position dispersion could be seen as a single, somewhat vague measure with many operational definitions. However, the very mathematics of these definitions were developed; they were termed 'measures', and we adhere to this terminology.

The common denominator for all the dispersion measures presented in this section is that they operate on *one* group of samples, either raw data samples or fixations, that do not take time or order into account, i.e. all samples are either recorded at the same time, or collapsed over time. These samples can originate from different sources such as:

1. Raw data samples or fixations from one or many participants looking at a stimulus. In this case, the dispersion in where people look derives from factors discussed previously in the book such as task, viewer idiosyncrasies, and type of stimulus. Typical research questions include how large the inter-participant fixation variability is, i.e. whether viewers look at similar positions, and thus whether their fixations are constrained to a limited part of the stimulus.
2. Raw data samples within a fixation from one or many participants looking at a stimulus. The variation within a fixation depends on factors such as the precision of the eye-tracker, the fixation stability of the viewer and, of course, which algorithm was used to define the fixation. Intra-fixational dispersion is used, for example, for clinical purposes to detect differences in fixation stability across different patient groups. Unlike data from the first type of source, data samples within a fixation are less prone to outliers, which have been removed during fixation detection.

11.2.1 Comparison of dispersion measures

Examples of three hypothetical spatial distributions of eye-tracking data are shown in Figure 11.1. In Figure 11.1(a), all the data samples are located very close to each other, and the dispersion is hence low. The opposite situation, giving a high dispersion, is illustrated in Figure 11.1(c) where data samples are spread out evenly over the display. However, in Figure 11.1(b) it is no longer transparent whether the dispersion is low or high. On the one hand, the data samples tend to cluster in two distinct groups, each with a low dispersion. If all samples are treated as a whole, on the other hand, the sum of distances between them is quite large. While most measures of dispersion give largely similar results along the extreme cases (very low or high variabilities), the results produced from the intermediate range depend heavily on which measure is used, as illustrated in Figure 11.1(d).

As shown in Figure 11.1, most of the measures yield similar and consistent results when the dispersion is very high or very low. When the scattering of data is somewhere in between

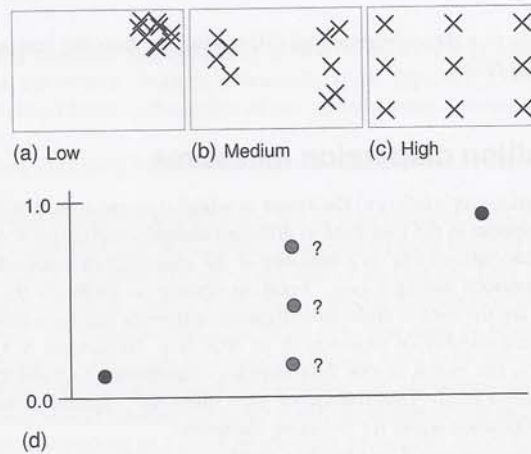


Fig. 11.1 For the position data in (a), dispersion is low, and for (c) it is high. But what dispersion value should the data in (b) give us?

low and high, however, less is known about how different measures behave. Here we provide a comparison of the different dispersion measures to give the reader an overview of how the definitions work (in particular with intermediate dispersion sources as input), before going on to describe the specifics of each measure of dispersion in full.

The properties of the dispersion measures which we will cover were investigated by implementing²⁹ each operationalization and comparing them using the data in Figure 11.2(b). The result can be seen in Figure 11.2(a), where values have been normalized to the interval $[0, 1]$ to better highlight differences across measures. Results show that some of the measures take more into account the fact that data with 'medium' dispersion forms two clusters, and therefore give lower dispersion values. Such examples are the nearest neighbour index (NNI), coverage, and average landing altitude. Conversely, standard deviation, range, and Kullback-Leibler distance (KLD) treat the medium- and high dispersion groups as being quite similar.

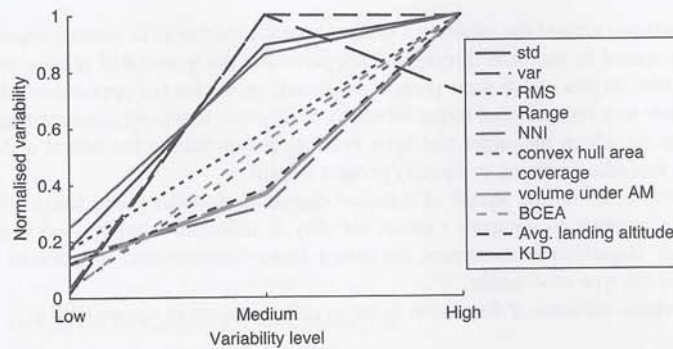
11.2.2 Standard deviation, variance, and RMS

Target question	<i>How much do the raw sample data or fixations in a set \mathcal{A} vary from the position mean or from sample-to-sample?</i>
Input representation	<i>A set \mathcal{A} of raw data samples, fixations, or angular distances θ between data samples</i>
Output	<i>Variance</i>

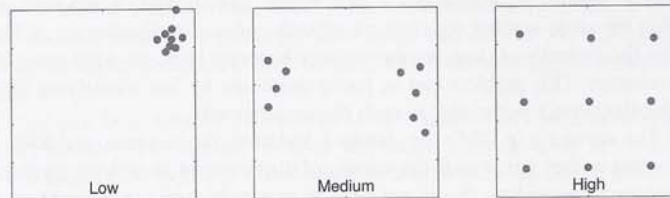
Standard deviation, variance, and root mean square are three statistical and mathematical ways of expressing variability in data. Standard deviation (SD) σ is defined as

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (11.1)$$

²⁹Our implementations, for practical reasons, differ slightly from how they are described in the text. All attention maps are generated with superimposed Gaussian functions using $\sigma = 0.10 \times$ stimulus width. For coverage, the attention map was cut to half the maximum height, and the BCEA was calculated for $k = 0.5$. A symmetric version of the KLD was used (Rajashekar *et al.*, 2004).



(a) Dispersion as calculated using the different operational definitions outlined in this chapter: Standard Deviation, Variance, Root Mean Square, Range, Nearest Neighbour Index, Convex Hull Area, Coverage, Volume under an Attention Map, Bivariate Contour Ellipse Area, Average Landing Altitude, Kullback-Leibler Distance. 'Low', 'medium' and 'high' corresponds to the synthetic data in (b), and a normalized variability closer to 0 indicates that the measure finds the positions to be less dispersed in space.



(b) Simple synthetic data sets representing fixation dispersions at two extremes (low and high), and the intermediate case where 'dispersion' is more subjective (medium). These data sets were input to the different dispersion measures, and position variability was calculated for purposes of comparison.

Fig. 11.2 The dispersion you come to report is very much influenced by the measure you chose. Simulated dispersion measures providing a heuristic for how your choice of dispersion measure will affect the dispersion output calculated. Values are normalized to the interval [0,1] to make relative comparisons easier. Note that the measures are in less agreement concerning low variability in position data, than high variability. This is a good index for the sensitivity of the chosen measure when, for instance, fixations are grouped in a spatially extended, yet local AOI on your stimulus display. It is logical that measures of dispersion in eye tracking are more sensitive to local differences in position since viewing is often restricted to monitors, and saccades are programmed to facilitate the acquisition of nearby visual information. Also note that KDL, one of the most widely used measures, finds dispersion to be lower in the 'High' case.

where \mathcal{A} contains N data samples $x_i, i \in [1, 2, \dots, N]$ and \bar{x} denotes the average of all x -values.

Di Russo, Pitzalis, and Spinelli (2003) examined fixation stability in a group of professional shooters and a control group. They calculated average standard deviations of eye movements during a fixation task, and found that the shooters were more stable in their fixations (i.e. had lower standard deviations) than the control group. Furthermore, the participants in the control group were less able to keep their eyes fixated over time, since their standard deviations tended to increase after 30 seconds of fixating the target. Similarly, Edelman and Goldberg (2001) used standard deviation as a measure of the spread of saccadic endpoints

(11.1)

est. All attention
or coverage, the
symmetric version

(i.e. fixations) around the mean, in a study where the discharge in primate superior colliculus was compared to saccadic direction. This particular study revealed greater neuron activity to saccadic targets which were present at saccade execution (as opposed to when a saccade was made to a remembered target location). Firing rate remained constant regardless of the duration for which the target had been present, demonstrating the neural underpinnings of greater saccadic precision to visually present stimuli.

Variance, σ^2 , is the square of standard deviation. Snodderly and Kurtz (1985) used the variance measure to compare fixation stability in macaque monkeys and humans, finding a greater dispersion in macaques, but also a larger between-trial variation in variance that reflected the type of stimulus.

A related measure of dispersion is the so called root mean square (RMS)

$$\theta_{\text{RMS}} = \sqrt{\frac{\theta_1^2 + \theta_2^2 + \dots + \theta_n^2}{n}} \quad (11.2)$$

where θ denotes the distance in degrees of visual angle. It is commonly used when calculating the precision of an eye-tracker (p. 34–41).

Due to their sensitivity to outliers, and their inability to recognize cluster formations in a data set, these three measures are typically used to estimate the dispersion in samples from a fixation. Consider, for example, the case where data clusters in two well-defined groups containing the same number of points on opposite sides of a stimulus (as in Figure 11.1(b)). Arguably, the dispersion is low, but the variance becomes large since the mean lies in between the two clusters. This problem can be partly overcome by first identifying the two clusters, and then calculating the variance in each cluster separately.

Just like skewness (p. 384), the standard deviation, the variance, and RMS are summary statistics, that is, they summarize the amount of dispersion of an underlying variable. As such, these measures themselves should not be used as variables in a statistical test. Nevertheless, there are specific tests that may be used to compare the variance in two or more groups. One such test is the Levene's test, which is often used to establish whether the assumption of homogeneity of variance, that is equal variance across groups, in an ANOVA is violated or not.

11.2.3 Range

Target question	<i>How large is the smallest box that covers the raw data samples, fixations or saccades in \mathcal{A}?</i>
Input representation	<i>A set of raw data samples, fixations or saccades \mathcal{A}</i>
Output	<i>Horizontal and vertical range /extension (pixels)</i>

Range (R), also known as 'extent' is calculated as the distance between points in the horizontal and vertical meridians (left-to-right, and up and down directions).

$$\begin{aligned} R_h &= \max(x) - \min(x) \\ R_v &= \max(y) - \min(y) \end{aligned} \quad (11.3)$$

Sometimes, but not always, the horizontal and vertical values are added to form an overall range $R = R_h + R_v$. Figure 11.3 illustrates the calculation. It can be seen that only four points (those that fall on the lines) are used in the calculation of range. This means that the distribution of other points within the boxed area does not affect the range, and it is therefore very

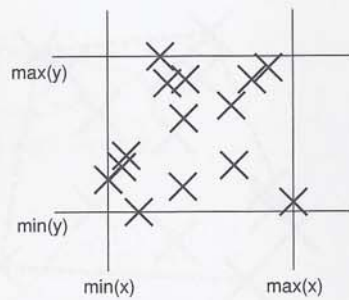


Fig. 11.3 Minimal box encapsulating all the fixations in \mathcal{A} .

sensitive to single outliers in the data. The maximum distance can be significantly larger if the maximal point becomes larger or the minimum point becomes smaller.

Range is a measure mostly used in human factors research, most specifically in the form of saccadic extent during car driving. Crundall and Underwood (1998) found that drivers have a saccadic extent varying from 38.7° to 82.4° in the vertical and 12.1° and 24.1° in the horizontal dimension. Saccadic extent is likely to be influenced by the *tunnel vision* that results from a heavier mental workload (see Godnig, 2003; Rantanen & Goldberg, 1999; Williams, 1988). For instance, when participants perform mental calculations and spatial imagery tasks while driving on highways, Recarte and Nunes (2000) observed significant decreases in saccadic extent.

Range has also been used to operationalize the variation in raw data samples during a fixation (this has also been termed 'fixation dispersion'), for instance in the I-DT fixation detection algorithm, to help define when a fixation begins and ends (see p. 154).

11.2.4 Nearest neighbour index

Target question	<i>To what degree are the points in \mathcal{A} randomly distributed?</i>
Input representation	<i>A set of fixations \mathcal{A}</i>
Output	<i>Distance (pixels)</i>

The *nearest neighbour index* measure is an operational definition of dispersion originating from Di Nocera, Terenzi, and Camilli (2006), who attempt to distinguish between 'randomness' and 'order' in a distribution of fixations. Its value is calculated in three steps. First, for each point $(x_i, y_i), i \in [1, 2, \dots, N]$ in \mathcal{A} , compute the Euclidean distance $d_{i,j}$ to the nearest neighbour among all the other points in \mathcal{A} . Then compute the average of all such distances $\frac{1}{N} \sum_{i=1}^N \min_j d_{i,j}$ and then divide that average by the expected value for random distributions $0.5\sqrt{\frac{A}{N}}$, where A is the area of the convex hull of the points. This gives a distance value D which can be calculated as

$$D = \frac{2}{\sqrt{AN}} \sum_{i=1}^N \min_j d_{i,j} \quad (11.4)$$

D approaches 1 when the distribution is more dispersed, whereas values smaller than 1 suggest more clustered distribution.

The originators of the measure calculate the NNI ratio for windows of duration 1 minute. The measure is used in human factors studies of mental workload. For instance, Camilli *et al.* (2008) showed that temporal attentional demand (items that need to be attended changing

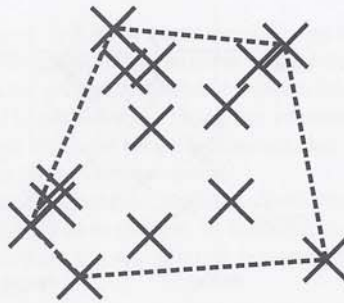


Fig. 11.4 The convex hull area is marked by a dotted line.

quickly in time) led to a more dispersed pattern, whereas visuospatial demand (the amount of information in space requiring attentional resources, irrespective of time) led to clustered pattern of fixations.

11.2.5 The convex hull area

Target question	<i>What is the minimal convex area that spans all points in \mathcal{A}?</i>
Input representation	<i>A set of data samples \mathcal{A}</i>
Output	<i>An area (pixels²)</i>

The convex hull area is an operational definition of dispersion as the smallest convex polygon containing \mathcal{A} , illustrated in Figure 11.4. There are several algorithms for calculating convex hull for any set of points, the best known of which may be the 'divide and conquer' algorithm by Preparata and Hong (1977).

The convex hull area was adapted to eye-tracking analysis by Goldberg and Kotval (1999) to allow for crude comparisons of the overall form of a scanpath. Goldberg and Kotval give examples showing that a convex hull is a better representation of scanpath shape than just circumscribing the collection of fixations with a circle. Nevertheless, like the related range measure, the convex hull may ignore large empty areas inside the hull, and is very sensitive to single outliers; both of these are aspects of distributions that attention maps and their Gaussian landscapes readily account for (see Figure 11.5).

Some researchers still use convex hulls to analyse the spatial distribution of their eye-movement data. For instance, Sullivan, Jovancevic, Hayhoe, and Sterns (2005) had a Star-gardt's disease patient, with bilateral central scotomata, perform three everyday activities: making a sandwich, catching a ball, and walking. Convex hulls were calculated from fixation data to visualize the extension of the preferred portion of visual field in each task. Ahlstrom and Friedman-Berg (2006) compare an air traffic controller weather interface with either a dynamic or a static representation of upcoming storms. Their data show that dynamic representations result in much smaller convex hull areas, and follow Goldberg and Kotval in interpreting this to mean that the static interface is poorly designed.

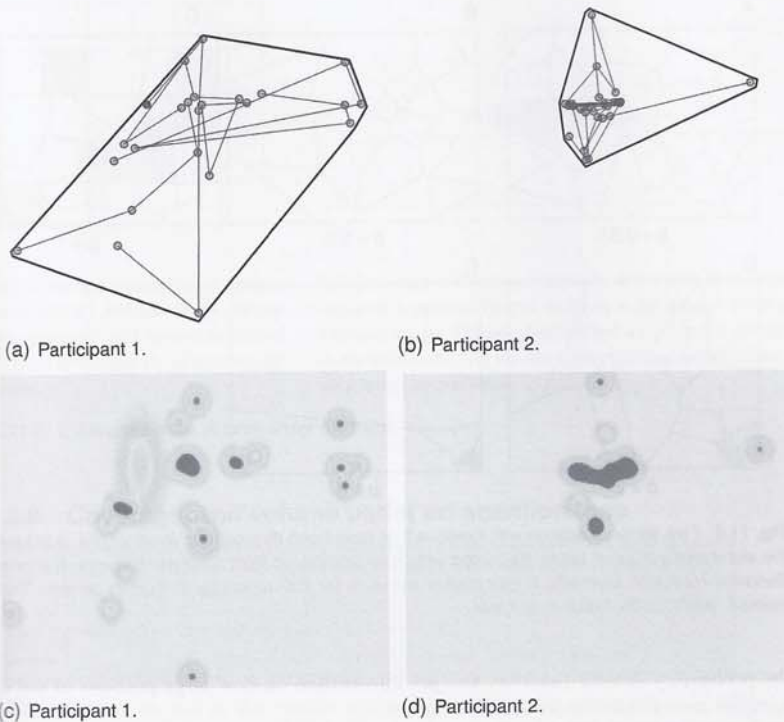


Fig. 11.5 Two participants viewing the Mona Lisa painting. Above, convex hulls around the scan path; below, heat maps for the same data.

11.2.6 Bivariate contour ellipse area (BCEA)

Target question	What is the dispersion measured as the area of an ellipse encompassing the majority of data samples in \mathcal{A} ?
Input representation	A set of raw data samples or fixations \mathcal{A}
Output	An area (pixels ²)

The bivariate contour ellipse area (BCEA) is calculated as the area of an ellipse that encompasses a given proportion $P = 1 - e^{-k}$ of points in a data set. It is calculated as

$$\text{BCEA} = 2k\pi\sigma_h\sigma_v(1 - \rho^2)^{1/2} \quad (11.5)$$

where σ_h and σ_v denote the horizontal and vertical standard deviation of the data samples, and ρ is the Pearson's correlation coefficient between the x and y samples. As shown above, k then decides the proportion of samples (in %) to include in the ellipse area.

Crossland, Sims, Galbraith, and Rubin (2004) criticize the BCEA on grounds of not being able to differentiate between one dispersed fixation and two fixations very close to one another. Instead they opt for using a mixture of Gaussian functions to represent raw data samples in a fixation, in essence producing miniature attention maps for the raw data sample distribution within the fixation. This allows the authors to find (sub-) fixations as maxima in

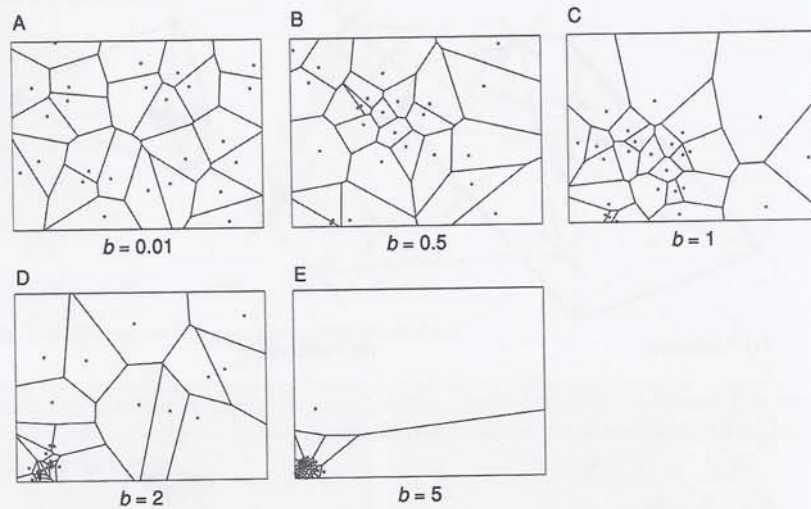


Fig. 11.6 Five stimulus spaces with fixations that have been divided into Voronoi cells. b -values indicate the skewness values of each. Reprinted with kind permission from Springer Science+Business Media: *Behavior Research Methods*, A quantitative measure for the uniformity of fixation density: The Voronoi method, 38(2), 2006, Eelco A. B. Over.

the probability density function, and get information on how large portions of the raw data samples are included under each maximum.

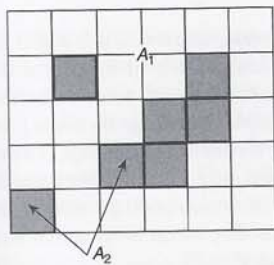
The BCEA has mainly been used to operationalize the fixation dispersion ('stability') concept for clinical applications (Crossland & Rubin, 2002; Tarita-Nistor, Gonzalez, Mandelcorn, Lillakas, & Steinbach, 2009), but also in more applied fields to quantify the inter-participant dispersion across viewers watching video (Brasel & Gips, 2008).

11.2.7 Skewness of the Voronoi cell distribution

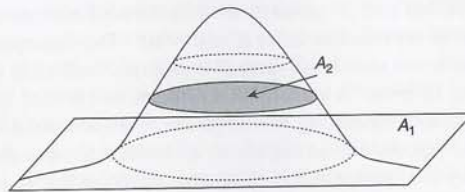
Target question	What is the dispersion in a very large number of fixations, \mathcal{A} ?
Input representation	A large set of fixations, \mathcal{A}
Output	A skewness value

The skewness of the distribution of Voronoi cell sizes was introduced by Over, Hooge, and Erkelens (2006) who consider it an operational definition of 'fixation density'. In this measure each fixation is given one cell, as Figure 11.6 exemplifies. When fixations are sparse, cells will be large, while crowded fixations give small cell sizes. When fixations are equally distributed, cell sizes will be the same, and have an unskewed distribution. Voronoi tessellation is a simple calculation that exists in several softwares.

Skewness is calculated with Equation (10.5) on page 316 with cell sizes as input. Over *et al.* fitted mathematical probability density functions to the observed cell-size distributions in order to estimate confidence intervals. Although the skewness value can be calculated also when \mathcal{A} contains few fixations, acquiring a confidence interval would require large numbers (at least 10,000) of fixations for high accuracy, Over *et al.* estimate.



(a) Coverage according to Cowen *et al.* (2002), defined as $\frac{A_2}{A_1}$, where A_2 spans all grid elements visited at least once, and A_1 all non-visited areas.



(b) Coverage in attention maps: A_2 is the area of the cutoff region at a selected height, while A_1 is the area of the entire stimulus image. Coverage is defined as $\frac{A_2}{A_1}$, but is sensitive to the slice level. The volume under the map summarizes all slice levels in one value.

Fig. 11.7 Coverage versus volume under the map.

11.2.8 Coverage, and volume under an attention map

Target question	How large is the area or volume covered by the attention map from \mathcal{A} ?
Input representation	An attention map made from \mathcal{A}
Output	An area (pixels ²), a volume (pixels ³), or a proportion

The coverage as well as the volume under an attention map are two closely related operational definitions of dispersion that deal with some of the drawbacks of extent and convex hull measures, but which have been too little used.

Coverage was proposed by Wooding (2002b) as a dispersion measure operating on attention maps to “measure the amount of the original stimulus covered by the fixations”. Given a height at which the attention map is ‘cut off’, Wooding defines the coverage as the area of the cutoff region (A_2) divided by the area of the whole image (A_1)

$$\text{Coverage} = \frac{A_2}{A_1} \quad (11.6)$$

In Figure 11.7(b), the cutoff height is set to half the maximum height and the cutoff area is represented by the shaded, grey region. However, a major problem is that the coverage value depends heavily on an appropriate selection of slicing/cutoff level.

The volume under the map measure overcomes the settings problem by extending the calculation to three dimensions: normalize the height of the attention map to unity (p. 248), and then calculate the *volume* under the attention map using summation or integration. In addition to making the slice level setting unnecessary, the volume approach gives a more sensitive measure, since areas at all different slicing levels are integrated into one single value.

An appropriate smoothing factor for the attention map (e.g. σ if using Gaussian based attention maps) still remains to be selected, and this choice will affect the coverage and volume under the map measures. See page 245 for suggestions of how to select σ . Parallel analyses with several σ s can be used to test whether the dispersion value is solely an effect of choosing the right σ , or if it reflects a systematic effect in your data.

Using a slightly different method to calculate coverage, shown in Figure 11.7(a), Cowen *et al.* (2002) first divided the stimulus into a grid, and then defined coverage as the number of

visited grid elements divided by the total number of elements. One dwell into a grid element was sufficient to count it as a 'visit'. They interpret the measure to be indicative of 'focused efficient searching' versus 'widespread inefficient search'. Cowen *et al.* describe their method as follows: "A screenshot was taken and loaded into a graphics package so that a grid, with squares measuring 30×30 pixels, could be overlaid on the screenshot of the page. The number of squares containing one or more fixations was manually counted. This number was divided by the total number of squares covering the page and then multiplied by 100 to produce a percentage" (p. 329). Similar methods to calculate the dispersion of fixations had been proposed already in the 1960s by Mackworth and Morandi (1967).

A related coverage measure is the *weighted search area* (WSA) proposed by Chi and Lin (1997), who argue that the WSA "can describe concentrations of visual search patterns and potentially quantify the amount of visual load (p. 255)" [of participants engaged in a simulated driving task]. The implementation is described in 10 detailed steps in the paper, and is based on the ideas to 1) divide the stimulus space into quadrants and 2) represent fixations in each quadrant with a vector taking both the position and duration of fixations into account.

Taking stimulus features into account, coverage has also been defined as the percentage of objects in a scene looked at for more than a minimum duration (Yoon & Narayanan, 2004a).

11.2.9 Relative entropy and the Kullback-Leibler Distance (KLD)

Target question	<i>How large is the dispersion in \mathcal{A}, when counted as the average distance between the attention map of each fixation in \mathcal{A} and the attention map of all the other fixations?</i>
Input representation	<i>A set of raw data samples or fixations \mathcal{A} per participant</i>
Output	<i>Dispersion value (bits)</i>

Entropy was explained on page 341. The Kullback-Leibler distance (KLD), sometimes also called the Kullback-Leibler divergence or relative entropy, is an operational definition of distance—not dispersion, but we will come to that—between two groups of data according to the known information theoretic measure with the same name (Cover & Thomas, 1991). Input data have the form of probability density functions (pdf:s), a form of normalized attention maps. If $p(x)$ and $q(x)$ denote the input pdfs for sets of eye-tracking data \mathcal{A} and \mathcal{B} , the KLD is defined as

$$\mathcal{D}(p||q) = \sum_{\mathfrak{X}} p(x) \log_2 \frac{p(x)}{q(x)} \quad (11.7)$$

over the alphabet \mathfrak{X} and is measured in bits. Zero-values in $p(x)$ results in a mathematical error (the logarithm of zero), and therefore Onat, Libertus, and König (2007) replaced all zero-values with a very small constant ($c = 10^{-9}$).

The KLD can be interpreted as the penalty in bits when assuming that the distribution of a random variable is $p(x)$ when it is in fact $q(x)$. It is used in information theory to measure the 'distance' between two probability density functions, although it is not a mathematically true distance, since it does not fulfil what is known as the triangle inequality, and is asymmetric. Symmetrization of the KLD can be achieved by calculating, for example, the harmonic mean $1/[1/\mathcal{D}(p||q) + 1/\mathcal{D}(q||p)]$. Another alternative symmetrization (and smoothing) of the KLD is called the Jensen-Shannon divergence (\mathcal{D}_{JS}). It is defined as

$$\mathcal{D}_{JS}(p||q) = \frac{\mathcal{D}(p(x)||\frac{p(x)+q(x)}{2}) + \mathcal{D}(q(x)||\frac{p(x)+q(x)}{2})}{2} \quad (11.8)$$

While the KLD is mathematically tractable it lacks intuitiveness; what does, for example, a dispersion of 14 bits mean? Moreover, a pdf cannot robustly be estimated from an attention map built from a few data samples, as pointed out by 't Hart *et al.* (2009).

The Kullback-Leibler distance can be transformed into a measure of dispersion, by calculating the KLD between the pdf of a single data point in \mathcal{A} , and the pdf generated from all other points in \mathcal{A} . This calculation is made repeatedly, by singling out each point in \mathcal{A} and comparing the KLD to the others. The final dispersion value is the average of all these similarity measurements. This is a very general method, that can be applied to any similarity measures to make a dispersion measure out of it, and has also been done with the average landing altitude measure.

The KLD measure was first used as a dispersion measure for eye-tracking data by Tatler, Baddeley, and Gilchrist (2005), and was later adopted by 't Hart *et al.* (2009), for example. Tatler, Baddeley, and Gilchrist generated a gridded AOI with squares of $2^\circ \times 2^\circ$; fixations were binned accordingly into these squares, and then normalized into a probability density function before computing the KLD. Their study compares the consistency in fixation locations between observers over time (measured as number of fixations), finding that it is high in the early phase, but decreases already after 3–4 fixations. Using the standard definition of entropy, Sawahata *et al.* (2008) found in a study of children watching TV programmes that dispersion varied with programme comprehension.

11.2.10 Average landing altitude

Target question	<i>What is the dispersion, measured as the average landing altitude, when dropping each data sample of each participant in \mathcal{A} onto the attention map of all the other participants?</i>
Input representation	<i>A set of raw data samples or fixations \mathcal{A} per participant</i>
Output	<i>A dispersion value</i>

The average landing altitude (ALA) dispersion measure is closely related to the similarity measure with the same name on page 373. The average landing altitude is calculated as the height at which the eye-tracking data of each viewer landed on the attention map created from all other viewers' eye-tracking data. Formally, dispersion is defined as

$$D_{ALA} = \frac{1}{N} \sum_{i=1,2,\dots,N} \frac{AM_{max}^i - AM^i(x_i, y_i)}{AM_{max}^i - AM_{avg}^i} \tag{11.9}$$

where AM^i denotes an attention map that has been generated by $N - 1$ data samples; all excluding the i^{th} sample. AM_{max}^i and AM_{avg}^i denote the maximum and average value of AM^i , respectively. $AM^i(x_i, y_i)$ represents the height at which data sample (x_i, y_i) 'lands'. Consequently, a random distribution gives a values close to 1, whereas a dense distribution has a dispersion close to zero.

The measure was developed by Nyström (2008) to overcome the lack in the KLD measure of an intuitive result unit. Inspired from the work by Itti (2005), Nyström used the measure to compare the dispersion across viewers' gazes before and after foveation of videos (where visual resolution is compressed at points at which the observer *did not* fixate), and found the dispersion to decrease when participants watched the foveated videos.

(11.8)

11.3 Position similarity measures

The similarity measures for position quantify similarity as the level of proximity between two sets of position data (raw data samples or fixations) with no particular order between them. They all take as input two sets of position data \mathcal{A} and \mathcal{B} , and they output a single value for that comparison. A single value output for a comparison may make it difficult to perform statistical tests determining whether \mathcal{A} and \mathcal{B} are significantly different from one another.

Like dispersion, similarity is also a vague concept for which there are several operational definitions. These definitions, rather than the concept, are termed 'measures'.

11.3.1 Euclidean distance

Target question	<i>What is the distance between two positions?</i>
Input representation	<i>Two position points (x_1, y_1) and (x_2, y_2), either as raw data samples, fixations or dwells</i>
Output	<i>Euclidean distance (pixels or degrees)</i>

The Euclidean distance, a much-used general measure in eye-tracking research, has also been used to operationalize similarity in position between two sets of data samples (raw or fixations). Euclidean distance between two points is defined as

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (11.10)$$

where (x_1, y_1) and (x_2, y_2) are the coordinates of the two points in two-dimensional space. Rao, Zelinsky, Hayhoe, and Ballard (2002) use the Euclidean distance as an operational definition of position similarity. Hart *et al.* (2009) use Euclidean distance to quantify consistency between observers at each time instant. However, they argue that the KLD is better seen over a whole trial, since the Euclidean distances summed over time would give a low similarity value when all observers visit the same positions, but in different order.

11.3.2 Mannan similarity index

Target question	<i>How similar in position are the raw data samples or fixations in \mathcal{A} to those in \mathcal{B}?</i>
Input representation	<i>Two sets of data samples \mathcal{A} and \mathcal{B}</i>
Output	<i>A distance (pixels)</i>

An early attempt to operationalize the similarity between two populations of eye-tracking data was presented in Mannan *et al.*, 1995, 1996 and Mannan, Ruddock, & Wooding, 1997. It is closely related to the nearest neighbour index. Given two sets of positions, \mathcal{A} and \mathcal{B} from raw data or fixations with M and N points, Mannan *et al.* used a measure where the squared Euclidean distance d^2 from each position i (of M) in \mathcal{A} to the closest position j (of N) in \mathcal{B} was calculated and vice versa. The similarity index was then calculated as the average of all squared distances

$$D_{\text{Mannan}} = \frac{1}{M+N} \left(\sum_{i=1}^M \min d_{i,j}^2 + \sum_{j=1}^N \min d_{i,j}^2 \right) \quad (11.11)$$

As pointed out by Tatler, Baddeley, and Gilchrist (2005) as well as Underwood *et al.* (2008b), among others, this type of distance-based measure has a number of severe limitations. Individual fixations may have disproportional impact on the overall similarity index,

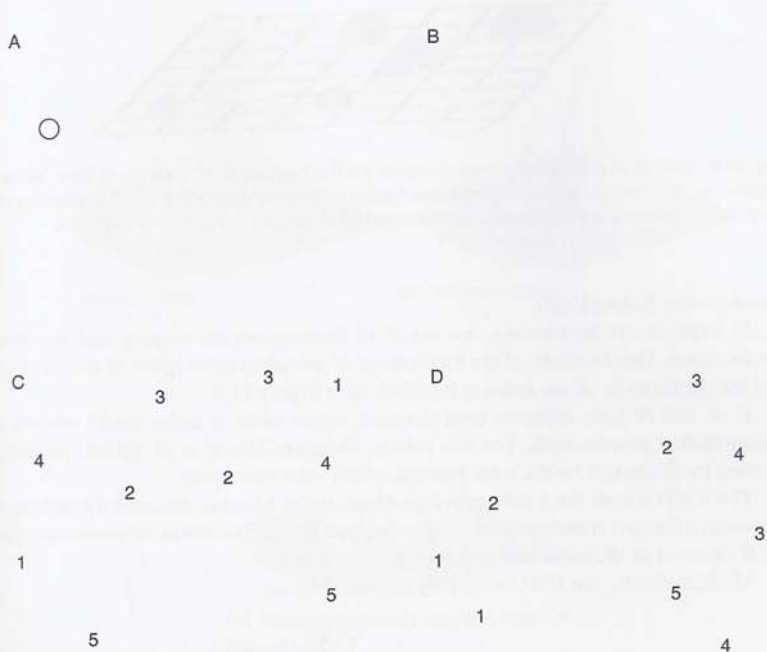


Fig. 11.8 Examples from Underwood *et al.* (2008b) pointing out limitations with the Mannan similarity index. A shows the basic case where fixations in \mathcal{A} (grey) can be mapped to the closest fixation in \mathcal{B} (white). Figure B exemplifies the problem that a single fixation in \mathcal{A} will be mapped on all fixations in \mathcal{B} and make the similarity value artificially low. Case C shows spatial mappings between the closest fixations in either distribution will ignore scanpath order, making the Mannan measure a pure position measure. Case D further exemplifies the need for scanpath simplification before alignment (pp. 273–278). With kind permission from Springer Science+Business Media: Knowledge-Based Patterns of Remembering: Eye Movement Scanpaths Reflect Domain Experience, 2008, Geoffrey Underwood.

look added page for back (downside backside)

and yield an index that is clearly not commensurate with the intuitive similarity. In extreme cases, the Mannan distance can be very unintuitive, as in Figure 11.8(B), where the positions in \mathcal{B} are packed in a small region somewhere in the stimulus, whereas \mathcal{A} covers the whole stimulus area. Since all distances to the closest point in \mathcal{A} are low, the similarity index would indicate a good match between \mathcal{A} and \mathcal{B} , although intuitively this is not the case. After the advent of improved similarity measures, the Mannan measure can be considered obsolete.

11.3.3 The earth mover distance

Target question	What is the cost of transporting the total fixation durations from \mathcal{A} to \mathcal{B} ?
Input representation	Two sets \mathcal{A} and \mathcal{B} of fixations with position and duration
Output	A distance/cost (pixels)

The earth mover distance (EMD) is a solution to the so-called transportation problem, which concerns minimizing the cost for moving any amount of matter from one set of source locations to another set of destinations. It can be solved by the Hungarian algorithm originally

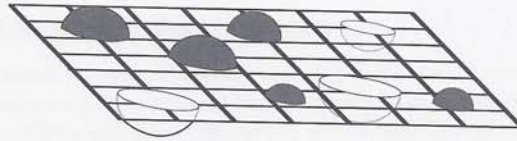


Fig. 11.9 Illustration of the earth mover distance: the five fixations in \mathcal{A} are piles of earth, while the three fixations in \mathcal{B} are holes. Both piles and holes have a volume corresponding to the fixation durations. The earth mover distance is the minimal cost (distance) for moving all the earth into the holes.

developed by Kuhn (1955).

In terms of eye movements, one set \mathcal{A} of fixations are the origins, and the other set \mathcal{B} the receivers. The durations of the fixations in \mathcal{A} are taken to be 'piles of matter' (i.e. earth), and the durations in \mathcal{B} are holes to be filled, as in Figure 11.9.

If \mathcal{A} and \mathcal{B} have different total duration, either earth or holes would remain after the transportation process ends. For this reason, Dempere-Marco *et al.* (2006) normalize each fixation by dividing it by the total duration of the set it belongs to.

The EMD solves the fundamental problem of the Mannan distance by making fixation durations a limited resource. This means that one single fixation in \mathcal{A} cannot be the closest to all fixations in \mathcal{B} , because it will long be consumed.

Mathematically, the EMD similarity is calculated as

$$D_{\text{EMD}}(\mathcal{A}, \mathcal{B}) = \frac{\sum_{i=1}^m \sum_{j=1}^n d(a_i, b_j) f_{ij}}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}} \quad (11.12)$$

where $d(a_i, b_j)$ is the distance in stimulus space, and f_{ij} the minimal transportation distance found by the algorithm.

Dempere-Marco *et al.* (2006) applied this measure to compare visual search patterns of radiologists interpreting CT images where lung disease was suspected, but argue that their EMD measure has a broader applicability as a general similarity measure.

11.3.4 The attention map difference

Target question	Where in a stimulus is the difference in attention map altitude between \mathcal{A} and \mathcal{B} large, and where is it smaller?
Input representation	Two attention maps \mathcal{A} and \mathcal{B}
Output	An attention map or a similarity value

The attention map difference is simple: just subtract the attention map for \mathcal{A} from the attention map for \mathcal{B} , or vice versa. The originator, Wooding (2002a), proposed that this measure was appropriate to quantify the similarity after both maps were normalized to unit height. Figure 11.10(c) shows the difference map between the attention maps in Figures (a) and (b). To obtain a single value representing the similarity, the average value of the difference map can be calculated.

There are at least three varieties of operational definitions for this measure:

1. Simple subtraction of one map from the other.

$$AM_{\text{diff}} = AM_{\mathcal{A}} - AM_{\mathcal{B}} \quad (11.13)$$

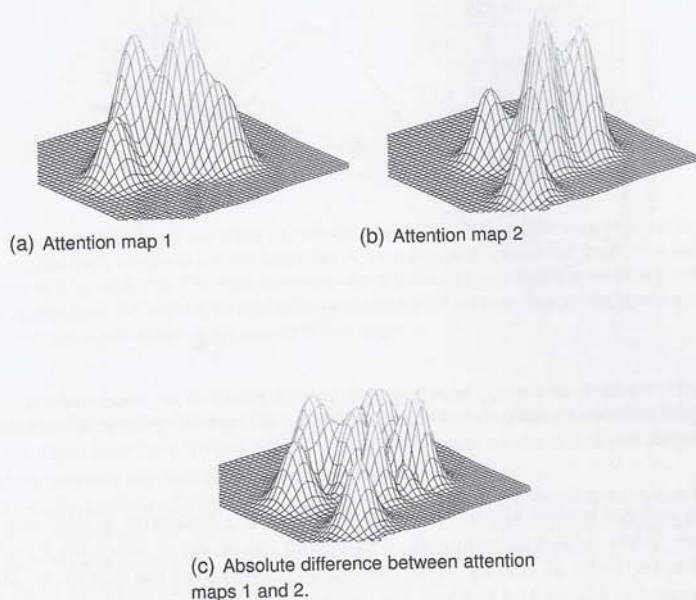


Fig. 11.10 Calculating similarity by subtracting two attention maps.

This gives a difference map that can have both positive peaks and negative valleys. It shows the direction of the difference in addition to the magnitude.

- The absolute difference between the two maps, which is always positive.

$$AM_{\text{AbsDiff}} = |AM_{\mathcal{A}} - AM_{\mathcal{B}}| \quad (11.14)$$

Figure 11.10 exemplifies this case.

- The *squared error* (SE), which squares the differences between the maps of dimensions $m \times n$.

$$AM_{\text{SE}} = (AM_{\mathcal{A}} - AM_{\mathcal{B}})^2 \quad (11.15)$$

Very often the mean squared error (MSE) is used instead of the squared error

$$AM_{\text{MSE}} = \sum_{i=1}^m \sum_{j=1}^n (AM_{\mathcal{A}}(i, j) - AM_{\mathcal{B}}(i, j))^2 \quad (11.16)$$

In spite of its simplicity, this measure appears not to have been used in actual research.

11.3.5 Average landing altitude

Target question	What is the similarity in position between the two groups \mathcal{A} and \mathcal{B} of fixations or raw data samples?
Input representation	Two sets of position data \mathcal{A} and \mathcal{B} containing raw data or fixations
Output	A similarity value

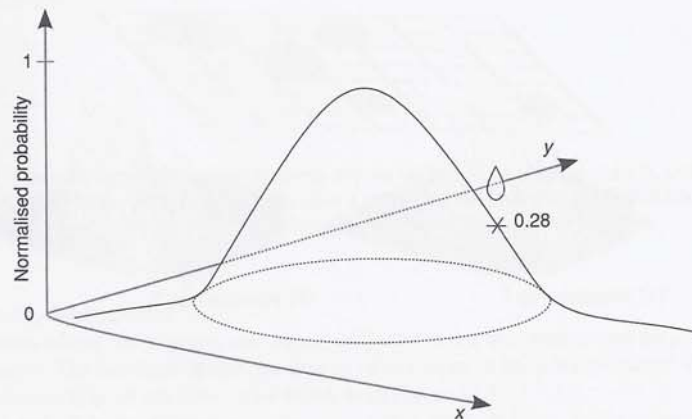


Fig. 11.11 Similarity as average landing altitude: for each fixation in \mathcal{A} , measure the landing height (for instance, 0.28) when dropping it on an attention map formed from \mathcal{B} , and then form an average over all such heights.

The average landing altitude has been used as both a dispersion (p. 369) and a similarity measure. When measuring similarity, it calculates the height at which the fixations (or raw data samples) from \mathcal{A} land on the attention map created from the data in \mathcal{B} . This principle is illustrated in Figure 11.11, where the raindrop represents one fixation. Consequently, the similarity can be defined as

$$S_{ALA} = 1 - D_{ALA} \quad (11.17)$$

where D_{ALA} is the dispersion value given in Equation (11.9). AM in Equation (11.9) should now be interpreted as an attention map generated from set \mathcal{A} , whereas (x_i, y_i) represent data samples from set \mathcal{B} .

11.3.6 The angle between dwell map vectors

Target question	How similar is the proportion of dwell time to AOIs between \mathcal{A} and \mathcal{B} ?
Input representation	Two matrices \mathcal{A} and \mathcal{B} of gridded AOIs with dwell time in each AOI cell
Output	A similarity value

The angle between dwell map vectors is a mathematically elegant similarity measure first used by Pomplun *et al.* (1996). It uses the gridded AOI representation from page 212, in which a grid is put onto the stimulus image and dwell time is calculated for each cell. To understand the principle, we need a simplistic example. Figure 11.12 illustrates how vectors are formed from the gridded AOIs in a stimulus with only two AOIs, and how angles are formed between the vectors.

Since dwell times are always positive, the angle between the vectors can vary from 0° to 90° . Similarity is then expressed as the cosine of this angle, with a value of one indicating that the vectors are identical (0°), and a value of 0 indicating that the vectors are maximally different (90°).

Note that this measure only compares how similar proportions of dwell time are in each matrix. In particular, \mathcal{A} and \mathcal{B} would be considered equal by this measure if each cell value

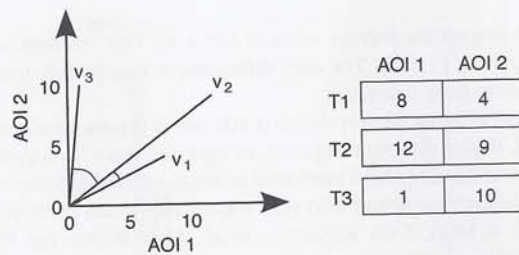


Fig. 11.12 Assume that there are three recordings, trial 1, 2, and 3 of data, now represented as T1–T3 in the figure. For each trial there are the same two AOIs in this simple example. There is a synthetic dwell time in each AOI for each trial. The *map vectors*, v_1 – v_3 are based on using the dwell times in the grid AOIs as vector coordinates. An angle is formed between each pair of vectors. The similarity value between two recordings—trials in this case—is the cosine of that angle.

in \mathcal{A} is twice the corresponding value in \mathcal{B} , since the two vectors would then be pointing in exactly the same direction. To test whether the dwell times in cells also equal in size, we need to calculate whether the two vectors are also of equal length.

DeAngelus and Pelz (2009) use dwell maps with normalized dwell time values (in %, with a sum of 1), which put the vectors on the unit circle. Their variety of the measure calculates the distance between (the tips of) the vectors, which roughly corresponds to the angle between vectors when the dwell map consists of normalized values.

Figure 11.12 shows gridded AOIs with only two cells. With matrices containing more cells, we have $n_x \cdot n_y$ number of cells, each with a dwell time value s_n within it. A vector $\mathbf{v} = s_1, s_2, \dots, s_{n_x \cdot n_y}$ is then formed for each image and group of viewers. The similarity between groups 1 and 2 is defined as the cosine of the angle θ between the $n_x \cdot n_y$ -dimensional vectors \mathbf{v}_1 and \mathbf{v}_2 , and can be calculated as

$$\cos \theta = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|} \quad (11.18)$$

Observe that when using a measures based on gridded AOIs, the output depends on the precise division of stimulus space into grids. Therefore Pomplun *et al.* (1996) calculated a final similarity by averaging over a number of different grid sizes.

11.3.7 The correlation coefficient between two attention maps

Target question	How similar are two sets of data samples \mathcal{A} and \mathcal{B} ?
Input representation	Two sets of data samples \mathcal{A} and \mathcal{B}
Output	A similarity value

The correlation coefficient between attention maps uses attention maps in the form of gridded AOIs. Given two attention maps $AM_1(x, y)$ and $AM_2(x, y)$, the correlation can be computed as

$$r_{pq} = \frac{\sum_{x,y} (AM_1(x, y) - \bar{AM}_1(x, y))(AM_2(x, y) - \bar{AM}_2(x, y))}{\left[\sum_{x,y} (AM_1(x, y) - \bar{AM}_1(x, y))^2 \sum_{x,y} (AM_2(x, y) - \bar{AM}_2(x, y))^2 \right]^{1/2}}$$

where $\bar{AM}(x, y)$ denotes the average value of $AM(x, y)$. This measure is closely related to the cosine from Equation (11.18). The only difference is that in Equation (11.18), the vectors have been shifted to have zero mean.

The advantage of using the correlation coefficient is that the results are intuitive and hence easy to interpret. Beside directly comparing two populations of eye-tracking data, the correlation coefficient and the KLD have been used to estimate the similarity between computationally generated *saliency maps* and attention maps from human observers (see Ouerhani, Von Wartburg, Hügli, & Müri, 2003; Rajashekar *et al.*, 2004; Rajashekar, Van Der Linde, Bovik, & Cormack, 2008).

11.3.8 The Kullback-Leibler distance

Target question	<i>How large is the position similarity between the two groups of position data \mathcal{A} and \mathcal{B}?</i>
Input representation	<i>Attention maps \mathcal{A} and \mathcal{B}</i>
Output	<i>A similarity value (bits)</i>

The KLD similarity measure, mathematically defined on pages 368–369, appears to have been first used with eye-tracking data by Rajashekar *et al.* (2004), Nyström *et al.* (2004) and Tatler, Baddeley, and Gilchrist (2005). Rajashekar *et al.* (2004) used the symmetric KLD to quantify the distance between fixation predictions and recorded fixations from humans. Dempere-Marco *et al.* (2006) used the KLD measure in the feature domain in a study on mammography radiologists, to examine how similar features looked at were to visual features characteristic of malign changes. Bestelmeyer *et al.* (2006) compared the restrictedness of scanpaths in schizophrenia and bipolar patients using the KLD measure, as a diagnostic tool. Comparing fixation distributions on videos between humans and monkeys, Berg *et al.* (2009) found large differences using the KLD between distributions in a permutation test (Monte-Carlo simulation), which reveals how our visual systems differ from our closest evolutionary ancestors.

Levy, Bicknell, Slattery, and Rayner (2009) compared fixation distributions on a sentence before and after reading an ambiguous word. Fang, Chai, and Ferreira (2009) used the Jensen-Shannon divergence to investigate the change in dispersion in fixations on a scene before and after two adjacent utterances.

11.4 Position duration measures

The position duration measures all concern how long participant gaze stays within a position. The position stayed within is always either that of a fixation or an AOI. We call the fixation-based position duration *fixation duration* and the dwell-based position duration *dwell time*. With the exception of IMSI, the other measures are all varieties of fixation durations and dwell times.

11.4.1 The inter-microsaccadic interval (IMSI)

Target question	<i>How long is the inter-microsaccadic interval (IMSI)?</i>
Input representation	<i>An inter-microsaccadic interval</i>
Output	<i>A duration value (ms)</i>

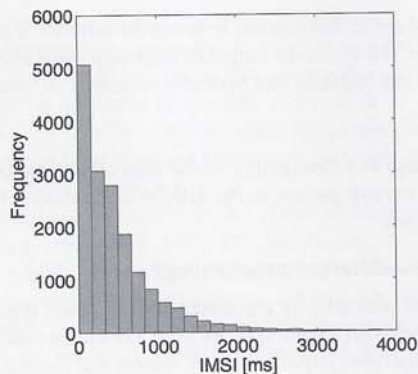


Fig. 11.13 Histogram over the distribution of inter-microsaccadic interval (IMSI), reproduced from Engbert (2006) with kind permission from Elsevier B.V.

The inter-microsaccadic interval (IMSI) is a very uncommon position measure, reported only by a few microsaccade researchers. The IMSI is to microsaccades what fixations are to saccades—periods of fixation-like ocular stability within the very small movements of microsaccades. Correctly detecting and calculating them puts high demands on both the eye-tracker hardware and filters and algorithms for event detection.

IMSI values have a distribution very similar to that of fixation durations, as Figure 11.13 shows.

11.4.2 Fixation duration

Target question	<i>For how long was the eye still in a position?</i>
Input representation	<i>A fixation</i>
Output	<i>The fixation duration (ms)</i>

Fixation duration is likely to be the most used measure in eye-tracking research. It is sometimes called ‘fixation time’, but also ‘dwell time’, or ‘dwell time of the fixation’, which may be confused with the most common use of the term dwell time as defined on page 386, which is the time from entering to exiting an AOI. Oster and Stern (1980) use the terms *saccadic reaction time* and *intersaccadic interval*.

There are a host of methodological issues surrounding the fixation duration measure, and we will go through them one by one.

The many different definitions

Informally, a fixation is defined as a period of time when the eye is relatively still (the oculomotor definition), but some definitions add visual intake as an additional criterion on fixations (the processing definition).

In reality, fixation durations are calculated by the fixation detection algorithms described in Chapter 5, which do not care about visual intake, and have different definitions of stillness. As a result, we have a situation in which fixation durations are solely defined by the event detection algorithms and their settings. For example, there are the ‘I-DT fixation durations’, the ‘velocity algorithm fixation durations’, and the ‘EyeLink fixation durations’, and many more varieties due to settings, and these are related but different fixations. Researchers in general,

however, have tended not to differentiate between the different algorithmic operational definitions, but surprisingly often treat the output of their own particular event detection algorithms as revealing generic true fixations that perfectly overlap with visual intake.

Data quality

As pointed out on page 161, the quality of the data affects fixation durations severely. Furthermore, if there is smooth pursuit in the data, most fixation detection algorithms will give faulty fixation durations.

Different fixations—different processing?

Different fixations are undoubtedly associated with different types of processing. In reading research, the first fixation on a word appears to be associated with lexical activation, and later fixations with discourse integrative processes. Inhoff and Radach (1998) note that it may be confounding to form averages over fixations that are that qualitatively different. This may be an issue for others than reading researchers, as the distinction between first and later fixations during scene viewing has been made by several researchers, for instance Henderson *et al.*, 1999. Reading researchers and others furthermore differentiate between look-ahead (progressive) fixations and look-back (regressive) fixations, which may or may not be qualitatively different. Also, Land *et al.* (1999) classify fixations into four distinct functional types during everyday activities (see p. 358).

The different types of processing of fixations *may* be reflected in their durations. For instance, McConkie, Reddix, and Zola (1992) show that fixations *below but not above* 140 ms are affected by lexical properties of the text read. Buswell (1935):142 noted that the earliest fixations in a picture are shorter (around 210 ms) than later (around 360 ms) fixations. This has later been interpreted as an early orienting period, followed by a more scrutinous inspection of informative details, which could motivate a division of fixations according to *ambient* and *focal* processing modes (Unema *et al.*, 2005).

Furthermore, Henderson and Pierce (2008) give evidence that one population of fixation durations is constant, while another is under the direct moment-to-moment control of the participants' ongoing scene analysis.

The participant idiosyncrasy

When a participant repeats a task, average fixation durations remain similar across trials; however different people have different average fixation durations (Andrews & Coppola, 1999, Rayner, Li, Williams, Cave, & Well, 2007, Johansson *et al.*, 2011). These authors conclude that there is an endogenous component that correlates ($r = 0.5-0.8$) with fixation duration, requiring an adequate experimental design as well as statistical handling of data (pp. 83-85).

Dependency between successive fixations

Fixation durations are not entirely independent of one another: in scene viewing, long fixations are more often followed by other long fixations, and short fixations by short fixations (Tatler & Vincent, 2008). In visual search, Hooge, Vlaskamp, and Over (2007) found fixations on difficult search elements to be followed by long fixations on the next element, irrespective of its difficulty. Importantly, independence is required by many statistical tests.

Attention

Just and Carpenter (1980) formulated the influential strong eye-mind hypothesis, according to which there is no appreciable lag between what is fixated and what is processed. If this hypothesis is correct, then when a participant looks at a word or object, he also simultaneously processes it, for exactly as long as the recorded fixation. As can be seen below, to a large

extent research supports at least a general eye–mind hypothesis.

However, visual attention—the spatial locus of intake and processing—moves slightly before the eye does. Deubel (2008) and others have shown that attention may be as much as 250 ms ahead of the eye, in particular in specially designed tasks called anti-saccade tasks (p. 305). It is not known whether this large temporal lag persists in more natural tasks, but most eye-tracking research is conducted and interpreted as though attention and fixation were synchronous events, and they probably are not.

It should be noted that there are several models of eye-movement programming in reading and scene perception (e.g. Henderson, 1992; Reichle, Rayner, & Pollatsek, 2004; Engbert, Nuthmann, Richter, & Kliegl, 2005) which attempt to capture and predict the proportion of fixation durations which can rightly be allocated to information processing, attention shifts, and saccade programming. The first two mentioned, by Henderson (1992) and Reichle *et al.* (2004) respectively, are strictly serial, arguing that attention does not shift until information processing is complete. However, Van Diepen and D'Ydewalle (2003) found that this does not hold across the board, since in scene viewing, masking the information visible in the visual periphery early during fixations using a gaze contingent window leads to increased fixation durations. This should not happen if a serial mechanism controls fixation duration, since information at fixation should always be processed first. The latter model cited, SWIFT, developed by Engbert *et al.* (2005), operates on the assumption that the attentional spotlight is more distributed, and a default timer regulates eye-movement triggering. Hence, SWIFT is a parallel model not a serial one, as attention (and thus fixations) can shift before information processing is complete.

There are many subtleties to the modelling of fixations and attention, and saccade programming. The main point is that there is a whole research literature on fixation duration and attentional shifts, regarding the two as distinguishable entities, therefore, if individual fixation durations are important for your results, it is worth bearing this in mind, and it should not just be assumed that the entirety of a fixation duration represents cognitive processing or 'visual intake'.

On the positive side, we can be fairly sure that attention and saccades are tightly coupled (Deubel & Schneider, 1996)—you can think of this like a rubber band, where stretching the rubber band to one point (the point where attention is allocated) means the other end of the rubber band (or the fixation point) will naturally follow. Therefore, when a participant executes a saccade to a target, you can be certain that attention has just moved to the same place. The fixation is only a short time behind attention, which indicates the metrics of the next eye movement.

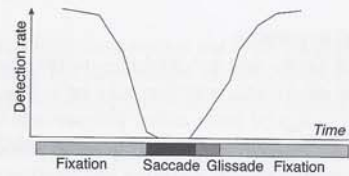
Note that in real-world tasks, outside the laboratory from which the data and theorizing in this section derives, fixating something does not entail close attentive processing of it, and does not guarantee a trace in working memory of all the features of the object looked at (Triesch *et al.*, 2003).

Finally, we must not forget that some processing trace of a fixated item may continue for a very long time after the eye has left the fixated position. This is evidenced by the fact that we learn from reading.

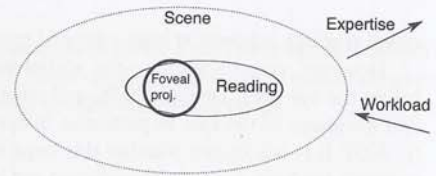
The duration of the intake period and saccadic suppression

Fixations produced by the event detection algorithms of Chapter 5 only concern the physical motion of the eye. The reason most researchers have for using fixation durations is that they are assumed to reflect perceptual intake and processing.³⁰ Generally speaking, this is a fair

³⁰Although many researchers have devoted their careers to understanding the *functional mechanisms* of fixation durations on eye-movement programming it its own right; how and when a saccade is triggered, therefore not using



(a) Principle of saccadic suppression. Short flashes that are produced during a saccade are not reported by participants, and some of this effect spills over to neighbouring fixations.



(b) Principle for functional visual field. Visual intake takes place from an area larger than the foveal projection, expertise increases the area, and a higher workload decreases it. The functional field is asymmetrical for reading Westernised texts, and information may be gathered from a larger area during scene viewing.

Fig. 11.14 Visual intake does not coincide perfectly with fixations.

assumption *if* the intake period equals the period of stillness detected by the algorithms.

Most eye-tracking researchers know that their participants are effectively blind during saccades, but it is less known that this 'blindness' spills over to part of the fixation, and hence affects our use of the fixation duration measure, as illustrated in Figure 11.14(a). Typically *pre-saccadic suppression* shuts down visual intake for 30–40 ms preceding the start of a saccade, while *post-saccadic suppression* follows thereafter for a duration lasting around 100–120 ms (Volkman, 1986). For a typical saccade of around 30 ms duration, some of the saccadic suppression spills over to the following fixation, so that in theory, intake and processing of the fixated position can start only after some 70–80 ms after the start of the saccade. The longer the saccade, however, the more of the suppression is consumed by the saccade, and the earlier the processing can start in the next fixation, which can thus be shorter. If fixation duration is used as a *precise* measure of processing, it is therefore advisable to measure also the duration of the preceding saccade, or at least refer to models of fixation control and saccade generation (e.g. Findlay & Walker, 1999, or some of the others mentioned in the previous section).

For a discussion as to *what* cognitive processes are suspended during saccadic suppression, see p. 321 and Irwin and Brockmole (2004). This is a large research area and full discussion is outside the scope of this chapter.

Glissades

Some authors argue that visual intake and processing starts not only directly when the fixation starts (the one detected by the algorithms), but already during the glissadic aftermaths of the saccade. For instance, Inhoff and Radach (1998) argue that there are good reasons for assigning the glissades to the fixation, since studies they quote have shown that brief flashes can be detected during the glissadic period. A counter argument is that the glissadic motion would smear the retinal image, making processing of fine texture difficult. As we saw on page 165, the glissadic velocity can be up to 130°/s, well above that of many saccades, so the smear is considerable, if intake is at all open. Event algorithms vary in what they assign the glissade to—saccade or fixation—and also in how systematically they assign them to one or the other. All in all, this contributes to making fixation duration values less comparable, at least between studies.

fixation duration simply as a measure which reflects processing of something else.

The functional visual field

How large is the area in which participants can take in meaningful information during a single fixation? In scene perception, this area has been given several names: 'functional field of view' (Mackworth, 1965), 'useful visual field' (Saida & Ikeda, 1979), 'functional visual field' (Nelson & Loftus, 1980), or 'visual span' (Reingold *et al.*, 2001), while in reading research, the functional visual field goes under the name 'perceptual span', (Rayner & Pollatsek, 1989). Engel (1971) used the term 'conspicuity area' in visual search to describe "the retinal locus within which the object to be searched for was noticed in a single 75 msec exposure". The principle and several important factors are illustrated in Figure 11.14(b).

It is clear that the foveal projection of 1.5° of visual angle is not the only area in which intake is made. When reading, for instance, the perceptual span is asymmetric, stretching 3 degrees from the point of fixation into the direction of reading, and hardly 1 degree backwards. When radiologists scan for lung nodules, their functional visual field stretches over 5 degrees of visual angle (Kundel, Nodine, & Toto, 1991), but in general in picture viewing, it appears to be at least 10 degrees across (Shioiri & Ikeda, 1989). As with reading, a kind of expertise, the differences in span depend on what you are looking for and what you are practiced in. The larger functional visual field in scene viewing compared to reading perhaps means participants need to process more information in one fixation, which could be one explanation for why fixation durations on average are longer in scene viewing. Note, however, there are many other possible considerations: reading is highly automated with familiar words, while scenes in scene viewing experiments are likely to be comparatively novel, thus lacking in the same degree of automation as reading; moreover, participants are likely to have clusters of long fixations on objects in scene viewing, in between periods of scanning where no distinct object is fixated and fixation durations are short.

Note that the visual field is larger horizontally than vertically. Thus, human contrast sensitivity is better in the horizontal periphery than in the vertical (Banks, Sekuler, & Anderson, 1991), and detection is better in the horizontal compared to vertical dimension (Engel, 1977).

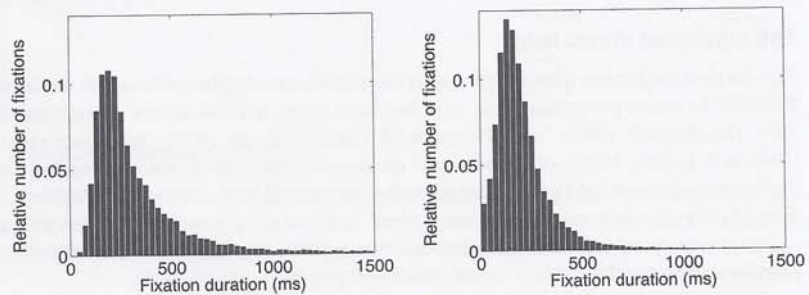
During many practical tasks, such as car driving, where peripheral vision is very important, the effective size of the functional visual field can be reduced by tasks or traffic situations that increase the *cognitive load* (Recarte & Nunes, 2000; Crundall, Underwood, & Chapman, 1999; Miura, 1992, 1990; Williams, 1988; Mourant, Rockwell, & Rackoff, 1969). For instance, car drivers' peripheral target detection decreases between 5 and 7 degrees of visual angle with increasing workload level, for all eccentricities (Crundall, Underwood, & Chapman, 2002). Furthermore, increased *driver age* further decreases the driver's peripheral detection between 8 and 24 degrees of visual angle (Gilland, 2004), while *expertise* in all tasks investigated have proven to increase the functional visual field.

Typical fixation duration values and how to interpret them

When you find the right algorithm and proper settings for your data, you will find that calculated fixations are frequently around 200–300 ms, but may be as long as several seconds (Karsh & Breitenbach, 1983; Young & Sheena, 1975), and as short as 30–40 ms, as exemplified in Figure 5.6 (p. 156). The distribution of fixation durations is not completely Gaussian; there is almost always a positive skew. Figure 11.15 shows the distribution of fixation durations in scene viewing (a) and during reading (b).

Average fixation durations definitely vary across different tasks and stimuli. Findings show one general pattern with several specific exceptions.

General finding—longer fixations equal deeper processing A longer fixation duration is often associated with a deeper and more effortful cognitive processing. This has been the conclusion in:



(a) Scene viewing—real still photographs from Nyström (2008). (b) Reading for 10–15 minutes. From case study 1 (p. 5).

Fig. 11.15 Distribution of fixation durations. Data recorded with a tower-mounted high-end eye-tracker at 1250 Hz; fixations calculated with BeGaze 2.1 velocity algorithm at $40^\circ/\text{s}$.

Reading Words that are less frequent, and would therefore require a longer lexical activation process, generally receive longer fixation durations (Rayner, 1998). More complicated texts give rise to longer average fixation durations, ranging from around 200 ms in light fiction to around 260 ms for physics and biology texts (Rayner & Pollatsek, 1989). More complicated grammatical structures give rise to longer fixation durations (Rayner, 1978). Also, longer fixation durations correlate with larger N400 amplitudes when taking ERP measurements (Dambacher & Kliegl, 2007), which is indicative of processing meaning and semantic content, particularly when the word is less frequently encountered.

Scene perception Out-of-context objects generate longer fixations than objects which fit the context (Henderson *et al.*, 1999; De Graef, Christiaens, & D'Ydewalle, 1990; Loftus & Mackworth, 1978). Finding the relevant information in blurred images increases fixation durations (Mackworth & Bruner, 1970).

Usability Harris and Christhilf (1980) found that pilots fixate longer on critical instruments from which information had to be extracted, rather than those requiring a mere check. Unema and Rötting (1990) found longer fixations when participants made more difficult mental calculations than when they made simpler ones. Stager and Angus (1978) could show shorter fixation durations with increased experience of a task. After comparing several experimental tasks, Oster and Stern (1980) conclude that fixation duration is a consequence of task requirements rather than a property of the saccadic system. Car-driver fixations when negotiating high-inclined curves on the road were longer than for non-inclined curves (Shinar, McDowell, & Rockwell, 1977), and more than twice as long in curves at night compared to daytime (Mortimer & Jorgeson, 1974). Experienced drivers have shorter fixations (262 ms) compared to novices (296 ms) (Laya, 1992). Also Goldberg and Kotval (1999) and many other applied researchers interpret longer fixations as an indication of the difficulty a participant has in extracting information from a display.

All this indicates functional links between what is fixated and cognitive processing of that item—the longer the fixation the 'deeper' the processing. However, the following are exceptions to this rule:

Longer fixations mean shallow processing In vigilance research, a long fixation is sometimes taken to indicate such a low arousal that participants are close to

daydreaming. This could be the reason why some have found that fixation durations of car drivers are longer on information-poor rural roads in comparison to information-dense urban roads (Chapman & Underwood, 1998). In this study however the I-DT algorithm was used with data containing smooth pursuit, and hence the concept of 'fixation' is different. Nevertheless, low arousal may in turn be the product of a non-demanding road requiring little search. In order to be able to use fixation duration as a measure of processing, it is therefore important to be able to argue that participants have not had tendencies to forget the task and start 'daydreaming'.

Higher stress results in shorter fixation durations In human factors, shorter fixations are indicative of a high mental workload (Unema & Rötting, 1990; Miura, 1990; Robinson, Erickson, Thurston, & Clark, 1972). Van Orden, Limbert, Makeig, and Jung (2001) developed a model using regression analyses from eye-movement data on a tracking task, showing that fixation duration was a robust and reliable predictor of tracking performance, again with short fixations correlating with high workload. The reader should be aware, however, that there is a distinction to be made between high workload which you complete successfully (giving longer fixations), and high workload which you struggle to engage with because you are too stressed (giving shorter fixations). Many short fixations across a web page are, according to Ehmke and Wilson (2007), indicative of the frequent usability problem, where a user goes to a page on the site, expecting to find specific details, but not finding them.

Expertise leads to longer fixation durations Expertise in a field such as chess, art, and goalkeeping result in longer (and fewer) fixations than for novices (Nodine, Locher, & Krupinski, 1993; Reingold *et al.*, 2001; Savelsbergh, Williams, Van Der Kamp, & Ward, 2002; Reingold & Charness, 2005). In this case the longer fixation does not mean more processing, but rather a different kind, which involves a larger visual span. With expertise it can be a matter of processing efficiency; fixations may be longer for the expert, but there are less of them compared to the novice because, with increasing skill, more information is extracted around the point of fixation making eye movements overall more efficient.

Neurological impairment means longer fixations Schizophrenia patients have longer fixations the more disturbed their thoughts (Ishizuka, Kashiwakura, & Oiji, 2007). Alzheimer patients make longer fixations when reading (Lueck, Mendez, & Perryman, 2000). Alcohol intoxication results in longer fixation durations (Moskowitz, Ziedman, & Sharma, 1976; Moser, Heide, & Kömpf, 1998). This should not be interpreted as more or deeper processing, but rather as indicative of a hampered processor. Longer fixation durations in infants have been associated with a poorer cognitive performance, both concurrently, and later in life (Colombo & Frick, 1999), in a line of research that uses not eye tracking but videotaping and direct observation.

Inspected stimulus moves quickly A few studies of inspection workers and internet users have noted very long fixations on a stimulus that just passes in front of the participant (Moraal, 1975). This is interpreted as a deliberate strategy of experienced viewers/inspectors facing a fast-moving stimulus, where time constraints make it more efficient than fixation-saccade sequences.

11.4.3 The skewness of the frequency distribution of fixation durations

Target question	<i>Do shorter or longer fixation durations dominate?</i>
Input representation	<i>A set of fixations from a trial or whole recording</i>
Output	<i>A skewness value for the frequency distribution</i>

Skewness refers to the degree of asymmetry in the distribution of fixation durations. The skewness calculation was introduced on page 315.

In human factors, this measure is taken to indicate differences in information acquisition, that may either be due to design or to task. Ellis and Smith (1985) found values between 1.84 and 2.64 during different phases of air traffic control work. Abernethy and Russell (1987) finds the value 1.83 for expert badminton players, compared to 1.77 for novices.

Harris, Tole, Ephrath, and Stephens (1982) evaluate two different designs of vertical speed indicators in aircraft, and conclude that the higher skew in one of them, because of more short fixations, indicating a lower mental workload. Some of the studies referred to by Rötting (2001, pp. 114–119) found differences in the measure as an effect of different phases of work, for instance as a pilot goes through the different checking phases before take-off. Megaw and Richardson (1979) show histograms with large differences in skew between inspectors of different materials. If it can be assumed that different fixation duration values indicate different cognitive processes, then the skew value reflects the relative prevalence of those processes.

11.4.4 First fixation duration after onset of stimulus

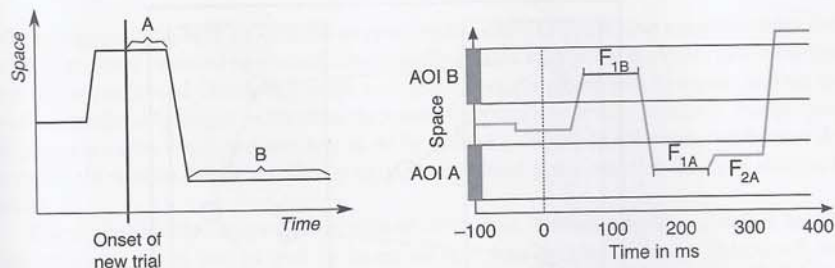
Target question	<i>How long was the first fixation on the stimulus?</i>
Input representation	<i>The first fixation on the stimulus after its onset</i>
Output	<i>The duration of the fixation (ms)</i>

The first fixation after onset has a particular status, as it coincides with the *very first* intake and processing of the attended part of the stimulus, and its duration reflects the immediate information processing. The initial fixation duration typically reflects a latency in the sense that it measures the time between the onset of a stimulus and the initiation of a saccade (see figures 11.16(a) and 13.1). There are important issues to consider when using first fixation durations:

First, how to measure fixation onset? It is a mistake to think that fixation onset is aligned with the stimulus onset, since the oculomotor fixation very often starts before the stimulus onset, and later only continues. When the stimulus software switches trial during a fixation, many recording softwares will split the fixation into two parts, as illustrated in Figure 11.16(a). If you have many short trials, this may affect the average first fixation duration significantly. In addition, it is likely that processing associated with the last fixation in one trial will spill over to the first fixation in the next trial. One way to alleviate this effect is to show a blank (noise) display between trials.

Should you use the first or the second fixations? Since the fixation position at stimulus onset has not yet been influenced by the stimulus content, the initial fixation is often excluded, and the *first fixation after the initial saccade* is counted as the first fixation 11.16(b). This way, the duration reflects the processing at the first actively chosen fixation position.

As pointed out on page 378, some care should be taken before forming averages between first and other fixations, as they may represent different cognitive processes.



(a) Fixations occurring on the cusp of a trial border, for instance when a new stimulus picture is onset, may be split in two, and the second half (A) falsely registered as the first fixation duration and position, while in reality (B) is the first complete fixation after trial onset.

(b) The two first fixations, F_{1B} in AOI B and F_{1A} in AOI A, with their durations indicated in black. Note that F_{2A} in AOI A is not a first fixation.

Fig. 11.16 Examples of position duration measures, and a problem with trial borders for first fixation durations.

11.4.5 First fixation duration in an AOI, and also the second

Target question	How long was the first (or second) fixation in an AOI?
Input representation	The first or second fixation on an AOI
Output	The duration of the fixation (ms)

The duration of the first fixation in an AOI, the *first fixation duration*, or just *FFD*, is interpreted as reflecting the time taken for fast processes such as recognition and identification. Note that despite the similarity in name, this is a very different measure compared to the previous one, first fixation duration after onset of stimulus. The current measure is specifically reserved for the first fixation *on a part of* the stimulus image, whereas the former referred to the very first fixation per se.

Thus, with this measure the AOI resides in the same stimulus that the participant has already been looking at for a period of time (see Figure 11.17). It is likely that the participant has seen the AOI in question using peripheral vision, and to a small extent processed it and items in its immediate proximity, in particular if the AOI is a word. For the previous measure, the first fixation is only preceded by enough processing to launch a saccade.

In reading, where this measure was first developed, it is considered to reflect the *lexical activation process*. The word properties that affect first fixation duration include word frequency, morphological complexity, metaphorical status, orthographic properties, the degree of polysemy, and other linguistic factors (Inhoff & Radach, 1998; Clifton *et al.*, 2007). The first fixation duration measure is now extensively used in reading research, second only to dwell time (also known as 'gaze duration').

There is also the *second fixation duration*, for instance F_{2A} in Figure 11.16(b). When a long word is hit by two subsequent fixations in the direction of reading, there is a systematic relationship between their landing positions and their durations. If the initial fixation hits the word on its initial few letters, it is short, and the subsequent fixation longer. If the initial fixation hits the centre of the word, it is longer, and the subsequent fixation shorter. According to Inhoff and Radach (1998), it is not clear whether this is an effect of different linguistic processes being computed. The second fixation duration is sometimes taken as a measure of

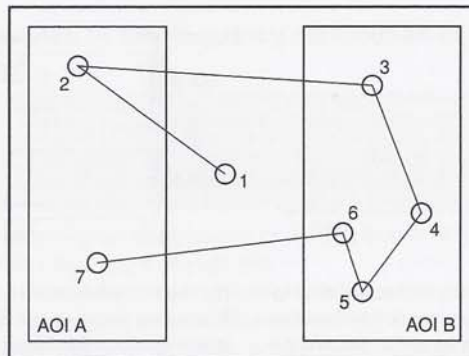


Fig. 11.17 Exemplification of five related position duration measures. *First fixation* after trial onset: 1. *First fixations* in AOI: 2 and 3. *Fixation durations* of 1–7 are calculated on a fixation-by-fixation basis. *Dwell time* is the time from entry to exit, namely the durations of: 2, and 3+4+5+6, and 7, respectively. For *Total dwell time* in AOI A, add the durations of fixations 2 and 7.

serial processing of long compound words (Pollatsek & Hyönä, 2005).

In scene perception, Van Diepen, De Graef, and D'Ydewalle (1995), Van Diepen, Wampers, and D'Ydewalle (1998) and Van Diepen (2002) show line drawings of scenes, finding that first fixation durations are longer when the fixated area is masked or degraded using gaze-contingent technology, which indicates that first fixation duration would work as a measure on visual information acquisition from the fixated area. Henderson *et al.* (1999); De Graef *et al.* (1990) showed that first fixation durations on semantically inconsistent and low-probability (hence more informative) areas in a picture are longer than for fixations on more plausible objects. In this case, first fixation duration is used as a measure not just of object activation, but of overall scene integration.

11.4.6 Dwell time

Target question	<i>For how long, measuring from entry to exit, did gaze remain inside the AOI?</i>
Input representation	<i>A dwell in an AOI</i>
Output	<i>The duration of the dwell (ms)</i>

A dwell is defined as *one visit in an AOI, from entry to exit* (p. 190). Terminology for the dwell time measure varies. In some parts of human factors research, the measure is called 'glance duration', and Loftus and Mackworth (1978) used the term 'duration of the first fixation' for the first dwell time in an AOI. Terms like 'observation' and 'visit' can also be found. In reading and some parts of scene perception research, dwell time is often called 'gaze duration', 'regional gaze duration', or even 'first-pass fixation time', and in psycholinguistics, Griffin and Spieler (2006) use the term 'gaze time'. Krupinski and Jiang (2008) use the term 'cumulative decision dwell time' for dwell time on lesions in medical images. Dwell time is used in most other eye-tracker-based research fields, and dwell is a more precise term than the ambiguous gaze.

The term 'attentional dwell time' is used for a completely different, non-eye-tracking measure, of the time it takes to release attention from a target that is being identified.

There are important distinctions between other measures, illustrated in Figure 11.17. First

of all, returns to the AOI are counted as new dwells. Also, as dwell time measures from entry to exit, not over repeated visits, it is conceptually related to fixation duration, and sometimes the two are confused. However, dwells tend to be more dispersed than fixations, and are typically considerably longer in duration, as it usually comprises several fixations. Furthermore, fixations are completely independent of AOIs and are calculated exclusively on the raw data samples themselves, while dwells can only be calculated if the stimulus has been divided by AOIs.

Dwell time is often defined as the sum of all fixation durations during a dwell in an AOI, but the measure can just as well be based on raw data samples. The raw data dwell time measure will include the durations of non-fixations such as blinks, saccades, and glissades, as well as fixations shorter than your minimal fixation duration criterion. One may argue that a lot of non-fixations means a lot of non-processing time, and that the raw dwell time measure is unsuited as a measure of cognitive processing. That is true in special cases, for instance if you were to have an AOI that is never looked at but is passed over by saccades a great many times. In general, however, this non-processing time is small (up to 20% on top of fixation durations), and should equal out across AOIs and across conditions. Also, considering the current state of the various fixation analysis algorithms (see Chapter 5), a measure based on fixations may contain an equal amount of imperfections.

Furthermore, a distinction is sometimes made between the first and the second dwell in an AOI, in analogy to first/second fixation duration in the AOI. The difference, of course, is that the second dwell is preceded by an exit from the AOI, while the second fixation is not.

Hyönä *et al.* (2003) present the reading measure 'extended first dwell time' (or, 'extended first-pass fixation time'), which is the first dwell time in an AOI, but including regression to other AOIs (previous text), assuming that the AOI is again returned to. This also works if the other region is a pictorial element. The idea is that regressive excursions are part of forming the understanding of the word regressed from, and should therefore be included in the processing time of an AOI, and that the dwell time should not be terminated at the regressive exit from it.

In many tasks, dwell time distribution is heavily right-skewed, as in Figure 11.18, which shows a histogram over dwell times on photographs in paper newspapers. Such data are commonly log-transformed before applying statistical tests (pp. 87–90).

Dwell time distributions can be constrained by constant trial durations. For instance, a trial duration of 3000 ms could give a distribution peak at 2700 ms for an AOI. As part of the variance can be considered to disappear outside of the 3000 ms, variance analyses (such as ANOVA) on the data may be inappropriate (p. 83).

The recorded dwell time for an object depends on the semantics of the object, and the task of the participant. The following research findings illustrate this:

Interest and informativeness Dwell time indicates interest in an object, or higher informativeness of an object. Friedman and Liebelt (1981) found that objects with lower probability of occurrence (defined as higher informativeness), were looked at longer than objects with high-rated likelihood of being present. Pieters, Rosbergen, and Hartog (1996) also observed that on second viewings of print advertisements, the dwell time to advertisement elements decreases. When the contents of an AOI changes considerably in the midst of a trial, the first return dwell time to the AOI (after the change) is larger (Ryan & Cohen, 2004). All this indicates a strong relationship between consecutive fixations on an item and how much you need to mine information from it.

Uncertainty and poorer situation awareness A higher dwell time may be indicative of uncertainty and poorer situation awareness. Ottati, Hickox, and Richter (1999) found that in a navigational task, novice pilots had a higher dwell time on the outside (through the

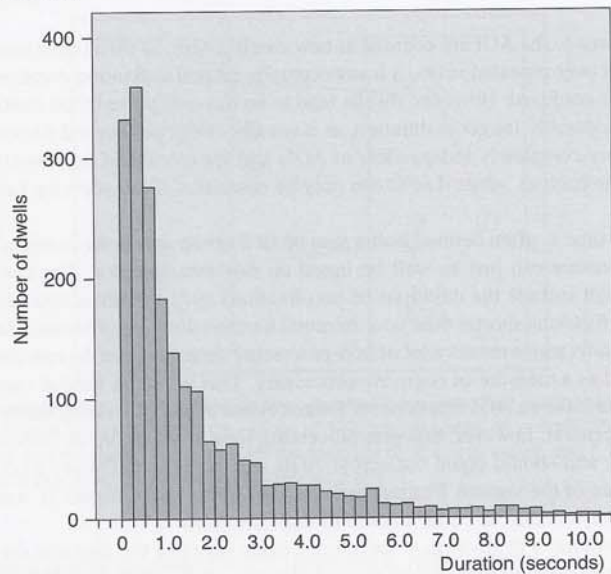


Fig. 11.18 Histogram with 2224 dwell times (unit: seconds) on newspaper photos during natural newspaper reading by 110 participants, using a headmounted eye-tracker at 50 Hz with Polhemus head-tracking. Bin size 250 ms.

window) than experienced pilots, which the authors attribute to uncertainty in locating navigational landmarks amongst pilots with less experience. Hauland (2002) in a large study of air traffic controllers, found that a higher dwell time on AOIs correlates with a poorer situation awareness.

Difficulty in extracting general information Longer dwell time may indicate difficulty in extracting information from a display, as put forward by Fitts *et al.* (1950), and Goldberg and Kotval (1999). Jacob and Karn (2003) note that dwell time is one of the most used measures in usability studies. In research on car driving for instance, a long-standing discussion is how long dwells to in-car instruments (radios, air control, and GPS etc.) can be without risk of accidents (Zwahlen, Adams, & De Bald, 1988; Rockwell, 1988)

Difficulty in extracting word information Rayner (1998), reviewing reading research using the fixation-based dwell time measure, concludes that dwell time ('gaze duration') is a good index both of word frequency—longer dwells relating to less frequent words—and of comprehension processes integrating several words. Dwell time on a word thus contrasts to first fixation duration, the other major reading measure. More generally, Rayner and Pollatsek (1989) argue that very fast cognitive operations, such as lexical activation and recognition, can be measured with first fixation duration, while slower cognitive processes affect dwell time. In spoken interaction, the dwell time on the interlocutor's (the speaker) mouth increases with the ambient noise levels, an indication that mouth movements play a role in hearing and understanding speech (Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998).

An upcoming conscious choice When participants compared abstract, unfamiliar shapes for attractiveness, and were asked to select one, they gradually increased dwell time on the item that was eventually chosen, up until it was finally selected. This has been termed

'the gaze cascade effect' (Shimojo, Simion, Shimojo, & Scheier, 2003).

In gaze-based interaction with computers, dwell time is the predominant criterion for deciding whether gaze on a button AOI should cause activation of the button function. In the basic version, a dwell time threshold of some 400 or 500 ms is required before activation. When the dwell time threshold is too low, buttons activate prematurely, and users experience what is known as 'the Midas touch': everything they look at activates (Jacob, 1991). Research in this area is intense at present, and several new designs exist that combine other selection criteria with dwell time (see for instance, Tall, 2008).

11.4.7 Total dwell time

Target question	<i>Over the whole the trial, how much time was spent in the AOI?</i>
Input representation	<i>A set of dwells on the same AOI</i>
Output	<i>The sum of dwell durations on the AOI (ms)</i>

Total dwell time is the sum of all dwell times in the one and same AOI over a trial (or other specified period). This may sound simple, but so far terminology for and usage of this measure is confusing. Rötting (2001, p. 120) uses the term 'gaze duration', which is also used in the reading research community for the common single dwell time. Journal papers and manuals also exhibit the terms 'cumulative dwell time', 'glance duration', 'gaze', 'total viewing time', 'total fixation time', 'fixation cycle', and 'time in zone'. Clifton *et al.* (2007) use the term 'total reading time' for "the sum of all fixations in a region, both forward and regressive movements" (i.e. total dwell time). Inhoff and Radach (1998) point out that although total dwell time (which they call 'total viewing durations') seems to be sensitive to linguistic processes that operate after the word has been identified, the measure should be refined by separating dwell time during first reading from dwells on the same word in subsequent readings.

Total dwell time subsumes the whole duration of a trial, and should therefore be sensitive to slow and long-term cognitive processes, but the lack of terminological precision in much of the literature before 2010, in particular the lack of distinction between single and total dwell time, makes it almost impossible to review what the measure has been used for. However, Henderson and Hollingworth (1999) conclude in their review that studies on scene perception "show a clear effect of the meaning of a scene region on gaze duration [which here means total dwell time] in that region, but a less clear effect on first fixation duration".

Reading-specific varieties of total dwell time include 'look-back fixation time' ('second-pass fixation time'), which is the sum of all dwells to a text AOI except the first one, and 'regression time', the sum of all dwells upon an AOI that follow a regression (Hyönä *et al.*, 2003).

When there are few AOIs, total dwell time suffers from the fixed trial duration restriction on variance analysis (p. 83) even more so than ordinary single entry dwell time.

11.4.8 First and second pass (dwell) times in an AOI

Target question	<i>How long was the first dwell in the AOI, and how long the second?</i>
Input representation	<i>The first (or second) dwell \mathcal{D} on an AOI</i>
Output	<i>The duration of \mathcal{D} (ms)</i>

First pass dwell time (also referred to as 'first pass gaze duration', 'first-pass fixation

time', and 'duration of the first fixation') is the reading research communities' term for the duration of the *first* dwell in an AOI, which may be a word, a region, or a sentence. In Figure 6.3 on page 191, the first pass dwell time in AOI 1 consists of two fixations, while the first pass dwell time in AOI 2 has a single fixation. When there is only one fixation, first pass dwell time equals first fixation duration (plus saccades, in the raw data option). Both AOIs have second pass dwell times, also, made up of 3 and 5 fixations each.

First pass dwell time has been proposed as a candidate measure for *early processing and object recognition*. For instance, Liversedge *et al.* (1998) argue that for long words, which are hit by many fixations, not only the first fixation involves early processing (lexical activation), but also several fixations thereafter combine to do this, and then the first pass dwell time is a better measure than the first fixation duration. Loftus and Mackworth (1978) find that the measure increases for semantically informative objects and Friedman (1979) that it increases for unlikely objects in its context. Henderson *et al.* (1999) found both first pass dwell time and second pass dwell time to be longer for semantically informative objects. However, Henderson and Hollingworth (1999) estimate total dwell time to be a better measure than first pass dwell time for studying object recognition.

Second-pass dwell time is defined by Hyönä *et al.* (2003) (under the names 'look-back fixation time' and 'second-pass fixation time') as the summed durations of all returning dwells to an AOI.

11.4.9 Reading depth

Target question	How 'deeply' is the text read?
Input representation	Text in an AOI and eye movement data
Output	A depth (in pixels) or proportion of the text looked at

Reading depth, also known as 'reading ratio', is a vague measure with several possible operational definitions. When using newspapers or other everyday written material, readers read only portions of the text, and may skip parts to continue with later text, as in Figure 6.5 on page 192. The purpose of the reading depth measure is to quantify how much of the text has been read. The following definitions have been used:

Centimetres Manually measuring how many centimetres have been read from scanpath visualizations and scene-overlaid videos is one option. This operational definition was used by Hansen (1994), who found the following relationship between text length and reading depth in newspapers: Triple the length of the text—e.g. from 20 to 60 cm—and you will have half as deep a reading—10% rather than 20% of the whole text. Although this entails an increase of 2 cm reading in 60 cm compared to the 20 cm text, this 2 cm gain costs 38 cm of additional unread text space. The variable type is ratio.

Dwell time divided by AOI area Holmqvist and Wartenberg (2005) and Holsanova, Rahm, and Holmqvist (2006) showed that broadsheet newspapers are read less densely (34 ms/cm² over all pages) compared to tabloid newspapers (50 ms/cm² over all pages), and that the most read article had an average value of 207 ms/cm², with the ads hovering around 5 ms/cm². The major advantage with this kind of measure is that it works for all sorts of combined stimuli, not just for text, as you can have pictures and words presented together. The variable type is again ratio.

The number of fixations per word in a text AOI If the entire newspaper article is read, we expect a value of about 0.8 (since not all words are fixated), but if a participant reads less of the text, we get a lower number. Poole, Ball, and Phillips (2004), and Poole

(2003) found the values 1.08–2.41 fixations per word in bookmark phrases that participants searched for on web pages. The variable type is nominal, as the data can only be counted; however, recalculation into a proportion is possible. Although not formally correct, ANOVA could be used if the data are normally distributed. A logit-transformation may be appropriate.

Dwell time per word in a text AOI Processing time is included in this operational definition of the measure. We can expect a value of around 200 ms/word when the text in the AOI is fully read, and lower values for a more shallow reading depth. This definition exists also in the reading research community in the form of two rare measures 'first pass reading times per character in a region' and 'total pass reading times per character in a region'. The variable type is ratio.

Ratio to baseline Record a full reading of the stimulus, and use as a baseline. If a supermarket customer reads on a food package for 2.5 seconds, and we have previously recorded a full reading at 20 seconds for the complete content on the package, then the 2.5 seconds corresponds to a reading depth of 12.5%. This operational definition is particularly useful for stimuli with very mixed content, such as food packages and information graphics. The variable type is ratio.

Note that re-reading adds to the reading depth measure, even if the same text is read over and over again, except in the centimetre length operational definition by Hansen (1994). On the other hand, all varieties except Hansen's have no demands on consecutive reading, not even that the participant sticks to conventional reading order.

Figure 11.19 shows one application of the reading depth measure from the newspaper reading studies in our lab. It shows that a longer text is not only read less deeply, but reader comprehension is also better for shorter texts.

11.5 Pupil diameter

Target question	<i>How large is the pupil?</i>
Input representation	<i>Raw data</i>
Output	<i>Pupil diameter or area (mm, camera pixels, mm², or pixels²)</i>

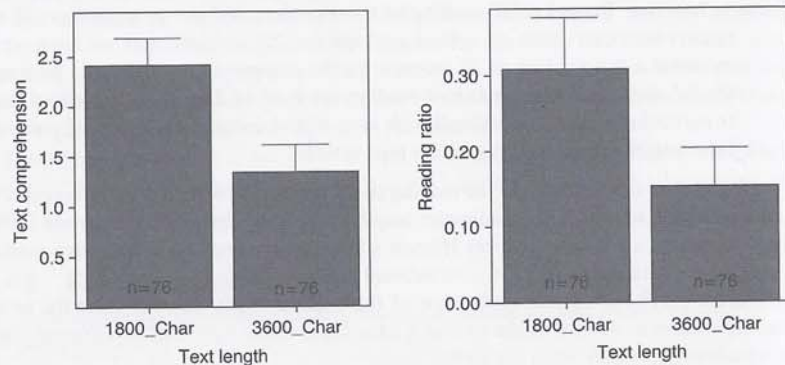
Pupil diameter ('dilation', 'size') is raw data provided as samples (in sample frequency). Values are typically given in pixels of the eye camera. Some eye-trackers can also report pupil diameter in millimetres after a simple calibration routine. When considering this measure as a property of eye position, it is important to point out that although in terms of the eye camera data points are recorded giving the pupil diameter when fixating a certain position, changes in pupil diameter may occur as a function of what has just been looked at, not what is presently being fixated (i.e. there may be some latency, see page 434). For the purposes of this chapter however, we deal with pupil diameter as it relates to the current position being looked at in space, as per the data recorded.

Operational definitions of pupil size

The recording software implements one of three different operational definitions. The simplest is to use the horizontal pupil diameter. The reason for measuring the horizontal diameter is that the vertical diameter is too sensitive to eyelid closure. With extreme gaze directions, however, the optical perspective may cause the horizontal diameter calculation to underestimate pupil diameter. Fitting an oval to the pupil image of the eye video and calculating the largest diameter, irrespective of direction, somewhat remedies the error due to gaze direction,



(a) Two identical newspaper folds, except that one article is shorter in the leftside version, and an advertisement fills the remaining space.



(b) Text comprehension is significantly better for shorter newspaper articles.

(c) Reading depth (ratio) is significantly lower for longer newspaper articles.

Fig. 11.19 In newspaper reading, article length influences both reading ratio (in this case number of fixations per word in the AOI) and text comprehension. Recorded using a head-mounted eye-tracker at 50 Hz with Polhemus head-tracking, and 40 participants on authentic but manipulated newspapers.

but this is more sensitive to eyelid closure. Calculating the area of the pupil is sensitive to both eyelid closure and gaze direction. These errors are known, but their magnitudes not systematically investigated (Klingner, Kumar, & Hanrahan, 2008; Pomplun & Sunkara, 2003).

Analysing pupil diameter in data recorded from a remote system may also introduce artefacts, since motion of the head closer to and further away from the camera also changes the pixel size of the pupil in the camera image. Measuring the camera-eye distance and applying some trigonometry can remedy this problem, but any noise or latencies that you have in the measured distance will be inherited in your pupil dilation measure. Therefore, pupil dilation is best recorded with a system that has a fixed distance between camera and eye. Beware of systems with automatic zoom in the camera, which can in itself cause large variations in recorded pupil dilation. Also, if your participant moves so much that now and then, part of the pupil is outside the eye camera image, data will not be valid.

Pupil size and luminance

When using pupil diameter as a measure of cognitive or emotional states, it is important to remember that the cognitive and emotional effects on pupil diameter are small and easily drowned in the large changes due to variation in light intensity. Varying brightness of the stimulus (screen) may easily introduce artefacts into the data. It is necessary to produce stimulus

slides with comparable brightness and contrast. A form of baseline can be achieved by letting each actual stimulus image be preceded for a period of 2–5 seconds with a slide with the same luminosity, either a homogeneous tint or a randomly scrambled version of the pixels in the following actual stimulus. In human factors studies with participants in situations with a natural variation in luminance, for instance aeroplane pilots, or when using web pages for stimuli, not to mention authentic advertisement videos, this is particularly difficult and data may have to be abandoned when luminance is not constant (Dehais, Causse, & Pastor, 2008).

Absolute pupil diameter is highly idiosyncratic; the correlation value, $r \in [0.811, 0.944]$, between the four conditions in the data of case study 2 on page 5 illustrate this. All pupil dilation values of a participant should be compared against an established baseline formed by fixating a blank screen for a longer period of time. Beatty and Lucero-Wagoner (2000) point out that normalized pupil diameter in % is inflated when the baseline pupil diameter is small, and recommend the absolute difference measure in millimetres.

Algorithms that compensate for changes in luminance have appeared, which analyse the variation in light on the stimulus monitor, using wavelets (Marshall, 2007) and principal component analysis (Oliveira, Aula, & Russell, 2009). Pupil dilation depends more on the light absorbed by the fovea, than on light hitting the peripheral parts of the retina (Piccoli, Soci, Zambelli, & Pisaniello, 2004), which could be used for further compensation.

Interpretations of the pupil diameter

The measure *can* be used to study a variety of cognitive and emotional states; note however that some commercial eye-tracking providers heavily over-estimate this possibility, and underestimate the difficulties involved (Bartels, 2009). Changes in pupil dilation are triggered by a variety of factors, which calls for a tight experimental design, if you want to make certain that the effect in pupil dilation is caused by one specific factor.

Mental workload increases pupil diameter Hess and Polt (1964) concluded that “changes in pupil size during the solving of simple multiplication problems can be used as a direct measure of mental activity”. Pupil dilation increased about twice as much (22 versus 11 per cent) when participants calculated 16 times 23, compared to 7 times 8. This general finding was replicated by Ahern and Beatty (1979), who found diameter changes of 0.1–0.5 mm (also page 436). Hyönä, Tommola, and Alaja (1995) showed that pupil diameters for three different types of translating vary as a function of the level of effort (4.20 mm – listening; 4.72 mm – shadowing; 5.22 – interpreting). Just and Carpenter (1993) found that sentences of varying syntactic complexity gave different pupil diameters when read. Kahneman and Beatty (1966) found larger pupillary responses when participants memorized more digits (0.1 mm versus 0.55 mm for 3 versus 7 digits). In the human factors field, pupil dilation is one in a family of measures used to examine mental workload and cognitive processing. Pupil diameter is often combined with blink rate and duration, fixation durations, saccadic extent, fixation rate, and dwell time, to estimate the cognitive requirements of different tasks (Brookings, Wilson, & Swain, 1996; Van Orden *et al.*, 2000, 2001). Although effects have been unclear when averaging over whole trials, data for the various phases of a task are clear: While blink rate, blink duration, and fixation duration all tend to decline as a function of increased workload (Van Orden *et al.*, 2000; Veltman & Gaillard, 1998), pupil dilation instead increases (Iqbal, Zheng, & Bailey, 2004; Van Orden *et al.*, 2000). Van Gerven, Paas, Van Merriënboer, and Schmidt (2002) found that mean pupil dilation is a useful event-related measure of cognitive load in research on education and learning, especially for young adults. Lying increases the pupil dilation, and attempts have been made to use the pupillary response as a lie-detector (Janisse & Bradley, 1980; Lubow

& Fein, 1996). The question remains, however, whether the pupil increases because of lying or because of a higher cognitive workload or stronger emotions in that measure situation. Dionisio, Granholm, Hillix, and Perrine (2001) concluded that “try[ing] to make their lies as believable as possible” was a more cognitively demanding task than truth-telling.

Emotion and anticipation increase pupil diameter Emotional and sexual arousal increase pupil dilation of the viewing participants in both males and females (Hess & Polt, 1960; Aboyoun & Dabbs, 1998). Partala and Surakka (2003) found larger average pupil diameter when participants listened to affect sounds, such as baby laughing and baby crying, compared to neutral sounds (office noise). Females responded more strongly to positive sounds, and males more strongly to negative ones. When anticipating to see answers to trivia questions that the participant reports being curious about, pupil dilation is larger if the participant was more curious about the answer (Kang *et al.*, 2009). We also appear to react to pupil sizes of others. For instance, pictures of women are rated as more attractive, by post-pubertal males, when their pupils are larger; this does not hold when women make the ratings (Bull & Shead, 1979). Of course, this was known centuries ago by women who used extracts from the highly toxic plant belladonna (meaning ‘beautiful lady’) to enlarge their pupils and increase their attractiveness. Harrison, Wilson, and Critchley (2007) show that a diminished pupil in faces causes participants who watch those faces to judge them as sadder, although not as expressing more fear, surprise, or disgust; also, diminished pupil promotes more empathy towards the faces. Not only has pupil size been found to be associated with emotional judgment, it is also a social signal that influences the pupil size of others—termed ‘pupil dilation mirroring’ or ‘pupillary contagion’ (Harrison, Singer, Rotshtein, Dolan, & Critchley, 2006).

Drowsiness and fatigue decrease pupil diameter This effect was found by Lowenstein and Lowenfeld (1962) and Yoss, Moyer, and Hollenhorst (1970), but not by Beatty (1982), who all used visual and auditory vigilance tasks. It is likely that the studies varied with respect to participant workload.

Diabetes decreases pupil diameter Patients with diabetes tend to have small pupil size, possibly because their pupillary sympathetic pathway is affected (Cahill, Eustace, & de Jesus, 2001).

Age decreases pupil diameter The resting pupil diameter was found to be smaller in the elderly group (mean age 69) at all three illumination levels, compared to younger (mean age 19) (Bitsios *et al.*, 1996).

Pain increases pupil diameter Chapman, Oka, Bradshaw, Jacobson, and Donaldson (1999) found that peak dilation increased significantly as pain intensity increased. Female participants show a greater increase at higher pain levels (Ellermeier & Westphal, 1995).

Drugs increase pupil diameter A large number of legal and illegal drugs increase pupil diameter, and pupil size is regularly used as a field indicator for drug intoxication.

11.6 Position data and confounding factors

A large variety of factors affect what positions in your stimuli your participants are likely to look at. As always, if you do not watch out for them, they may turn up as confounds in your experiment, but if you systematically utilize these factors, you may make new discoveries.

Table 11.1 Some participant factors that influence the positions they look at.

Factor	Likely effect	Sample reference
Alcohol	Missed event, tunnel vision	Buikhuisen and Jongman (1972)
Medication	Fixation dispersion, and saccade and smooth pursuit parameters	O'Driscoll and Callahan (2008)
Schizophrenia	Restricted dispersion	Loughland <i>et al.</i> (2002)
Autism	Eye and face avoidance	Klin, Jones, Schultz, Volkmar, and Cohen (2002).
Phobias	Avoidance	Pflugshaupt <i>et al.</i> (2007).
Eating disorder	More looks at their own unappealing body parts	Jansen, Nederkoorn, and Mulken (2005).
Obesity	Look more at food when fasted	Castellanos <i>et al.</i> (2009); Nijs, Muris, Euser, and Franken (2009).
Sexuality	Looks at body parts of either gender	Rupp and Wallen (2007); Tsujimura <i>et al.</i> (2009)

Take alcohol as an example. The drug influences many parts of the brain and causes participants to miss task-important events, reduces their functional visual field, and induces tunnel vision (Buikhuisen & Jongman, 1972). If you have no control over blood alcohol levels with your participants, you have a confound that may overturn your results. But you may also control alcohol levels, and make a systematic comparison between levels or to sober participants.

In this section, we briefly point out a number of possibly confounding factors for position results in eye-tracking-based research: the participant himself, and the drugs and medication he uses, his cultural background, the task given to him, and the experiences he has, as well as the central bias effect with monitor-based, and research on features of the stimulus itself.

11.6.1 Participant brainware and substances

Participants vary in their brainware and in what substances they consumed before arriving in your lab. This not only makes eye tracking very interesting in the study of clinical groups, it also makes medication a possible confounding factor in many studies. Table 11.1 lists a number of participant factors.

11.6.2 Participant cultural background

The participant's *cultural background* would appear to be another possible confounding factor for experiments that use position measures. Chua, Boland, and Nisbett (2005) showed that participants from an American culture tend to look more at focal objects, and participants from a Chinese culture more at the background, when both are shown the same pictures with a focal object and a complex background. However, this result could not be replicated by Rayner, Castelano, and Yang (2009); Evans, Rotello, Li, and Rayner (2009); Rayner, Li, *et al.* (2007).

However, Blais, Jack, Scheepers, Fiset, and Caldara (2008), and Miellet, Lingnan, Matthew, Rodger, and Caldara (2009) found that East Asians looking at faces look more at the nose than Western Caucasians, and that this was *not* due to a larger functional visual field. Also, McCarthy, Lee, Itakura, and Muir (2006) found that Canadians and Trinidadians who think about the answer to a question from an interlocutor tend to look up, while Japanese partici-

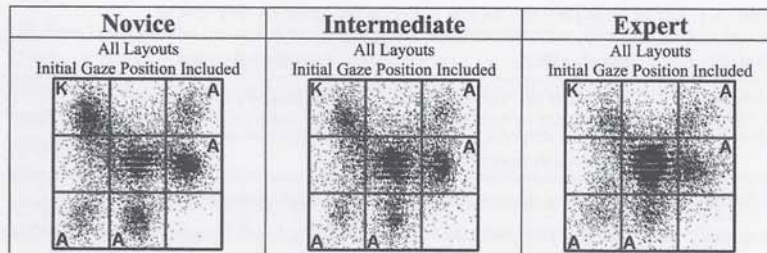


Fig. 11.20 Raw data sample from participants with varying levels of proficiency in chess, as they evaluate chess board with attackers (A) and a king (K). The figure suggests a lower dispersion of experts. Reprinted from Eyal M. Reingold, Neil Charness, Marc Pomplun, and Dave M. Stampe, *Psychological Science*, 12(1), copyright © 2001 by SAGE Publications. Reprinted by Permission of SAGE Publications.

pants look down. Canadians but not Japanese altered their gaze behaviour when knowing that they were observed (McCarthy, Lee, Itakura, & Muir, 2008).

11.6.3 Participant experience and anticipation

Another possibly confounding factor for position measures that may need to be controlled for is the participant's *experience* with the task. A host of studies show that expertise gives rise not only to a more task efficient selection of gaze positions, but also to a superior perceptual processing from a larger functional visual field, stretching further out from the fixation point, than it does for novices (although not overriding the physiological restrictions at the retina, visual pathway and visual cortex). Such effects of experience has been shown for painters (Vogt & Magnussen, 2007; Nodine *et al.*, 1993; Antes & Kristjanson, 1991), drivers (Mourant & Rockwell, 1972), during the diagnoses of electrical circuits (Van Gog, Paas, & Van Merriënboer, 2005), of dental and mammography X-rays (Van Der Stelt-Schouten, 1995; Krupinski, 1996), chess (Reingold *et al.*, 2001; Reingold & Charness, 2005) and basketball players (Memmert, 2006). As part of *training*, for instance in sign language or to drive, gaze positions gradually change (Emmorey, Thompson, & Colvin, 2008; Mourant & Rockwell, 1970; Mourant *et al.*, 1969).

Why is experience so important? Gaze is generally anticipatory, reflecting the participant's probabilistic model of the world. For instance, if you meet a pedestrian coming from the other direction, and you judge her to be collision-prone, you will look earlier and more at her than if you feel safe with that person (Jovancevic-Misic & Hayhoe, 2009). In ball sports, the experts look at the bounce point 100–200 ms before the ball reaches it (Land & McLeod, 2000; Ripoll, Fleurance, & Cazeneuve, 1987), giving them time to confirm their prognosis for the continued trajectory, which in turn they need for quick action. Expert goalkeepers facing a penalty shot look at the legs and face of the kicker (Savelsbergh *et al.*, 2002) as part of their preparation.

11.6.4 Communication, imagination, and problem solving

Not controlling *what is said to participants* is very likely to alter the position data you collect from them, as shown already by Yarbus (1967) in the example with the unexpected visitor. Both speaking and listening heavily influences the position of fixations (Holsanova, 2008; Griffin & Bock, 2000; Tanenhaus *et al.*, 1995).

Other people's gaze will also alter the positions to which your participants will look. In general, people are well aware of other people's gaze direction, and it affects where they

look themselves. In a series of studies on gesture perception, Gullberg and Holmqvist (2006, 1999) showed that speakers who fixate their own gestures attract significantly more gazes from the listener to these gestures, and this effect is used by magicians to govern the audience gazes (Kuhn & Tatler, 2005). Communication between two people becomes easier and mutual problem solving is quicker if they can see each other's gaze positions (Velichkovsky, Pomplun, Rieser, & Ritter, 1996).

An increased *mental workload during speech planning* or actively recalling a memory often causes participants to look away from their interlocutor, computer monitor etc. This is known as 'gaze aversion'. Doherty-Sneddon, Bruce, Bonner, Longbotham, and Doyle (2002) showed that 5-year old school children inconsistently avert gaze, but that it increases dramatically during the first years of primary education, reaching adult levels by 8-years of age. Glenberg, Schroeder, and Robertson (1998) showed that the amount of time spent looking away from an interlocutor increases with task difficulty, and that participants gave more correct answers when averting their gaze. Together, this indicates that gaze aversion may be functional.

Speech affects eye movement even if objects spoken about and looked at just represent the mental images produced by the participants (Polunin, Holmqvist, & Johansson, 2008; Loetscher, Bockisch, & Brugger, 2007; Zangemeister & Liman, 2007; Johansson *et al.*, 2006; Laeng & Teodorescu, 2002; Brandt & Stark, 1997). When solving geometrical and graphical problems, the ability to imagine the functional role of part objects in the solution depends on where participants look, as shown by Grant and Spivey (2003); Yoon and Narayanan (2004b). In a study on ambiguous pictures, including the Necker cube, Pomplun *et al.* (1996) show that gaze position coincides with the interpretations participants subjectively experience.

11.6.5 Central bias

The vast majority of eye-tracking research, even if it purports to generalize to all visual activities, is made using stimuli presented on single monitors. With most stimuli—text being the large exception—participants show a marked tendency to fixate the centre of the screen more than any other part. Tatler (2007) attributes this observation to one of three possibilities:

First, the center of the screen may be an optimal location for early information processing of the scene. Second, it may simply be that the center of the screen is a convenient location from which to start oculomotor exploration of the scene. Third, it may be that the central bias reflects a tendency to re-center the eye in its orbit.

Fehd and Seiffert (2010) point out that looking steadily at the centre of scenes allows a participant to keep a better overview of the multiple objects in it compared to looking around. Interestingly, monkeys appear to have less central bias (Berg *et al.*, 2009), which suggests that specific human expectations and experiences cause it.

Researchers unaware of the central bias effect may erroneously generalize a result from monitors to real-life behaviour.

11.6.6 The stimulus

Elements in the stimulus itself may inadvertently attract participant gaze and become a confound in studies. For instance, accidentally having people or faces on a stimulus picture where the research question refers to the other visual elements (in a park, for instance) is very likely to alter results, because people and faces attract participant gaze. The 'visual search' paradigm has attempted to systematically investigate what other so-called 'bottom-up features' (colour, motion, orientation, for instance) in a picture attract visual attention, or in other words make an object 'pop out'. Wolfe and Horowitz (2004) review a large number of

visual features that may or may not attract participant attention.

Some eye-tracking studies have indeed given support to the idea that bottom-up features are important to human choices of gaze positions. Reinagel and Zador (1999) and Parkhurst and Niebur (2003) report a higher luminance contrast around fixation than in randomly chosen regions, while Baddeley and Tatler (2006) conclude that high-frequency edges are good predictors of fixation locations.

The Itti and Koch (2000) saliency model is a hotly debated computational model of some of the assumed neurological principles that employ bottom-up (pre-attentive) features to select fixation targets. Implementations are available on several web pages and in some of the dedicated analysis softwares for eye-movement data. Competing but less known models include the target acquisition model (TAM) (Zelinsky, 2008), the gaze-attentive fixation finding engine (GAFFE) (Rajashekar *et al.*, 2008), the contextual guidance model (Torralba, Oliva, Castelhana, & Henderson, 2006), and the neurodynamical cortical model (Deco & Rolls, 2004). None of these are as well investigated as the saliency model, but Nyström (2008) shows that the GAFFE model does somewhat better than the saliency model in comparison to human data. Itti (2005) demonstrates a better than chance similarity between human eye-movement data and the output of the saliency model, however.

Models can be seen as useful and necessary steps in the evolution of our scientific understanding of what people look at, but the cost for false negatives could be very large in some specific applications. For instance, radiologists scanning X-ray images appear to use other features compared to those emphasized in the visual search paradigm (Krupinski, Berger, Dallas, & Roehrig, 2003).

Also, the dominance of bottom-up features is not supported by Chen and Zelinsky (2006) who show that top-down guidance of eye movements in a search task always prevails over bottom-up saliency features (colour coding of singular elements among grey-scale objects). Similarly, by blurring some parts of stimulus pictures and increasing contrast in others, Einhäuser *et al.* (2008, 15) show that "a visual search task can override and actively countermand sensory-driven saliency in naturalistic visual stimuli". Also Henderson, Brockmole, Castelhana, and Mack (2007) argue that Itti's bottom-up models of saliency do not account for human eye movements, while others have shown that central (Tatler, 2007) and oculomotor biases (Tatler & Vincent, 2009) can explain eye-movement data better than the saliency model.