

8 Scanpaths—Theoretical Principles and Practical Application

Previous chapters explained how to process raw data samples into fixations, saccades, and smooth pursuit (Chapter 5), how to divide stimulus space into areas of interests (AOIs) (Chapter 6) that give rise to events such as dwells and transitions, and also how to build representations of overall spatial distribution called attention maps and their visualizations such as heat maps (Chapter 7). In this chapter, we discuss the *scanpath*—a trace of a participant's eye-movements in space and time—and its events and representations.

Using and analysing scanpaths raises many questions, some very practical, and some deeply abstract. This chapter consists of the following sections, and we suggest selective reading:

- The first Section 8.1 (p. 253) presents a formal definition of scanpaths and relates it to visualizations and representations of scanpaths.
- Section 8.2 (p. 255) provides condensed hands-on advice for research with scanpaths.
- In Section 8.3 (p. 256), we present the most common usages of scanpaths.
- Common events that occur in scanpaths are defined in Section 8.4 (p. 262): The *back-track*, the *regressions*, the *look-back*, the *look-ahead*, *local versus global* scans, and the *reading versus scanning* events.
- Section 8.5 (p. 268) describes by what means a scanpath may be represented: *strings*, *Euclidean vectors*, and sequences of *attention maps*.
- Scanpath representations are typically used for the purpose of comparing two or more scanpaths. The principles for scanpath comparison are outlined in Section 8.6 (p. 273).
- Section 8.7 (p. 278) discusses whether scanpaths can be related to specific cognitive processes. It addresses scanpath theory and the related role of memory and task to scanpath planning and inhibition of return. The section further discusses the average scanpath and the challenging issue of how to develop and evaluate better scanpath comparison methods. This section ends with open issues in scanpath comparison.
- Finally, Section 8.8 (p. 284) summarizes the chapter and the scanpath events and representations that we will use throughout the rest of the book.

8.1 What is a scanpath?

The term 'scanpath' originates from the work by Noton and Stark in the early 1970s (Noton & Stark, 1971a, 1971b). Other common terms for scanpaths are 'scan pattern', 'search pattern', 'scan sequence', 'gaze sequence', 'fixation track', 'inspection pattern', and 'eye-movement pattern'. About 70% of the journal papers and 84% of Google hits write scanpath as a single word ('scanpath'), and the rest use separate words ('scan path').

Noton and Stark's scanpath term refers to the fairly abstract concept of a fixed path that is characteristic to a specific participant and viewing pattern. In contrast, the term scanpath is today used to very concretely describe how the eye physically moves through space, typically

but not exclusively for one participant. In agreement with the physical definition of other common terms in this book, we define a scanpath as the *route of oculomotor events through space within a certain timespan*. This assumes that the ‘path’ has a beginning and end, and therefore a length.

The most accurate estimate of a scanpath that an eye-tracker can provide is the spatial coordinates of a participant’s gaze on a stimulus taken every $1/F_s$ second, where F_s is the sampling frequency of the eye-tracker. Space is usually confined to two dimensions x and y , and a scanpath of length L can then be described by a sequence of coordinates $S_i = (x_i, y_i), i = 1, 2, \dots, L$. This is what we can see if we plot a raw data sample scanpath.

A *scanpath function* $f(\cdot)$ is given one or many scanpaths as input and computes a representation of those scanpaths. For example, one output from f would be a sequence of attention maps which can then be seen as a Gaussian-based scanpath function, which encodes the probability that a certain spatial position will be a part of the scanpath. Another output would be vectors representing the fixations and saccades of which a scanpath is comprised. Figure 8.1 shows four different scanpath representations of the same data.

The most common previous definitions of scanpaths have followed Noton and Stark (1971a) in saying that scanpaths consist of a sequence of *saccades*. *Fixation positions* are part of several scanpath definitions, but *fixation durations* are seldom utilized in any of the measures of scanpaths. *Smooth pursuit* is completely absent from most researchers’ working definition of what a scanpath is, with rare exceptions, such as Boccignone, Caggiano, Marcelli, Napoletano, and Di Fiore (2005), who conduct an analysis of eye-tracking data on video stimuli.

The minimum requirement of a scanpath representation is that it is a sequence that *takes ordinal information into account*. This means that any representation of a specific scanpath must transform into a representation of *another* scanpath whenever the order between elements in the representational sequence is changed.

Static and dynamic visualizations

Recorded scanpaths are typically projected onto the stimulus or an empty space representing the stimulus. Visualizations of data that are not projected onto a 2- or 3D stimulus space, for instance a space–time diagram of (x, y) -coordinates, are not generally considered to be scanpath visualizations.

Most analysis softwares offer scanpath visualizations in a number of varieties, and the possibility to export or print them. Scanpath visualizations are either *static*, of which previous chapters have had many examples, or *dynamic*, which comes out well on computers but not on paper. Both these allow for direct inspection of the data from a single participant and a single trial. Most software packages support viewing four types of static visualizations: raw data sample scanpaths, which depict the entire set of raw (x, y) coordinates; fixation-based scanpaths, with fixations plotted either with or without circles of different size to indicate their duration, and with the option to print sequence numbers next to fixations.

In the static visualizations, the dynamic aspect of the scanpath is supported by connection lines and fixation numbering. Without these, the static scanpath visualization is reduced to a density plot: the same set of unconnected points that underly the attention maps of Chapter 7. The dynamic visualizations in the software additionally emphasize the sequential order of scanpaths in a variety of ways depending on the manufacturer. The basic dynamic scanpath is a single gaze cursor that is played back. This type of visualization is often used to replay participants’ eye-movement data in order to elicit verbal data from them (pp. 99–108). Moreover, this type of visualization may also be depicted for many participants at a time. Scanpaths of multiple participants typically become very cluttered, except in dynamic gaze replay, whereby multiple gaze cursors are played back against a stimulus.

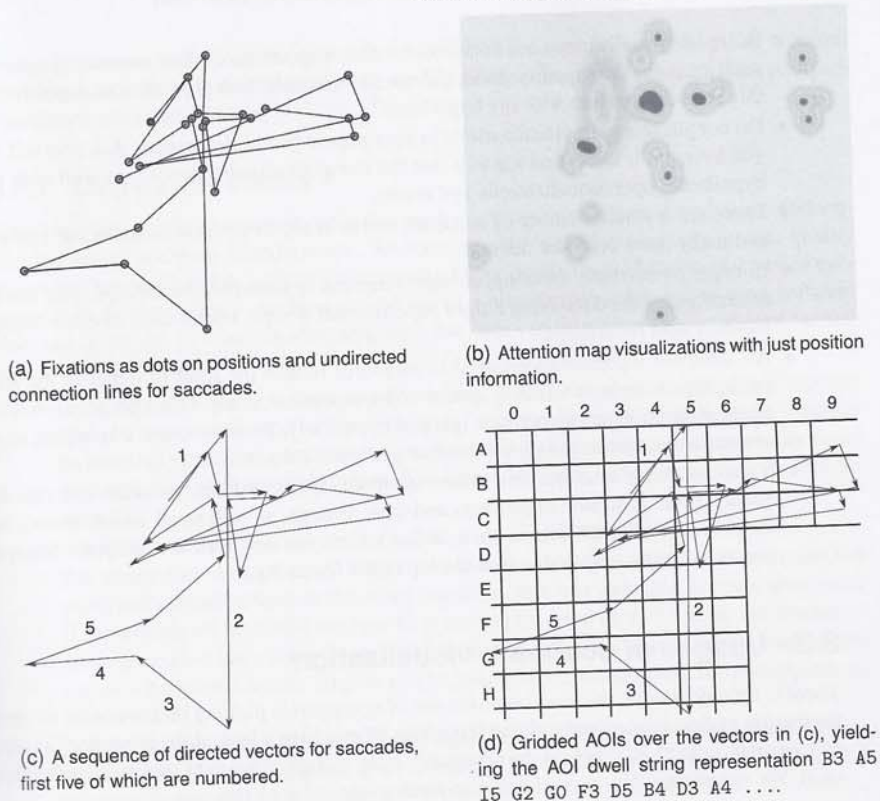


Fig. 8.1 Visualization of the four different scanpath representations of the same data.

Scanpath representations

The three scanpath representations that are more than sole visualizations are all designed to cope with what visualizations are poor at; namely providing a representation of data that can be used for computational analysis and statistics. The AOI strings—in particular the dwell string—are the most common non-visualizing scanpath representation, but sequences of Euclidean vectors and attention maps are also beginning to be used.

Besides visualization, the commercial analysis software packages of 2010 are only just beginning to add functionality for scanpath analysis. If you want to use the existing scanpath measures, you therefore need to export data such as raw data samples, fixations, saccades, and/or AOIs, and then implement the measures yourself. The alternative is to use non-commercial applications (e.g. West, Haake, Rozanski, & Karn, 2006; Tsai, 2010; Foulsham, 2010; Cristino, Mathôt, Theeuwes, & Gilchrist, 2010).

8.2 Hands-on advice for using scanpaths

Scanpaths are used so variably that only a few general pieces of advice can be offered. If you plan to use scanpaths as visualizations, or calculate statistics from scanpath representations you should consider the following issues:

- Scanpath visualizations are excellent for first inspections of data, answering questions such as: is the data quality good, did the fixation detection algorithm do a good job, is this recording in line with my hypothesis?
- Do not put scanpath visualizations in your papers just as decoration. Ask yourself why you have put it there, and see to it that the scanpath visualization aligns well with your hypothesis, operationalizations and results.
- There are a whole number of scanpath events ready to be used in statistical analyses, and many more could be defined.
- In order to attribute meaningful interpretations to individual scanpaths, you need to disambiguate the data using a tight experimental design, verbal data, or other complementary data recordings.
- All scanpath representations used in measures reduce the level of detail in the scanpaths, for example in terms of spatial and temporal accuracy. Other properties such as fixation duration are sometimes ignored completely. Be sure to use a scanpath representation that retains the properties that you want to measure.
- If you are using measures that utilize scanpath representations, be aware that raw data quality, event detection algorithms and their settings, as well as all issues around AOI identification may introduce noise in the values you get from the measure. Scanpath events and representations are at the top of the hierarchy.

8.3 Usages of scanpath visualization

There is little doubt that the most common use of scanpaths is plotting them in order to check the quality of data immediately after a recording. “Let’s have a look at the recording” is often synonymous with “Let’s look at the scanpath”, even though other data quality visualizations exist. We can recapitulate several down-to-earth usages of scanpath *visualizations* as:

Data quality checks In Chapters 4 and 5, we made ample use of scanpath visualizations for this purpose. It shows you if there are offsets or poor precision in the data, and whether fixation and saccade detection worked properly.

Preliminary impression of the data Often a scanpath is visualized to receive a quick first impression of where the participants looked, and in which order. During piloting, such inspection can be used to check whether the task elicits the desired eye-movement behaviour, and to give a first impression of whether your hypothesis will be supported.

Offset compensation If your data has an offset, some software allows you to perform manual offset compensation and drift correction watching scanpath visualizations, as described on page 224.

Manual data analysis Scanpath analysis by visual inspection is hopefully decreasing as methods and software become more capable, but has been the main form of analysis in previous years (e.g. Josephson & Holmes, 2002; Holsanova, 2001; Buswell, 1935).

Illustrating scanpaths in publications There are a variety of reasons why scanpaths are shown, from showing off some good data on the background of a stimulus, to using the scanpath visualization to clearly demonstrate the experiment and/or the analysis.

Cued retrospective thinking aloud Participant’s thoughts may be difficult to access through eye-movement data and tight experimental designs alone. It has become increasingly common to show the scanpath to a participant just after his data has been recorded, and ask him to retell what he was thinking of during the initial inspection of the stimulus. This method is described in-depth on pages 99–108.

In the next sections, we present and discuss three of the most common usages of scanpath visualizations: checking the data quality, analysing the data manually, and exhibiting scanpath visualizations in publications.

8.3.1 Data quality checks

Checking data quality is undoubtedly the most common, yet informal and quite undocumented use of scanpath visualizations. We have already looked extensively at data quality issues in Chapters 2 and 4, and only reiterate the major points here. Visual inspection of eye-movement data for quality checking should not be a problematic issue, unless your software does not allow you to visualize scanpaths with raw data samples. Data quality checks can tell us a lot about the data:

- Is the data accurate or is there an *offset* in any part of the stimulus image? Offsets can often be seen as mismatches between actual data and semantic entities in the stimulus picture that were likely fixation targets given the task. Accuracy tests are easier to do with text than with general images. The upper scanpath in Figure 8.2(a) provides an example.
- Are there many *optic artefacts* in the data, as in Figure 8.2(a)?
- Are many *data samples lost*? This could indicate that the eye image is poor and the pupil and corneal reflection therefore cannot be properly detected. It could also occur if the participant is closing his eyes or is turning his head away from the eye-tracker.
- Is the *precision* low (noise levels high) in the recording, as seen in the spread of raw samples contained within fixations? For example the fixations of six participants in Figure 8.2(b).
- Is the event calculation algorithm doing a proper job? Plot the raw sample scanpath next to the fixation-based scanpath, and compare. Figures 8.2(c) and 8.2(d) from Chapter 5 provide an example where fixations were lost during fixation and saccade calculation.

8.3.2 Data analysis by visual inspection

There are many reasons why researchers would want to use scanpath visualizations for their data analysis.

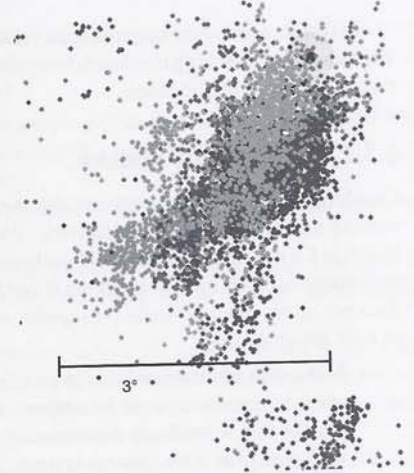
First of all, a manual, participant-by-participant analysis may be what is needed for *pedagogical reasons* in the publication, for instance in Buswell (1935), who starts his book with a long commentary to the scanpath visualization from the data of the participant called “Miss W.”, focusing on the order and position of individual fixations, and on what has not been fixated. This manual analysis and the description of the scanpath serve a particular pedagogic purpose that quantitative analyses could not easily provide.

In other cases the *software is inadequate*. For instance, Buswell (1935) reports statistics using what we today call gridded AOIs, constructed by manual analysis from scanpath visualizations. It took many decades until software supported such AOI analyses.

Even fairly recently, researchers have had to retreat to manual analysis because *statistical tools are inadequate*. For instance, Josephson and Holmes (2002) give a thorough description of the string-edit method (p. 348), and additional methods that can be used to further a string-edit analysis. They conclude, however, that statistical results based on string-edit calculation are currently not possible, and end up eyeballing their data. Similarly, Tzanidou, Minocha, and Petre (2005) discuss the problem of identifying metrics that compare scanpaths across many participants and many stimuli, concluding that there is no such measure, and decid-

The Marketer: Alchemist, Magician, Sorcerer and Medicine Man
 Is marketing a part of a science? Perhaps marketing is more like
 sorcery. Think of a sorcerer collecting ingredients from different sources
 and mixing them into a potion accompanied with the magical effect of a
 flash of light and the illusion to follow. To some extent this fits with
 Culture's vision of a marketer as a mixer of ingredients. Of course
 sorcery and magic are mythical and the real world of marketing
 may help to dispel some of the myths surrounding marketing.

The Marketer: Alchemist, Magician, Sorcerer and Medicine Man
 Is marketing a part of a science? Perhaps marketing is more like
 sorcery. Think of a sorcerer collecting ingredients from different sources
 and mixing them into a potion accompanied with the magical effect of a
 flash of light and the illusion to follow. To some extent this fits with
 Culture's vision of a marketer as a mixer of ingredients. Of course
 sorcery and magic are mythical and the real world of marketing
 may help to dispel some of the myths surrounding marketing.



(a) Above a scanpath with corner offset due to problems during calibration; and below extreme optic artefacts due to mascara.

(b) Noisy recording showing raw samples (scanpath without connection lines) from six participants fixating a corner point on the stimulus monitor. Notice that the noise is large and oblique.

The Marketer: Alchemist, Magician, Sorcerer and Medicine Man
 Is marketing a part of a science? Perhaps marketing is more like
 sorcery. Think of a sorcerer collecting ingredients from different sources
 and mixing them into a potion accompanied with the magical effect of a
 flash of light and the illusion to follow. To some extent this fits with
 Culture's vision of a marketer as a mixer of ingredients. Of course
 sorcery and magic are mythical and the real world of marketing
 may help to dispel some of the myths surrounding marketing.

The Marketer: Alchemist, Magician, Sorcerer and Medicine Man
 Is marketing a part of a science? Perhaps marketing is more like
 sorcery. Think of a sorcerer collecting ingredients from different sources
 and mixing them into a potion accompanied with the magical effect of a
 flash of light and the illusion to follow. To some extent this fits with
 Culture's vision of a marketer as a mixer of ingredients. Of course
 sorcery and magic are mythical and the real world of marketing
 may help to dispel some of the myths surrounding marketing.

(c) Raw data samples. Fixations are seen as black blobs.

(d) After fixation and saccade analysis, some fixations have disappeared.

Fig. 8.2 Data quality checks using scanpaths.

ing to analyse their own scanpath data over web pages using visual inspection and manual categorization.

Arguing that the types of analysis provided by the academic fields of eye tracking are not useful for usability research, Ehmke and Wilson (2007) set out to find types of scanpaths that coincide with a specific cognitive process, but since so little is known about scanpaths and possibly meaningful subscans, manual analysis of scanpaths against retrospective interviews—that is verbal data—appears to be the only way forward.

As Holsanova (2001, 2006, 2008) studied the eye-movements of participants freely describing pictures, she examined patterns in scanpaths and speech that jointly indicated specific cognitive processes, using a manual transcription method to finds alignment between sub-scans and spoken items. More precisely, Holsanova (2001) built what are known as multimodal score-sheets, in which several tiers of temporal data share one common timeline. As shown in Figure 8.3, one tier was the sequence of dwells in AOIs, while another tier listed the

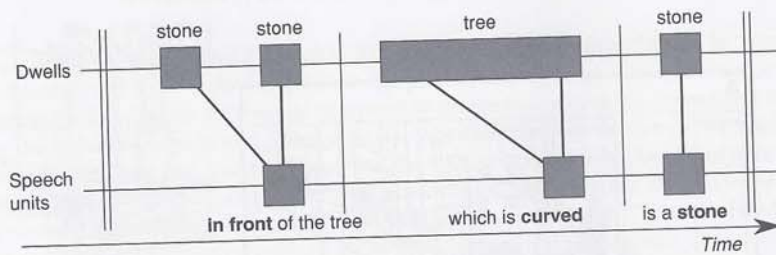


Fig. 8.3 Multimodal score-sheet resulting from manual co-analysis of synchronized speech and eye-movement data. The final dwell on the 'stone' AOI is a lookback, coinciding with the participant naming the stone, and thus most likely due to the planning and development of speech. Adapted from Holsanova (2001).

development of speech units over time. As this was a picture description task, gaze travelled across the same AOIs that speech referred to, and so the multimodal score-sheet indicated patterns of temporal alignment of speech to gaze.

Using an adapted variety of the method in Holsanova (2001), Johansson *et al.* (2006) compared scanpaths during scene perception with scanpaths during subsequent imagery, and needed to take into account shrinking and repositioning of scanpaths, as well as synchronization to speech. Again, this could only be done manually.

Land, Mennie, and Rusted (1999) recorded scene videos with overlaid gaze cursors of participants making tea. Their analysis of task behaviour led to a multimodal score-sheet similar to Figure 8.3, but instead with tiers representing actions rather than speech. This video-based analysis could only be made manually.

How should a manual analysis be carried out? In all the above cases, manual scanpath analysis aims at finding the sequence of fixations or dwells in AOIs, that is a list of the objects looked at, with information about when and for how long gaze stayed there. Today many softwares can output this list, but not when data consist of gaze-overlaid scene videos, and not always with dynamic stimuli or in mental imagery studies where AOIs are difficult to define. Since in these cases the scanpath needs to be played back, the coding methods for gaze-overlaid scene videos in Chapter 6 (p. 227) can be utilized, with small adaptations to take advantage of the more advanced playback control features in scanpath visualization software.

8.3.3 Exhibiting scanpaths in publications

All quantitative results in Buswell (1935) and Yarbus (1967) were based on manual analysis from scanpath visualizations. Both publications make extensive use of scanpath illustrations, both for illustration of data and presentation of actual results.

Many authors still publish selected scanpath visualizations in their papers and books, for a variety of reasons. First, some publications simply address an audience new to eye-movement data and the scanpath is needed to explain what eye-movement data are. Second, when the presented analysis includes an important *qualitative description* of the scanpath (Buswell, 1935; Holsanova, 2008), a scanpath visualization is obviously needed. *Difficulties in operationalizing the concept* under study in terms of computational measures could be another major reason why we find scanpath visualizations in research publications. Methodology papers and papers that discuss measures, such as Goldberg and Kotval (1999), Rötting (2001), and Underwood, Humphrey, and Foulsham (2008a) present many scanpath visualizations to illustrate concepts and operationalizations, and the same is true in this book.

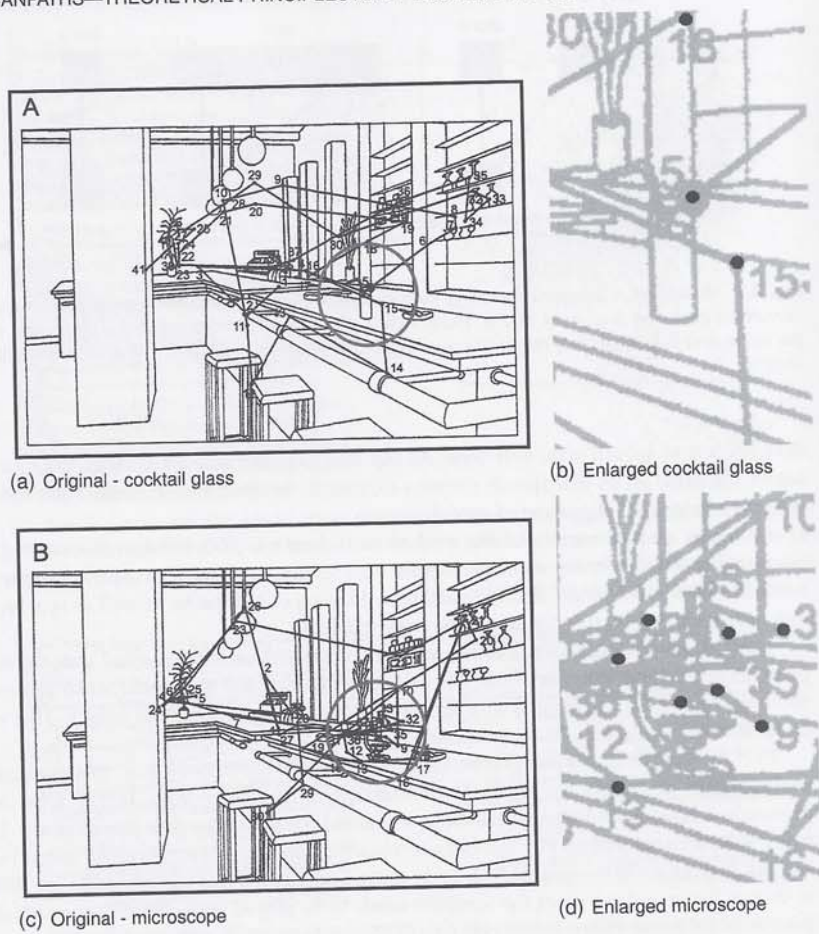


Fig. 8.4 This “typical viewing pattern” in fact proves to be an outlier. The circle in (a) and (c) denotes the cocktail glass and microscope respectively, which are then blown up with dimmed background in (b) and (d) to indicate fixation clustering (note, this is a visualization only, and fixation dots in (b) and (d) may not reflect the exact number of fixations in the recorded data). Reprinted from *Journal of Experimental Psychology: Human Perception and Performance*, 25(1), John M. Henderson, Phillip A. Weeks, and Andrew Hollingworth, The Effects of Semantic Consistency on Eye Movements During Complex Scene Viewing, pp. 210-228, Copyright (1999), with permission from Elsevier.

In other cases, when the audience of the paper knows about eye tracking, when the operationalizations are clear and quantitative results do not rely on visualizations, why then use scanpaths for illustrations?

Our first example is taken from Henderson, Weeks, and Hollingworth (1999). It has been selected because it is a well-known paper with many citations, but also because the scanpath visualization the authors choose to include in the paper is in line with their hypotheses, but deviates from the statistical results. More precisely, they present two “typical viewing patterns” (stated in the last paragraph of page 213 in their article) over almost identical bar room scenes, but where a cocktail in one is replaced by a microscope in the other. Their Figure

2, here reproduced as Figure 8.4 depicts this. The concept under investigation is 'semantic consistency', and it is a natural hypothesis that participants will look earlier, more often and longer at the inconsistent microscope compared to the consistent cocktail. Indeed, this is also what the included scanpath visualization shows. When counting the fixations in the scanpaths of their Figure 2, we find 7 or 8 fixations near the microscope, but only one or two next to the cocktail glass. However, a scanpath visualization is not a result. For one thing, a scanpath is descriptive data from one single participant, usually in one single trial. Therefore, Henderson *et al.* (1999) make a statistical analysis of the data, over several participants, using a number of AOI measures, including number of fixations on the target, number of entries, first-pass and second-pass dwell time, and several others.

After we have seen the impressive example in the scanpath visualization of the authors' Figure 2, the reported overall difference in number of fixations is somewhat disappointing: 9.7 versus 10.7, which was found to be a non-significant difference. The large difference in the scanpath visualization from this particular participant is not even close to the reported nonsignificant 9.7 versus 10.7 averages. In fact, this "typical viewing pattern" is an outlier.

Maybe scanpath visualizations are often superfluous additions to the results presented in current-day journal papers. They may distract some readers to look for confirmation of the hypothesis, and even to start counting fixations in the single scanpaths visualization, rather than to look at the overall data analysis and the quantitative results. The authors appear to add these scanpaths with unclear but probably quite different intent, for instance proof that data were actually recorded and not just made up, or that data quality—as judged from the included scanpath—was good, or perhaps just as decoration or in tribute to Buswell (1935) and Yarbus (1967).

In our second example, however, we should trust the published scanpath illustration rather than the given quantitative data. In schizophrenia research, there has been much discussion about 'restricted' scanpaths to facial stimuli, which have consistently been reported in studies of schizophrenia patients (for instance, Green, Waldron, Simpson, & Coltheart, 2008; Green, 2006; Benson, Leonards, Lothian, St Clair, & Merlo, 2007; Loughland, Williams, & Gordon, 2002, just to mention a few). Similarly, participants with social phobia tend to exhibit 'hyperscanning' when watching faces, (Horley, Williams, Gonsalvez, & Gordon, 2004). 'Restricted' and 'hyperscanning' are definitely characteristics relating to scanpaths. Figure 8.5 shows examples with such (fabricated) data.

Operationalizing 'semantic (in)consistency' with reference to 'looking earlier, more, and longer', as did Henderson *et al.* (1999) in our first example, is relatively uncomplicated. There is a direct connection between the concepts used in their hypothesis, and conclusion, and the measures used. There is one AOI only, and the measures in use are simple and well-known, and nothing is really said about the general shape of the scanpaths that motivates a scanpath visualization in the paper. In contrast, operationalizing 'restricted scanpaths' puts much larger demands on data representations and measures, and eye-tracking-based papers in schizophrenia and social phobia research are consequently richly decorated with scanpath visualizations illustrating restrictedness and hyperscanning.

Green *et al.* (2008) and other schizophrenia researchers attempt to construct quantitative tests that approximate a 'scanpath restrictedness' measure, using a combination of two measures:

1. *Scanpath length* (p. 319) should be shorter for more restricted scanpaths in almost all cases. There are unfortunate exceptions however. Many short fixations at the circumference around the nose could add up to quite a long scanpath, while a small number of long fixations at the eyes and mouth could give a shorter scanpath length.

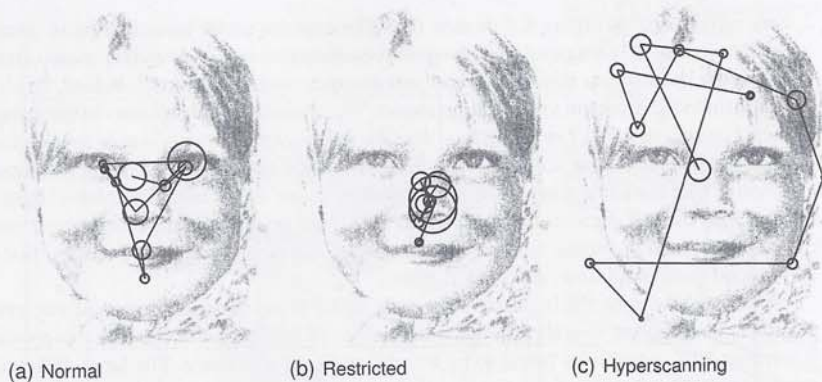


Fig. 8.5 Examples of normal, restricted and hyperscanning scanpaths on the well known face image used by Yarbus (1967), here faded to increase scanpath visibility. The scanpaths are fabricated for the purpose of illustration.

- Therefore, the authors additionally compared *number of fixations* (p. 412). In particular, if the number of fixations were equal and the scanpath lengths still differ, the probability increases that the shorter scanpath would indeed be restricted (in the sense of the presented scanpath visualizations).

Operationalizing the ‘restrictedness’ concept using the combination of scanpath length and number of fixations is not intuitive, but requires careful mathematical thinking of why this operationalization could work. The scanpath visualizations are included to explain what ‘restrictedness’ really is, and in effect define the concept ostensively, rather than through the measures.

Generally, if it is difficult to make the operationalization of the study intuitive with the use of scanpaths and their visualization, then it is advisable to shop around for other measures in collections such as Part III of this book. For instance, the variability measure known as convex hull (p. 364), the Kullback–Leibler similarity measure (p. 376) and related measures on pages 359–376, could have provided the authors with a single measure that would be closely aligned with the restrictedness concept.

8.4 Scanpath events

Just as fixations and saccades are eye-movement events that can be detected from raw data samples, *scanpath events* are temporally restricted patterns that occur in eye-movement sequences. Typically, scanpath events comprise subscans of length two or more in a sequence of fixations, saccades, or other events (such as smooth pursuit). A scanpath event can be associated with any type of form or pattern, and therefore an almost unlimited number of scanpath events may exist. Reading researchers, and some usability researchers have defined the few scanpath events that have been used in published research, and these are the ones discussed here.

8.4.1 The backtrack

A *backtrack* is the specific relationship between two subsequent saccades where the second goes in the opposite direction of the first.

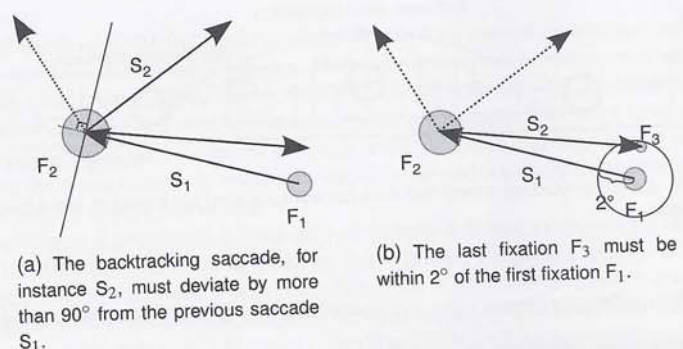


Fig. 8.6 Two definitions of backtracking. Dashed lines indicate saccades that are *not* counted as backtracks. In (a) the definition by Goldberg and Kotval (1999), and in (b), the one by Renshaw *et al.* (2004).

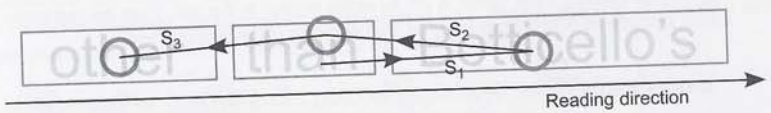


Fig. 8.7 Backtracking versus regressions: Saccade S_2 is a regression, because it moves in the opposite direction to the text, and also a backtrack because it moves in the opposite direction to the previous saccade. Saccade S_3 is only a regression, but not a backtrack according to Murray and Kennedy (1988). If a fourth saccade S_4 continued backwards, S_2 – S_4 would count as backtracks according to Murray and Kennedy (1988).

There are two operationalizations of backtracking saccades outside of reading research. The original definition by Goldberg and Kotval (1999) counted all saccades deviating more than 90° from the previous, making many saccades backtracks. Renshaw, Finlay, Tyfa, and Ward (2004) gives a more restrictive definition of backtracking, which requires the fixation ending the backtracking saccade to be within a minimal distance (set to 2°) for the event to count as a backtrack. Figure 8.6 illustrates the two operationalizations. When comparing their own definition to that of Goldberg and Kotval, Renshaw *et al.* found, in data collected from a usability evaluation task, that their more restrictive measure was more sensitive, arguing that it is a better measure for finding differences in usability studies.

In reading research, backtracks have been defined as “sequences of three or more left-going saccades, each of which is no greater than 13 character spaces in extent. Typically, these comprise a series of short saccades directed to a sequence of words, which are, as a result, inspected in reverse order.” (Murray & Kennedy, 1988). It is interesting to note that Tatler and Vincent (2008) found that reversal of direction in a scanpath is preceded by a longer than average fixation.

8.4.2 The regression family of events

The closely related term *regression* refers to events that are similar to backtracks but not the same. In order to be a regression, the saccade needs to move in the opposite direction to the text, but not necessarily the opposite direction to the previous saccade. Figure 8.7 illustrates the difference.

Regression events exist in different sizes: an *in-word regression* is a small movement

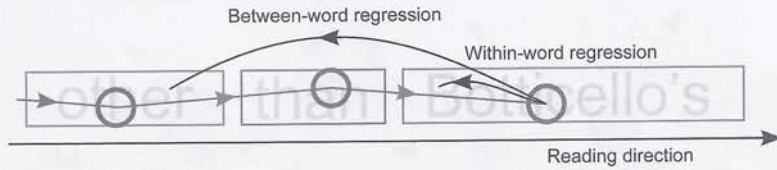


Fig. 8.8 A short in-word regression that does not leave the word just looked at, and a long between-word regression.

4
broom many times over and provides the replicants with arms and
hands to carry the buckets. Having given the command to the brooms
to fill the well he promptly falls asleep only to be woken by the flow of

(a) Fixation number 20 at 'given' is the foremost fixation in the text. Then the regression starts with a saccade back to 'broom many' at the beginning of the previous line.

4
broom many times over and provides the replicants with arms and
hands to carry the buckets. Having given the command to the brooms
to fill the well he promptly falls asleep only to be woken by the flow of

(b) The regression scanpath continues as the participant re-reads one and a half lines, until fixation number 49 at 'the', just before 'command', passes fixation number 20.

Fig. 8.9 A regression scanpath is a reading event, defined as going back in the text (a) and re-reading a passage. The regression scanpath ends when the point of departure from forward reading is passed and the participant resumes reading left-to-right (b).

backwards within a single word, a *between-word regression* moves further back in the sentence, to a previously fixated word (Figure 8.8).

The *regression scanpath* is a scanpath event in reading research. Figure 8.9 shows a regression scanpath starting after fixation number 20. Not until fixation number 48 is gaze back at the same word, so this regression scanpath has a length of 28 fixations.

Hyönä, Lorch Jr, and Kaakinen (2002); Hyönä *et al.* (2003) furthermore differentiate between *re-inspections* (when re-reading parts of the currently fixated sentence) and *look-backs* (when re-reading parts of another previously read sentence). Each of these are events from reading data.

8.4.3 The look-back and inhibition of return

Both backtracking and regressions differ from *look-backs*. Look-backs are operationalized as saccades to AOIs already looked at; they are also known as 'returns' and 'refixations'. The term look-back more pertains to spatially extended viewing behaviour, outside the field of reading research, for example searching for a target in a visual array or a link on a web page.

Look-backs are closely related to the concept of *inhibition of return* (Posner, Rafal, Choate, & Vaughan, 1985), which is the observation that attention is unlikely to be re-directed to previously inspected areas within a transient temporal window—in short, we do not look back to places we have just looked. Although considered a well-established phenomenon,

Smith and Henderson (2009) found only little support for a general inhibition of return mechanism in scanpaths during scene viewing. They provide empirical evidence that saccades frequently go back to previous fixation locations. It therefore seems that inhibition of return is not sustained enough to account for extended viewing scanpaths. This is not to undermine the importance and validity of inhibition of return however. There is a large corpus of literature examining the phenomenon. Its effects are more subtle and are likely masked when viewing visually complex stimuli like scenes.

"Returning to a previously fixated item constitutes a failure of memory", Gilchrist and Harvey (2000) write, but working memory has a limited temporal capacity, and it should matter how long ago the AOI was previously looked at for fixations there to count as look-backs. If the AOI content is no longer in working memory, is a look-back then different from looking at an object that has not yet been fixated? With this in mind, Mennie, Hayhoe, and Sullivan (2007) limited the operational definition of look-backs to within a 10 s window (after a reach and grasp sequence had been completed). Gilchrist and Harvey (2000) instead measured the number of "between-refixation intervals"—the periods between visits of the same AOI—at different durations in a visual search task, finding them to be below 8 seconds for one participant, and below 3 seconds for the two others.

8.4.4 The look-ahead

Look-aheads are saccades forward towards, and resulting fixations upon, objects that will soon be used, picked up, or in other ways be part of future planned actions.

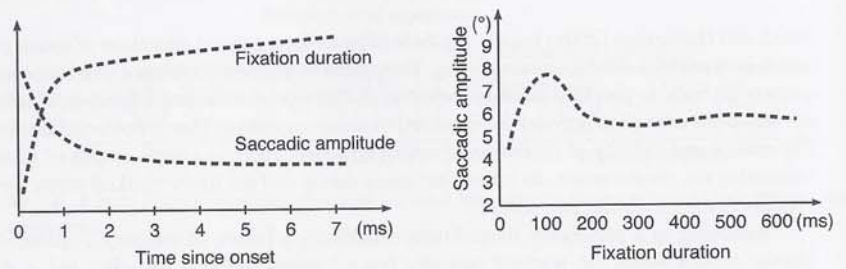
Mennie *et al.* (2007) define look-ahead fixations according to the following spatial and temporal properties. When a fixation lands upon an AOI corresponding to a container within the "10 second period before the initiation of a reach from the workspace to that container" it counts as a look-ahead fixation. They emphasize that guiding fixations coinciding with the reach are excluded from the look-ahead category.

Although not providing a precise definition of look-aheads, Pelz, Canosa, Babcock, and Barber (2001) propose that "look-ahead fixations represent a strategic deployment of attentional and visual resources to optimize information gathering during natural tasks". In this sense, look-aheads are highly task dependent. Pelz *et al.* (2001) reported that, in a hand washing task, 3% of all the fixations were look-aheads compared to only 1% in a less complex control task.

8.4.5 The local and global subsfans

The idea of local scanpaths is that spatially confined saccades with a small amplitude—Zangemeister, Sherman, and Stark (1995) suggest the thresholds 1.6°, 4.6°, 7.9°, or 11°—belong to a local scan of the particular details in a small patch of the stimulus. Larger saccades are assumed to belong to global overview scanning. Viewers tend to alternate between global and local scans when inspecting visual scenes, and tend to start with a global overview scan directly after onset of the scene.

Overview scans appear to be associated not only with longer saccadic amplitudes but also with shorter fixation durations (Unema, Pannasch, Joos, & Velichkovsky, 2005). These authors, amongst others, propose that global and local scanpaths are indications of two different types of cognitive processing: *ambient* and *focal*, respectively. The distinction is based on the well-known observation that fixation durations gradually increase from the time of stimulus onset, while saccadic amplitudes decrease. Participants initially exhibit scanning of the salient features of the image—long saccades, short fixations (ambient processing)—and later inspect the local areas in more detail—using short saccades and long fixations (focal



(a) Fixation duration and saccadic amplitude over bins of time

(b) Fixation duration against saccadic amplitude

Fig. 8.10 Short fixation durations combined with long saccades are characteristic of ambient processing, according to Unema *et al.* (2005), while longer fixation durations and shorter saccades are indicative of focal processing. These principal drawings are adapted from data presented by Unema *et al.* (2005), and Tatler and Vincent (2008).

processing).

The local versus global distinction indicates that clustering, for instance on saccadic amplitude, may separate data into events belonging to two qualitatively different categories. The need for a threshold is its major drawback.

On a more theoretical level, Groner, Walder, and Groner (1984) argue that local scans are not necessarily comprised of short saccades, and global scans long saccades. Rather, local scanpaths are seen to be driven by cognitive processing real-time, whereas global scanpaths are driven by an *overall* search strategy, or purpose. The former is bottom-up mediation of scanpaths, while the latter reflects top-down control.

8.4.6 Ambient versus focal fixations

Ambient and focal are two different cognitive states that a fixation with its preceding saccades is believed to correspond to. The distinction is operationalized as illustrated in Figure 8.10: a fixation with duration below a threshold d following a saccade with amplitude above threshold a is ambient (overview scanning), while the fixation is focal (focused inspection) when the fixation duration is above d and the preceding saccadic amplitude is below a . The thresholds do not have clearcut settings, but Velichkovsky, Rothert, Kopf, Dornhöfer, and Joos (2002) argue that d should be in the vicinity of 250 ms, and a around 4° . In principle, an event detection algorithm could categorize fixations into ambient and focal.

Alternative operationalizations exist. The *saccade/fixation ratio* from Goldberg and Kotval (1999) is argued to compare search time (saccades) to processing time (fixations). This variety uses saccade duration rather than amplitude, which is of little importance, as both correlate strongly. The global to local (g/l) ratio measure on page 338 quantifies the overall amount of global scanning versus detailed inspection, but does not classify the fixations.

The interpretation stems from Buswell (1935), who noted that the earliest fixations in a picture are shorter (around 210 ms) than later (around 360 ms). Also, saccadic amplitudes are longer in the initial scan and decrease over time (Figure 8.10). Several studies have repeated both these findings, and interpreted them as indicative of an early orienting period, followed by a more scrutinous inspection of informative details.

Unema *et al.* (2005) argue that a model for saccade generation with two visual processing systems, ambient and focal processing, can be supported with data on this relation. Ambient processing is characterized by long saccades and short fixation duration, corresponding to

the peak in 8.10(b). Ambient processing is thought to be a process that is bottom-up driven and which creates an overview for later focal processing. Their speculation is that ambient processing is linked to what is known as the where/how system of the dorsal stream. Focal processing, which takes place at the flatland in Figure 8.10(b), conversely reflects top-down-driven scrutinous inspection of details which may be associated with the corresponding 'what' system.

When participants searched for possibly camouflaged military vehicles in photos, their mean saccade amplitude decreased and mean fixation duration increased gradually as a function of the ordinal saccade and fixation number (Over, Hooge, Vlaskamp, & Erkelens, 2007), independent of whether they knew that the target was the only unknown part of the scene or not. Over *et al.* interpret the result as showing that the coarse-to-fine search strategy is used even when it is not optimal.

8.4.7 The sweep

Aaltonen, Hyrskykari, and Riih  (1998) define a 'sweep' as a *sequence of saccades that move in the same direction*, and compare downward and upward scanpaths (sweeps) of varying sizes. Aaltonen *et al.* do not present a computational method for detecting sweeps, but various computational operationalizations are possible (e.g. detecting sweeps has similarities with detecting reading; see the next section).

8.4.8 The reading and scanning events

Reading and scanning events can be loosely defined as scanpath patterns that correlate with the cognitive processes of reading and scanning (see Rayner & Fischer, 1996 for a distinction between reading and scanning). In practice, such patterns are detected by algorithms that follow a set of predefined criteria, much like those detecting fixations and saccades. As with the detection of other scanpath events, reading detectors identify physical properties of scanpaths, and do not assess cognitive processing in the participant; it is indeed possible to move the eye in a reading-like pattern, even though the mind is occupied with something else.

The simplest reading detectors use criteria for saccadic amplitude, which is quite short during focused reading, whilst when scanning across pages saccades are longer. More advanced versions of reading filters add requirements on saccades such that they must be horizontal in direction (within certain bounds), or fixation durations such that different timings correspond to reading rather than scanning.

Figure 8.11 shows the principle for a simple reading detector. It assumes that the current fixation is located at position $(curr_h, curr_v)$, and imposes an area in which the next fixation must land for it to be part of a reading event. The area, indicated by a rectangle, spans about two words ahead horizontally, and a word back (in case of in-word regressions); there is also some margin for upwards and downwards movements along the line. A second return sweep detector finds long backward movements corresponding to line length; this also detects slightly downward movements, again with a delimiting area around the expected landing position at the start of the next line. Reading is assumed to have been detected if at least three fixations (two intermediate saccades) fulfill the detection conditions. The reading detector developed by Holmqvist *et al.* (2003) compared the amount of reading in paper newspapers compared to internet newspapers. It was found, in contradiction to popular expectation, that traditional newspapers give rise to more reading and less scanning, with the opposite being true of online news.

Several other detection algorithms have been published:

1. Attempting detection of online reading, Campbell and Maglio (2001) use three criteria

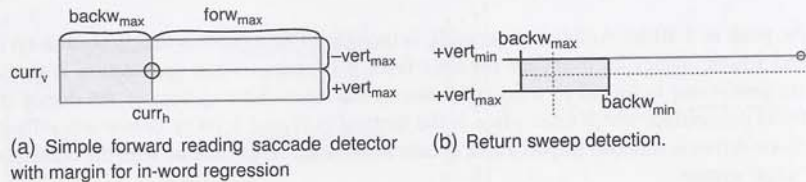


Fig. 8.11 Spatial requirements for the simple algorithmic reading detector used by Holmqvist *et al.* (2003). Successive saccades should move within or to restricted boxes to count as reading.

on saccades: amplitude (long versus short), direction (right, left, up, and down), and axis (x versus y). Using these criteria, each recorded saccade is given a total score that, when summed over all combinations of criteria, can then be compared to a threshold for reading. The authors report high classification accuracy and a detection latency of 1000 ms.

2. Simola, Salojärvi, and Kojo (2008) trained a 9-state discriminative hidden Markov model to differentiate between eye-movement data from three different reading tasks: (i) simple word search, (ii) finding a sentence that answers a question, and (iii) choosing the subjectively most interesting title from a list of ten titles. Their model has a 60% accuracy in determining the correct type of task.
3. Kollmorgen and Holmqvist (2009) used Markov models to train a version of the reading filter from Holmqvist *et al.* (2003) to detect reading in eye-movement data recorded from participants writing on computers (data shown on p. 291). This was also used to analyse the interplay between reading and writing activity in a number of other studies (see Johansson, Johansson, Wengelin, & Holmqvist, 2008; Wengelin *et al.*, 2009; Johansson, Johansson, Wengelin, & Holmqvist, 2010). The 6-state hidden Markov model has a precision/recall of 0.88/0.87 on validation data.
4. Based on the work by Campbell and Maglio (2001), Buscher and Dengel (2009) implement a reading and skimming detector. Again, saccades are scored based on saccadic amplitudes such that a sequence of long saccades is likely to reflect skimming over the text.

8.5 Scanpath representations

Even a scanpath built from raw data samples contains so much information that calculations quickly become computationally complex. Therefore, several *representations* of scanpaths have been developed that take only selected aspects of the oculomotor behaviour included within scanpaths into account. In the following sections, we will describe how sequences of *symbols* (typically letters representing AOIs), *Euclidean vectors*, and *attention maps* are used to represent scanpaths.

Besides containing voluminous amounts of data, scanpaths generated from data samples may contain information that is largely irrelevant to the experimental questions at hand. The goal when building the scanpath representation is therefore to retain as much of the relevant information as possible while allowing the desired visualization or calculation. AOI strings, for example, have mainly been developed to render string-edit comparisons (p. 348) of scanpaths possible. However, this comes at the cost of decreasing the spatial and temporal resolution of the scanpath, since the positions of data samples are replaced with letter strings corresponding to larger areas.

Scanpath representations reside at the top of the hierarchy. A small error in the simpler components (such as fixation detection criteria) travels all the way up, thus affecting higher representations (such as scanpath length). For instance, Green (2006) calculated *scanpath length* on her data as the sum of all saccadic amplitudes. She used the I-DT algorithm with a minimal fixation duration setting of 200 ms however, which strongly reduces the number of detected fixations as well as the distribution of their durations (p. 159). Green includes a raw data sample scanpath (the sum of all distances between successive samples) "to ensure that potential group differences in fixation scanpath length did not simply reflect group differences in the number of fixations." (her page 87), or, we could add, the settings of the algorithm. Overall, the reported saccade-based scanpath length is about half the raw sample scanpath length. As another example, a scanpath represented by gridded AOIs is uniquely defined by the grid size. Consequently, scanpaths that are judged as similar using one grid size may become dissimilar when using a smaller or larger grid size. Choosing an appropriate scanpath representation with carefully defined fixations, saccades, AOIs, and/or other entities is of crucial importance for the outcome of a study.

Three formal scanpath representations have been devised for algorithmic comparison of pairs of scanpaths, and the calculation of a few other measures. These are: strings, vectors, and attention maps. They make use of the representations defined in Chapters 5–7.

8.5.1 Symbol sequences

Symbol sequences refer to a string of symbols, typically letters, that represent selected aspects of a scanpath with or without relation to AOIs. By far the most common type is the *AOI string*, where each symbol represents either fixations or dwells in an AOI. Other types of strings are based on properties such as *fixation duration*, *saccade amplitude*, or *saccade direction*. These representations are mainly developed for the purpose of calculating scanpath similarity using the string-edit (Levenshtein, 1966), or related methods.

AOI-based fixation and dwell strings

Several scanpath measures, the most known of which is the string-edit measure, represent scanpaths with a string of fixations or dwells in areas of interest (AOIs), as defined in Chapter 6. The important difference between *gridded AOIs* and *semantic AOIs* should be taken into account in this type of representation. Figure 8.12 shows both varieties. The gridded AOIs are constructed by putting a grid of equally sized areas across the stimulus, ignoring whatever semantic parts the stimulus consists of. When a scanpath runs over the gridded AOIs, each fixation or dwell is replaced by the name of the AOI it hits. The scanpath in Figure 8.12(a) will be represented by the string: A6 C5 F0 I1 J1 K2 I3.

Each letter is a representation of a single fixation position within a whole AOI area, which is known as *spatial downsampling*. As a consequence, a small difference in gaze pattern would be enough to alter the string, but interestingly, other small differences would result in the *same* string. That is, some small differences matter, whereas others are ignored. Fixation and dwell string representations thus introduce a form of noise in your scan path data that may occlude actual results.

A division of stimulus space into semantic AOIs adopts the natural semantic parts of the stimulus. In Figure 8.12(b), the semantic AOIs are taken from Josephson and Holmes (2006), who used them to analyse viewing behaviour on a television screen. Semantic AOIs have different sizes, so in Figure 8.12, the scanpath example would be represented by the string MMTCCCHGM (where M = 'Main' etc., and each letter denotes a fixation). The three Ms in the string represent fixations with very different positions in the stimulus, which means we have a very coarse position representation. This can be motivated if the 'Main' AOI is a

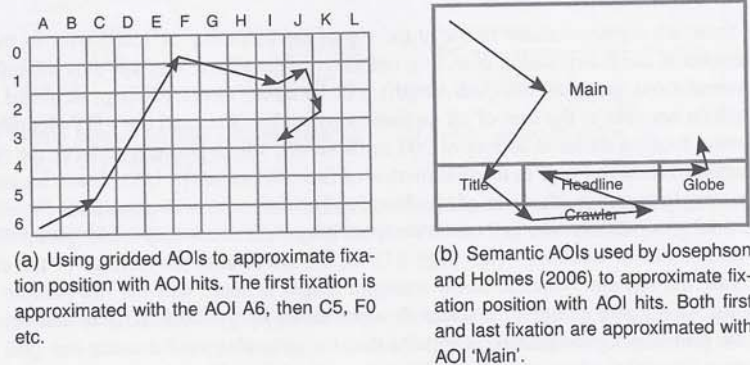


Fig. 8.12 Gridded versus semantic AOIs for representing scanpaths.

semantically homogeneous area from the viewpoint of the hypothesis, analysis and theory. Otherwise, *very* different scanpaths will be represented as equal when using semantic AOIs.

As an alternative to gridded and semantic AOIs, data driven methods where recorded eye-movement data are used to define AOIs are only beginning to emerge for scanpath representations (see Santella & DeCarlo, 2004; Hooze & Camps, 2009). As pointed out earlier (pp. 219–220), data-driven AOI representations should be used with appropriate care.

Since our sample scanpath has two subsequent fixations in each of 'Main' and 'Crawler', we have a repetition for M and C in the string MMTCCCHGM. Brandt and Stark (1997) use a variety of the AOI-string representation of scanpaths that ignores such repetitions, or in other words, they represent the scanpath with dwells rather than with fixations. The dwell-based AOI-string for the same scanpath would be MTCHGM. Removing consecutive repetitions is referred to as "compressing" the string; several fixations (CC) are replaced by one dwell (C).

The major advantages of the AOI string representation of scanpaths are that it retains an approximate *sequence* representation of the order of fixations, and that the string is a rough representation of the *shape* of the scanpath. The major drawback is the reliance on AOI segmentation, which necessarily introduces noise in the measures that rely on it.

Saccade amplitude and direction based strings

While being most frequently used, AOI strings represent scanpaths only in relation to spatial areas. If we are more interested in other aspects of a scanpath, strings can be constructed from other properties of eye-movement data such as the amplitude and direction of saccades, or the durations of fixations.

Gbadamosi (2000) and Zangemeister and Liman (2007) developed a string representation of scanpaths that combines one number for amplitude with another for direction. They use 16 (hexadecimal) numbers for each, as illustrated in Figure 8.13. Each saccade is therefore represented by a pair of (hexadecimal) numbers, and a scanpath with a sequence of pairs, such as D6 23 71 28 73 B3 54. This representation of scanpaths does not require a segmentation of space into AOIs; in fact, it completely abandons positional and semantic information, and instead focuses on representing our subjective perception of the overall shape of a scanpath. Still, it is based on segmentation (of direction and amplitude) and therefore again introduces noise when used in measures.

Letter strings with fixation durations have been proposed in the literature (Jarodzka, Nyström, & Holmqvist, 2010; Goldberg & Helfman, 2010), although they are only just beginning to appear (see Cristino *et al.*, 2010 and p. 353).

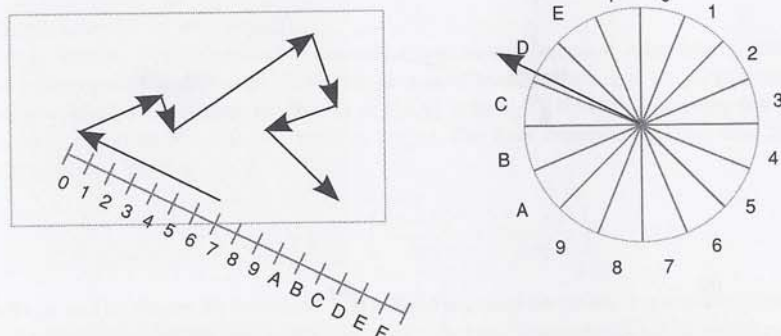


Fig. 8.13 To the left, the amplitude (length) of the first saccade is measured using a hexadecimal (16 unit) ruler. To the right, a discreet 16 region segmentation of saccadic direction, and the same first saccade has been placed so its direction is measured. With a 6 unit amplitude and a D for direction, the first saccade is represented by D6, where the first (hexadecimal) number in each pair represents direction and the second amplitude. The whole scanpath is represented by the string: D6 23 71 28 73 B3 54.

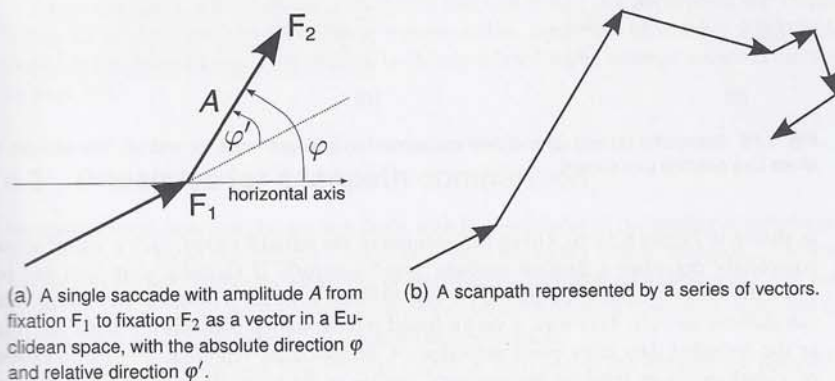


Fig. 8.14 We can think of a scanpath as a sequence of vectors in space (b). Each vector is described by its origin F_1 and endpoint F_2 . We can immediately derive its amplitude A , its angular deviation from the previous saccade ϕ' , and its direction ϕ relative to the horizontal axis.

8.5.2 Vector sequences

Because of the reliance on the string-edit similarity calculation, few non-symbol representations have evolved. However, Jarodzka, Nyström, and Holmqvist (2010) argued that Euclidean vectors were suitable to represent ideal saccades, i.e. those that take the shortest distance between subsequent fixation positions.

A vector is a mathematical entity with length (A) and direction (ϕ), as shown in Figure 8.14(a). Apart from curvature, it thus approximates two major properties of a saccade. In addition, the vector representation does not require a segmentation of space, but retains the original detail of the ideal saccade and hence introduces no additional noise in future calculations. As shown by Jarodzka, Nyström, and Holmqvist (2010), a similarity calculation can nevertheless be made, using geometric calculations instead of letter identity checks.

Using vectors for single saccades, a scanpath can be represented as a *sequence of vectors*,

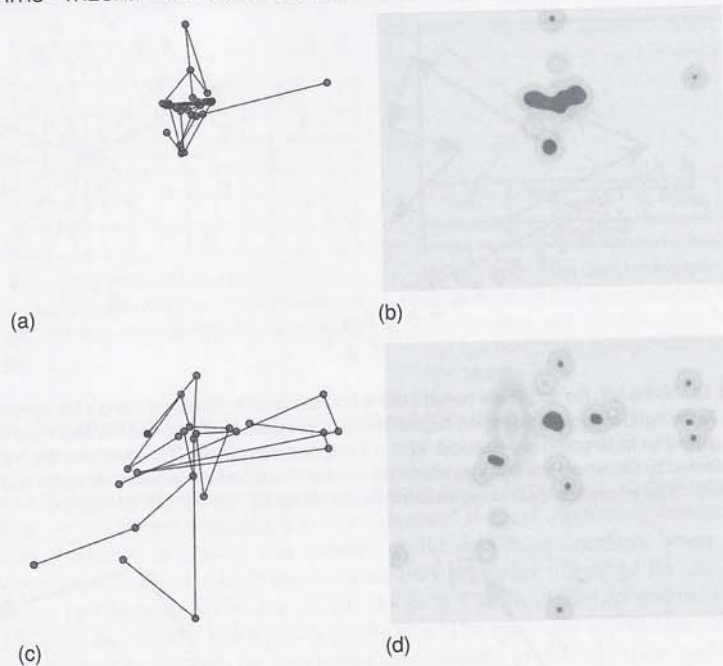


Fig. 8.15 Scanpaths (a) and (c) and their corresponding attention maps (b) and (d). The stimulus is the Mona Lisa painting (not shown).

as shown in Figure 8.14(b). Given the position of the initial fixation, such a vector sequence completely describes a fixation–saccade based scanpath. If saccadic start- and end-points explicitly encode fixation positions, we can attribute a duration to them, and utilize them in calculations as well. This way, a vector-based representation contains the same information as the recorded data after event detection. A vector-based representation can additionally be visualized in an intuitive manner very similar to the scanpath visualizations from most manufacturer software.

8.5.3 Attention map sequences

Attention maps, introduced in Chapter 7, are two-dimensional functions or ‘maps’ that represent the spatial distribution of a scanpath(s). Examples are shown in Figure 8.15 where the left pictures illustrate scanpaths connecting fixation sequences with lines, and the pictures on the right illustrate the corresponding attention maps. Attention maps are generated from either data samples or fixations, and ignore the order of these. According to the scanpath definition in this book, an attention map therefore does not qualify as a scanpath representation, because it does not signify the route, or order, of oculomotor events through space.

Nevertheless, the attention map measures of pages 359–376—including the Mannan distance measure developed by Mannan, Ruddock, and Wooding (1996)—are often referred to as ‘scanpath metrics’ in the literature (for example, by Cerf, Harel, Huth, Einhäuser, & Koch, 2009; Boccignone *et al.*, 2005). Similarly, in their overview over what they refer to as scanpath analysis metrics, Underwood *et al.* (2008a) discuss both the Mannan distance measure and fixation maps, before settling for the string-edit analysis, which uses fixation and dwell

string representations of scanpaths.

It is however straightforward to generate a *sequence of attention maps* from a scanpath, thus retaining order information. In the simplest case, each position $S_i = (x_i, y_i)$ in a scanpath can be replaced with an attention map $G = f(x, y)$, where $f(\cdot)$ denotes a function that generates an attention map G from the position (x_i, y_i) . The most common function used for this purpose is the Gaussian

$$f(x, y) = \exp - \left(\frac{(x - x_i)^2}{2\sigma_x^2} + \frac{(y - y_i)^2}{2\sigma_y^2} \right) \quad (8.1)$$

where σ_x and σ_y denote the horizontal and vertical standard deviation, respectively. Given just one scanpath, its attention map equivalent carries the same spatio-temporal information, since peaks in the attention maps correspond directly to positions in the scanpath. An alternative is to think about the attention map sequence as a probabilistic scanpath that encodes the probability that a certain spatial position will be visited by a participant's gaze. However, as more scanpaths are used to build the attention map sequence, the 'path' in each individual scanpath is lost, but not the ordinal, collective eye-movement behaviour. In this sense, an attention map sequence does not violate our definition of what a scanpath is.

More generally, we can define a *scanpath function* $g(x, y)$ that represents any mapping from a set of scanpaths to an alternative representation, regardless of whether it is based on Gaussian functions or not. This concept is closely related to the average scanpath discussed on page 282.

8.6 Principles for scanpath comparison

We should stress here that this section deals with the *principles* of comparing scanpaths only, and by this we mean the different processing steps necessary before scanpath comparison can be carried out. The exact measures of quantitative scanpath comparison are covered in detail in Chapter 10, where we also discuss other measures of movement in eye-tracking data, besides the movement inherent in a scanpath representation.

According to Henderson and Ferreira (2004, p. 42), one of the reasons for reduced research interest into 'scan patterns', as they refer to them, is the lack of appropriate scanpath similarity measures for calculating commonalities in viewing behaviour both within- and between-participants: "it may be that with more careful examination, more consistency in scan patterns will be found, especially as analysis tools become more sophisticated". Indeed, scanpath comparison is needed in fields from neurological schizophrenia research (Benson *et al.*, 2007), over to mental imagery (Johansson *et al.*, 2006) and scene perception (Underwood, Humphrey, & Foulsham, 2008b), but also including applied areas such as evaluating web page design (Josephson & Holmes, 2002), and in expertise task analysis (Jarodzka, Scheiter, *et al.*, 2010). See the unresolved issues section towards the end of this chapter (p. 278) for further discussion.

Methods to compare two or more sequences have long been used in bioinformatics to align DNA and protein sequences. In this field they have been used to test, for example, whether mutated sequences have evolved from the same original sequence, or to match a test sequence against a database.

In eye-movement research, the majority of sequence comparison methods have been used to compare pairs of scanpaths comprising AOI-strings. For example, Brandt and Stark (1997) compared scanpaths from participants viewing checkerboard patterns with scanpaths from subsequent visual imagery of the same stimuli. Similarly, Foulsham and Underwood (2008)

compared scanpaths from participants viewing the same pictures twice; the initial viewing was followed by a later recognition test.

Scanpath similarity calculation is typically performed in four sequential steps:

1. *Representation*. A scanpath representation is chosen that best fits the experimental questions at hand.
2. *Simplification* is performed to remove parts of the scanpaths that are irrelevant to the hypothesis or that decrease subsequent computational power.
3. *Alignment* aims to optimally map entities between scanpaths.
4. *Calculation* of similarity between the aligned scanpaths then generates the final similarity score(s).

Only the last step is required to actually compute a similarity score, whereas the first three steps are performed to prepare the scanpaths to best address the experimental questions.

As noted, we discuss the principles of scanpath comparison here, and leave the implementation details of specific algorithms, as well as the calculation of similarity scores, for Part III.

8.6.1 Representation

The first step concerns choosing a suitable representation of the scanpath. The most common representations are data samples (x, y, t) , AOI letter strings, vector sequences, and attention maps sequences. By definition, a representation does not retain all oculomotor information contained within the scanpath, but provides a more crude approximation of some of its features.

8.6.2 Simplification

Scanpaths are sometimes subjected to a process that typically is called *compression*, or *simplification*. Compression further simplifies the representation such that detail considered superfluous by the designer of the measure—such as several small fixations next to one another in Figure 8.16—is removed, and the output is a simplified version of the original scanpath. For AOIs strings, ABC would be a compressed version of the original string ABBBC, since the repetitive Bs have been removed. This type of simplification is suitable when order of AOI visits is of relevance, but not when dwell time or number of fixations within an AOI is of importance.

8.6.3 Sequence alignment

Sequence alignment originates from bioinformatics and aims to arrange two or more sequences in such a way that their elements (e.g. amino acids) optimally match. For very short sequences, the optimal match is usually easy to identify simply by looking at the sequences. However, as they get longer, alignment become an increasingly harder problem that needs to be addressed computationally. Then optimality is defined by a set of scoring parameters, and alignment is reduced to the problem of finding the optimal solution (e.g. the maximum or minimum score) based on the scoring scheme. Common scoring parameters consist of costs to insert, delete, and substitute sequence elements or blank spaces (gaps).

Alignment is typically *global* or *local* in nature. Global alignment intends to optimally match whole sequences with each other, whereas local alignment aims to identify shorter subsequences that are shared between the sequences being aligned. As an example, consider the following sequences:

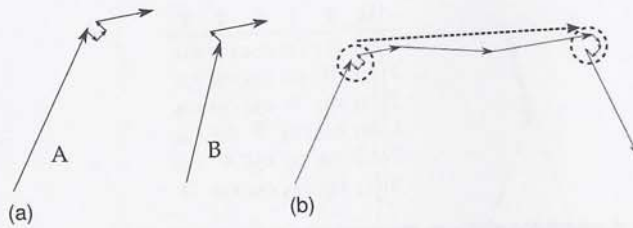


Fig. 8.16 Simplification of a vector-based scanpath representation as a prerequisite for successful alignment. In (a), four fixations in scanpath A with small saccades between them should be aligned to two fixations in scanpath B. Our intuition tells us that these small saccades are irrelevant for the comparison, and algorithmic simplification provides a firm tool to justify this judgement. Figure (b) shows two types of simplifications: small saccades are removed (dashed circles), and sequences of saccades in the same direction are removed (dashed arrow).

Global	FTFTALILLAVAV	S ₁
	F--TAL-LLA-AV	S ₂
Local	FTFTALILL-AVAV	S ₁
	--FTAL-LLAAV--	S ₂

Fig. 8.17 Sequence alignment between a short and a longer scanpath represented as strings. *Global* sequence alignment creates several gaps, indicated by -, in the shorter AOI fixation string, while *local* sequence alignment maximizes the length of substrings. Example from Wikimedia Commons.

S₁: FTFTALILLAVAV
 S₂: FTALLAAV

These are aligned according to Figure 8.17. Gaps or blank spaces, denoted by -, have been introduced at locations where matching between elements has not been possible. They correspond to one participant looking at some extra AOI, for instance, that the other participant skipped. A *gap penalty* is then paid for every such unmatched element. Notice that the number of possible alignments becomes huge as the sequence length increases, and that the optimal alignment is not necessarily unique.

A concise way to express the cost associated with matching one sequence element to another is the *substitution matrix*. Table 8.1 shows such a matrix for the sequences in Figure 8.17. Thus, the cost of matching I in one string with T in another is given by the value c_{35} . The diagonal is zero, and the reason for this is because there should be no cost associated with replacing an element with itself. Most substitution matrices are symmetric, that is $c_{ij} = c_{ji}$, but nothing prevents you from assigning higher costs for alignments in one direction and building an asymmetric similarity measure. The substitution matrix is particularly useful since it encodes the relationships between sequence elements. Translated to AOIs, this means that costs can be directly related to how similar two AOIs are with respect to each other in terms of spatial (e.g. Euclidean distance) or psychological (e.g. semantic similarity) aspects. Together with the gap penalty, the substitution matrix defines the optimal sequence match.

The matching process often utilizes a *comparison matrix*, in which the two strings being compared form the column and row divisions in the matrix, as shown in Table 8.2. Again, substitution costs can be added that allow the computation of an optimal path through the

	A	F	I	L	T	V
A	0	c_{12}	c_{13}	c_{14}	c_{15}	c_{16}
F	c_{21}	0	c_{23}	c_{24}	c_{25}	c_{26}
I	c_{31}	c_{32}	0	c_{34}	c_{35}	c_{36}
L	c_{41}	c_{42}	c_{43}	0	c_{45}	c_{46}
T	c_{51}	c_{52}	c_{53}	c_{54}	0	c_{56}
V	c_{61}	c_{62}	c_{63}	c_{64}	c_{65}	0

Table 8.1 A substitution matrix for the sequences in Figure 8.17. The cost of aligning I in one string with T in another is given by the value c_{35} . The Levenshtein measure has $c_{ij} = 1$ for all cells except the diagonal, where $c_{ij} = 0$.

	F	T	A	L	L	L	A	A	V
F	c_{22}	c_{25}	c_{12}	c_{24}	c_{24}	c_{24}	c_{12}	c_{12}	c_{26}
T	c_{25}	c_{55}	c_{15}	c_{45}	c_{45}	c_{45}	c_{15}	c_{15}	c_{56}
F	C_{22}	c_{25}	c_{12}	c_{24}	c_{24}	c_{24}	c_{12}	c_{12}	c_{26}
T	c_{25}	C_{55}	c_{15}	c_{45}	c_{45}	c_{45}	c_{15}	c_{15}	c_{56}
A	c_{12}	c_{15}	C_{11}	c_{14}	c_{14}	c_{14}	c_{11}	c_{11}	c_{16}
L	c_{24}	c_{45}	c_{14}	C_{44}	c_{44}	c_{44}	c_{14}	c_{14}	c_{46}
I	c_{23}	c_{35}	c_{13}	C_{34}	c_{34}	c_{34}	c_{13}	c_{13}	c_{36}
L	c_{24}	c_{45}	c_{14}	c_{44}	C_{44}	c_{44}	c_{14}	c_{14}	c_{46}
L	c_{24}	c_{45}	c_{14}	c_{44}	c_{44}	C_{44}	c_{14}	c_{14}	c_{46}
A	c_{12}	c_{15}	c_{11}	c_{14}	c_{14}	c_{14}	c_{11}	C_{11}	c_{16}
V	c_{26}	c_{56}	c_{16}	c_{46}	c_{46}	c_{46}	c_{16}	c_{16}	C_{66}
A	c_{12}	c_{15}	c_{11}	c_{14}	c_{14}	c_{14}	c_{11}	c_{11}	c_{16}
V	c_{26}	c_{56}	c_{16}	c_{46}	c_{46}	c_{46}	c_{16}	c_{16}	c_{66}

Table 8.2 A comparison matrix for the sequences in Figure 8.17. The path for local alignment is outlined with Cs.

comparison matrix, from the upper left corner to the lower right corner. With the optimal path for local alignment, denoted by C in the comparison matrix, vertical and horizontal movements are treated as gaps, whereas diagonal movements represent matches.

The popular string-edit algorithm uses three types of operations for sequence matching: insert, delete, and substitute (Levenshtein, 1966), usually with equal weights. Therefore, it can be seen as optimal sequence matching to use a very simple substitution matrix containing only ones in the cells c_{ij} .

Recently, more flexible algorithms have been adapted into eye-movement research and scanpath comparison, mainly the Needleman-Wunsch algorithm (Needleman & Wunsch, 1970) and Dijkstra's algorithm (Dijkstra, 1959). The Needleman-Wunsch algorithm is a method to align two sequences globally and allows a flexible scoring scheme. It performs best for sequences with already high similarity and similar lengths. Dijkstra's algorithm operates on representations known as *graphs* to find the shortest path from one node in the graph to another. The use of Dijkstra's algorithm presupposes that the comparison matrix is first transformed to a graph representation.

Note that a substitution matrix is only of practical use when the number of possible sequence elements is finite and small. In the case of sequences with infinite alphabets (i.e. an unlimited number of unique elements), costs have to be calculated directly in the comparison matrix. This is the case for example when scanpaths are represented by a sequence of vectors,

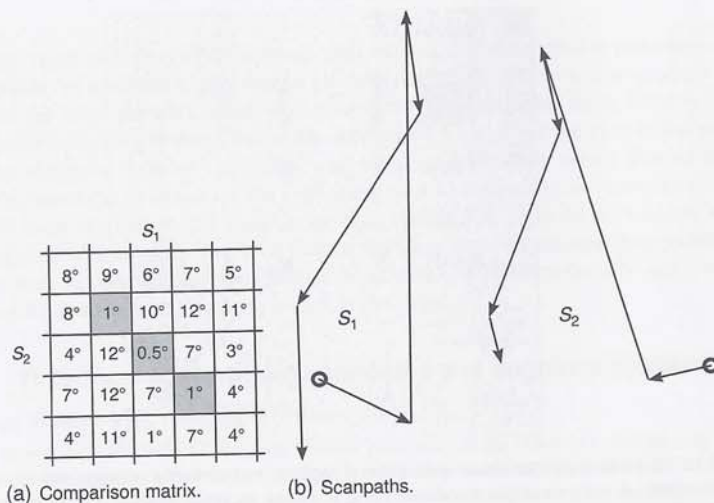


Fig. 8.18 Pairwise vector comparison of all saccades in two vector representations $S_1 = \{u_1, u_2, u_3, u_4, u_5\}$ and $S_2 = \{v_1, v_2, v_3, v_4, v_5\}$ of scanpaths. Circles denote the onset of a scanpath. The comparison matrix (a) shows, for each pairwise vector comparison, the length of the differential vector $\|u_i - v_j\|$ in degrees of visual angle. If the two vectors are similar in amplitude and direction, this value is low. Figure (b) shows the scanpaths used in the comparison. Grey matrix cells indicate consecutive mappings that would produce a good alignment for a subset of the saccades.

which can have a large number of possible amplitudes and directions. Figure 8.18 illustrates a situation where saccade vectors u_i and v_i are compared to each other and the substitution cost equals the length of the differential vector $\|u_i - v_j\|$.

Closely related to the comparison matrix is the visualization known as a *dot-matrix plot* or simply *dotplot*. Instead of assigning each value in the matrix with a substitution cost, it adds a dot at positions where elements match, while keeping other matrix cells empty. This is visualized in Figure 8.19, which shows a dotplot for self-similarity where one string is matched with itself to point out commonly *recurring* subsequences. Two identical strings with a unique set of elements (i.e. without recurrences) would generate a diagonal in an otherwise empty dotplot. A collection of dots located off the main diagonal represent matching subsequences. However, more than being a method to align and quantitatively compare sequences, dotplots are typically used to subjectively inspect and interpret similarity.

8.6.4 Calculation

Finally, we *calculate the similarity* between the two scanpaths. Given the aligned scanpaths and the similarity metric, this is usually a simple and quick operation. For example, when the optimal alignment represented by a path in the comparison matrix is found, the similarity score is typically found by summing all cost and gap penalties along the path.

Scanpath similarity is a complex multidimensional concept, yet all similarity measures proposed so far have had at their heart the design criterion that they should output a single similarity value between 0 and 1. This could be a problem when scanpaths are very similar in some aspects, but dissimilar in others. For a multidimensional approach, which takes account of scaling, spatial and temporal offset, as well as fixation duration amongst other components which are important in the comparison of scanpaths. Details are given in Jarodzka, Nyström,

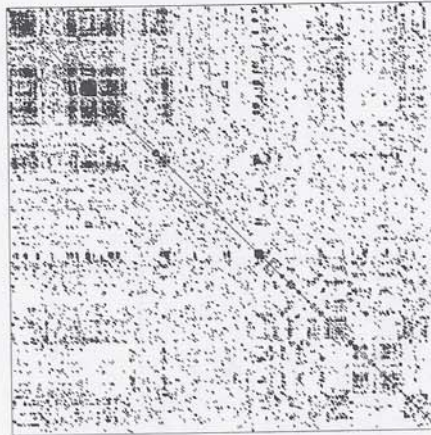


Fig. 8.19 Dot-matrix plot for visual evaluation of regional self-similarity between two identical strings. Similar repetitive patterns within the same string are seen as black regions. This is particularly common in the first 25% of string, in the upper left corner of the plot. From Wikimedia Commons (File:Zinc-finger-dot-plot.png).

and Holmqvist (2010) and on page 354.

8.6.5 Pairwise versus groupwise comparison

The comparison measures that exist for scanpaths, irrespective of their representations, are mainly pairwise and output a single value. Methods to compare groupwise similarity between scanpaths are only beginning to surface, in particular those that allow statistical testing of similarity. Feusner and Lukoff (2008) attribute the lack of statistical methods to the fact that a scanpath itself does not produce a numerical value, but only a pairwise comparison between two scanpaths. This disqualifies the use of traditional statistical methods such as the *t*-test, which require one value for each group entity (scanpath).

A way to approach groupwise scanpath comparison is first to calculate the *average scanpath* for each group (p. 265), and then compare this pair. However, if the comparison only results in a single value, statistical analysis is still not possible.

Feusner and Lukoff (2008) provide one solution to this problem by calculating the average pairwise similarity between (d_{between}) and within (d_{within}) two groups of scanpaths. Then they calculate the difference between these two values,

$$d^* = d_{\text{between}} - d_{\text{within}} \quad (8.2)$$

for all possible group divisions where one group has n scanpaths and the other one has m scanpaths ($m + n$ is fixed), yielding a symmetric distribution of d^* with zeros mean. If the groups contain random scanpaths, we would expect a d^* -value around zero. Significance tests of group similarity are then possible using permutation tests.

8.7 Unresolved issues concerning scanpaths

There are several important but unresolved issues that it is useful to be aware of when working with scanpath representations of eye-movement data. One is the question of whether there

is a direct relationship between scanpath patterns and specific cognitive processes. Another is the validity of scanpath theory, and its prediction that participants will re-produce a spatial model of the same scanpath when looking at an identical stimulus anew. Closely related is the issue of scanpath planning; that is the question of what drives the eyes to the successive locations along the route of a scanpath, and whether *inhibition of return* (Posner & Cohen, 1984) decreases the likelihood of the eyes going back to their previous location. Another unresolved issue is whether it is possible, or even meaningful, to build an 'average scanpath' from a group of scanpaths. The final unresolved issue revolves around open problems concerning how to compare the similarity of two scanpaths computationally and statistically, adopting the framework(s) outlined earlier in the chapter.

8.7.1 Relationships between scanpaths and cognitive processes

It is easy to agree with Yarbus (1967) who wrote:

Eye-movements reflect the human thought processes; so the observer's thought may be followed to some extent from records of eye-movements (the thought accompanying the examination of the particular object). It is easy to determine from these records which elements attract the observer's eye (and, consequently, his thought), in what order, and how often.

Often, Yarbus' findings raise the question of whether "thought processes" can reflect more specific cognitive states such as interest, difficulty, or confusion? Is there a specific scanpath pattern that directly and uniquely corresponds to a cognitive process?

There is indeed a general consensus that scanpaths are determined to a large extent by idiosyncratic cognitive factors (as claimed by Choi *et al.* (1995) and others). Very few studies however have targeted what a specific cognitive process (like a "thought") looks like when appearing in a scanpath. The attempts by Goldberg and Schryver (1995b), Goldberg and Kotval (1999), and Salvucci (1999) to infer intent from eye movements, led to the definition of several new measures, but none of them could be directly and systematically linked to specific cognitive processes.

Usability analysts using eye tracking are in need of a method to investigate the relationship between scanpaths and cognitive processes; Ehmke and Wilson (2007) argue that the measures used by academic researchers are of little help to someone looking at scanpaths for signs of interest, confusion, hesitation, or poor computer interface design. They argue that the usability analyst needs to be able to draw concrete conclusions from scanpaths. For instance when seeing a scanpath comprised of "many short fixations across a page where information might be expected", can she conclude that "expected information is missing"? Is it scientifically justified to ask the participant if they expected content to be present at certain points along the route of this scanpath?

There are several dangers to such manual assignment of cognitive processes to scanpaths. For instance, it is not unlikely that a scanpath described as "many short fixations across a page" will co-occur with many different cognitive states, and will not be uniquely determined by any one of them. The usability analyst may easily fall prey to guesswork, and the participant to inadvertently confirming the guesses (compare p. 105). Moreover, many of the scanpath-based concepts (such as 'regularity'), referred to by Ehmke and Wilson (2007), are vague and difficult to accurately capture in a measure. Trying to relate vague scanpath-based concepts to guessed cognitive processes is not scientific, and is unlikely to be revealing for research purposes.

While it is very difficult, if possible at all, to define a general relationship between a cognitive process and a prototypical scanpath pattern, there are ways to associate a cognitive process with a scanpath that is specific to a situation. One way is to use an experimental

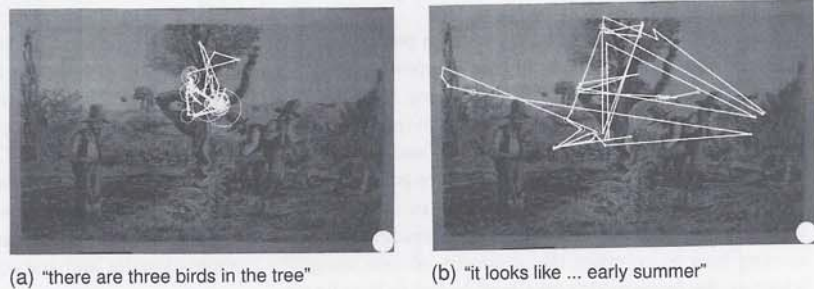


Fig. 8.20 Two scanpaths and concurrent speech. Reprinted from *Human Cognitive Processes*, 23, Holsanova, J., Discourse, vision, and cognition, Copyright (2008) by, and re-printed with kind permission from, John Benjamins Publishing Company, Amsterdam/Philadelphia, www.benjamins.com.

design that constrains the number of possible interpretations of the scanpath. This requires control over stimuli, task, and even background of participants, however. Another option is to include complementary data sources, such as speech, neurophysiological data, or body-movement data.

For example, Holsanova (2008, 2006, 2001) had participants describe complex scenes and then segmented the spoken discourse into a series of what Chafe (1994) calls “idea units”; sections of thought expressed as speech and delimited by prosodic features, speech timing, and a particular form of words called “discourse markers”. The flow of idea units in speech is considered to correspond to the flow of thought in the speaker’s mind. Holsanova then temporally matched the idea units to scanpath patterns, so that it became clear what participants looked at while speaking a particular idea.

Figure 8.20 shows two of the clearest simple patterns from Holsanova’s research. In Figure 8.20(a), “three birds in the tree”, a limited picture element, corresponded very well to the idea in speech. In Figure 8.20(b), the idea in speech, “it looks like early summer”, was not located at any particular position, but spread out in the image, at places where there is evidence of summer. These examples suggest a tight coupling between sub-scanpaths and cognitive processes in free description, complementing the more well known visual world paradigm of Tanenhaus *et al.* (1995) and others. However, it still only means that scanpaths reflect ideas, not that ideas or cognitive states can be uniquely identified from scanpaths. It also shows that the particular stimulus partly determines the scanpath, so it might be difficult to find general stimulus-independent scanpath patterns for the same cognitive processes.

What does Holsanova’s research tell us about interpreting all those scanpaths from non-experimental freeviewing tasks that have no concurrently recorded speech? Without the idea units from speech and their temporal on- and offsets, we have no method to find the on- and offset in the scanpath, i.e. event, that corresponds to the start and end of a thought, and we have no content of the thought either. Without speech or other complimentary data, we are simply left guessing.

8.7.2 Scanpath Theory

It is rare for discussions about scanpaths to circumvent scanpath theory as devised by Noton and Stark (Noton & Stark, 1971a, 1971b). From two reported studies, which use two participants in the first and four in the second experiment, the authors conclude that when a participant looks through an image already seen, the remembered spatial model (which the authors term “the feature network”) from the first viewing directs him to look at the stimulus in the

the same way the second time around. In essence the scanpath theory predicts that:

... for each pattern with which he is familiar, each person has a fixed and characteristic path which he follows from feature to feature, both when viewing the pattern and when matching it with its internal representation during recognition. (Excerpt from Noton & Stark, 1971b.)

Note that the scanpath according to this theory is a fixed, theoretical path that has a representation in the brain. The main implication of scanpath theory is that scanpaths will be re-capitulated, not being driven by image content, but by a stored internal representation.

Scanpath theory has been re-considered, for instance by Henderson and Ferreira (2004), who review a number of studies in contradiction to the predictions of Noton and Stark. In particular, participants can recognize a scene without making eye-movements (Biederman, Rabinowitz, Glass, & Stacy, 1974), and there is usually very little sequence similarity between repeated viewings by the same participant; even if looking at the same positions a second time, the order is different, according to these authors. In fact, some authors completely refrain from the term 'scanpath' in order to dissociate themselves from scanpath theory, for instance Underwood, Foulsham, and Humphrey (2009) and Henderson (2003).

On the other hand, scanpath theory is to some extent supported by empirical evidence. When recording eye movements from participants viewing pictures during encoding and later recognition, Foulsham and Underwood (2008) found higher than chance similarity between scanpaths. Moreover, *mental imagery* studies from the late 1990s and onward repeatedly find that participants shown a blank screen and asked to retell a previously shown scene, largely reiterate the same eye-movement sequences on the blank screen as when viewing the scene itself (Zangemeister & Liman, 2007; Johansson *et al.*, 2006; Brandt & Stark, 1997). In these studies scanpath theory has been replaced by newer theoretical explanations, however.

8.7.3 Scanpath planning

Research on *scanpath planning* investigates whether, when, and to what extent participants plan their scanpaths ahead, and which information is involved in this planning.

An appealing argument in favour of scanpath planning is the frequent occurrence of one or a series of short fixations (< 100 ms); considering that the time it takes to program a saccade by far exceeds these 100 ms, some of the saccades in the sequence must be pre-programmed (Becker & Jürgens, 1979).

Zingale and Kowler (1987) proposed that a scanpath results from an organized plan that is retrieved from memory before the scanpath is initiated. Using simple visual arrays, they asked participants to fixate a sequence of 1–5 static targets and found the latency of the first saccade to increase with the number of targets. Zingale and Kowler attributed the increase to the additional planning required to encode a longer scanpath.

Findlay and Brown (2006) identify three possible scanning strategies which they go on to test empirically: *systematically directional*, raster-like as in reading or back and forth scanning; *locally perceptual*, based on low-level information acquired at the current point of fixation; and *globally perceptual*, considering global stimulus features such as shape. Asking participants to sequentially fixate each item in a visual array, they found empirical evidence only for the strategies to scan items in a raster order (directional), and to use the global external contour to guide eye-movements, making saccades towards the centre, but still using the contour as a guide. Supposedly, a systematic strategy such as a raster scan can be pre-planned, but there is no need to encode the whole scanpath into memory (Gilchrist & Harvey, 2006). In contrast, a scanpath guided by global contour requires some initial global processing of the stimulus, where the scanpath is first encoded and stored in some way.

Both Zingale and Kowler (1987) and Findlay and Brown (2006) argue that it is likely

that bottom-up information can modify the scanpath, and make it deviate somewhat from the pre-planned scanpath. In visual search, for example, backtracks were found to be more common to items that resembled the visual search target (Peterson, Kramer, Wang, Irwin, & McCarley, 2001). Interestingly, using very reduced stimuli with a small set of equal-sized objects, some researchers have suggested that when there is no bottom-up information to distinguish between objects, attentional deployment is random (Horowitz & Wolfe, 2001) and no memory is retained of what objects have (and have not) been visited.

An important part of scanpath planning is to keep track of previously visited locations, and prevent revisits to these. This mechanism is known as *inhibition of return*, and is argued to serve as a 'foraging facilitator' in visual search (Klein & MacInnes, 1999), preventing inefficient returns to areas already inspected. It has been suggested that at least five spatial objects can be kept in memory (Synder & Kingstone, 2000), and that this information is used when planning future scanning (Findlay & Brown, 2006).

Other research has concluded that memory not only guides the selection of targets along a scanpath (Gilchrist & Harvey, 2000), but that this memory can also prevent *revisits* of earlier targets for fairly large sets of objects (Dickinson & Zelinsky, 2007). This research has developed into the question of whether visual search is like foraging, carried out over chunks of space that can then be successively dismissed.

Using real-world stimuli and tasks, such as letting participants make sandwiches, avoid non-animate targets, or meet other people approaching in a narrow hallway, Land and Hayhoe (2001); Hayhoe and Ballard (2005); Rothkopf, Ballard, and Hayhoe (2007), and others show that scanpath planning is tightly connected with the ongoing task and the immediate visual surroundings. In particular, *look-ahead* fixations during real-life tasks, for instance looking at the jam jar four or five seconds before it is actually time to grab it, show how the eye-movement system is used in the planning of future sub-activities in the overall task (Mennie *et al.*, 2007). In this view, task-driven plans lie behind virtually all eye movements, and random scanpaths would only occur when participants feel they have no task, and bottom-up features are free to pull the eyes—even then, however, we quickly apply meaning to what we look at.

8.7.4 The average scanpath

Although some researchers find the 'the average scanpath' difficult to calculate due to the unique unfolding of each scanpath (for instance, Hornof, 2007, p. 317), others make an attempt to take a number of scanpaths and build an average from them (Josephson & Holmes, 2002; Hembrooke, Feusner, & Gay, 2006; Torstling, 2007).

Josephson and Holmes (2002) first calculate the pairwise string-edit distance between all AOI-based scanpaths, and then define 'the most central scanpath' as the sequence with the smallest average distance to all other sequences. Using multiple sequence alignment, Hembrooke *et al.* (2006) construct the average scanpath from global similarities among the scanpaths.

A slightly different definition of the average scanpath was suggested by Holsanova *et al.* (2008). It is illustrated in Table 8.3 with a hypothetical example including five AOIs and five participants. The AOIs in this table have been ranked in the order that they were visited by each of the five participants. Participant 1, for instance, looked at AOI 1 first, then at AOI 3, etc. Revisited AOIs are not counted, only first visits. AOIs that are not visited by a participant receive the highest remaining rank or the average of the highest remaining ranks.

Finally, a sequence of attention maps can be considered to represent an average scanpath in the sense that each attention map in the sequence describes spatial distribution for all participants.

The term 'averaging' implies calculating one single entity out of several, so that the single

Table 8.3 Average scanpath as the average order of entry for five AOIs seen by five participants (fictitious data).

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

average is somewhere near the middle, and accepting that the variance information from the averaged group of scanpaths is lost. This makes the average scanpath alone unsuited for statistical calculations. Moreover, if two participants consistently look at different sides of a monitor, their average scanpath will be in the centre of it, where none of the participants have ever looked. These are severe limitations when using averaging scanpath representations and measures.

8.7.5 Comparing scanpaths

The general principles for scanpath comparison were outlined earlier in the chapter. However, many challenges still need to be addressed. At a conceptual level, there are a large number of desirable pairwise scanpath comparisons. For instance, we would like to detect the degree to which:

- the *overall shape* is the same between two scanpaths, and whether both scanpath shapes exhibit the same temporal sequence. The string-edit measure has approximated a shape comparison for some cases, while many measures completely ignore temporal order.
- two scanpaths are *similar in shape but different in scale*. A half-sized but otherwise identical scanpath should be considered similar by a measure to be useful in for instance mental imagery studies, but no current measure can do this.
- two scanpaths have a difference in *spatial extent*. This can be studied with measures utilizing attention map representations of scanpaths.
- there is a *similarity in position but reversal of order*. This could in principle be detected using the sequence alignment method described above.
- one participant executes his scanpath faster than another participant, although to the exact same positions, by investigating how the *temporal alignment* differs.
- the *fixation duration profiles* between two scanpaths differ, even if position and sequence order is identical.
- *similar sub-scans* exist in either of the two scanpaths (even though the sub-scans may appear in a different order).

Any proposed scanpath comparison measure must eventually be validated. There are two main methods for validating scanpath similarity: *absolute* and *relative*. Absolute validation can only be made by comparing the output from the proposed measure to a baseline that expresses the true similarity between scanpaths. One such baseline can be established by showing people a large number of scanpath pairs and asking them to judge the similarity for each pair on some scale. However, setting up this baseline both requires and allows us to answer many open questions: Should we show static or dynamic scanpath pairs (i.e. is dynamics a part of scanpath similarity)? Should our judges rate the similarity using a single

or multiple scales? Is human judgement of the similarity between pairs of scanpaths really systematic across individuals? Is it possible for a human to judge the degree of similarity between any two scanpaths at all?

Validating relative similarity is much easier. Take for example any scanpath and create distorted versions of it by adding an increasing amount of noise. Since we know that scanpath similarity between the original scanpath and its distorted version should increase with the level of noise, this is something that should be reflected in a valid scanpath similarity measure. Still, we cannot tell whether the absolute differences in similarity between noise levels are sensible.

Given a substitution matrix, another open issue is how to choose appropriate costs. Although the matrix offers a flexible scoring scheme, it is up to the researcher herself to choose costs that are suitable to the experiment in hand.

Finally, it is important to distinguish between scanpath comparison in picture- and video viewing; objects, and therefore fixations, in video stimuli are largely associated with a particular temporal span. Therefore, scanpaths recorded from videos do not critically require sequence alignment prior to comparison. Moreover, since the duration of a video is always fixed, different sequence lengths are not a large problem in the comparison.

8.8 Summary: scanpath events and representations

Scanpath events are specific subscans that occur within a limited chunk of a scanpath. Six such events have been defined in this chapter:

- **backtracks**, saccades going in the opposite direction to the previous one.
- **regressions**, which exist as *in-word* and *between-word* regressions, *regression scanpaths*, and *re-inspections*.
- **look-backs**, which are also known as *returns*.
- **look-aheads**, saccades towards items that are important in the immediate action plan.
- **local versus global**, a categorization of the scanpath into two types of subscans; this is very similar to *ambient versus focal*.
- **sweeps**, a sequence of saccades in the same direction.
- **reading versus scanning**, events in reading of larger texts such as newspapers.

Furthermore, a scanpath can be represented in a number of ways:

- **Sequences of symbols** aim to represent selected features of a scanpath by means of symbols. The most different types of symbol sequences are
 - * **Fixation strings** where fixations are represented by letters denoting names of the AOIs where they reside. An example of a fixation string is MMTCCCHGM.
 - * **Dwell strings**, which are fixation strings where consecutive, repetitive fixations are merged. The dwell string of the example string above would thus become MTCHGM.
 - * **Direction/amplitude strings** of saccades such as D6 23 71 28 73 B3 54, where the first hexadecimal digit in each pair is segmented saccadic direction and the other digit segmented saccadic amplitude.
 - * **Duration strings** where symbols represent quantized fixation duration. However, these have been used very sparingly in the literature.

- **The vector sequence** represents a scanpath by a sequence of Euclidean vectors. Typically, the vectors represent saccades, and the start and end position of saccade vectors represent fixations.
- **Attention map sequences** represent one of many scanpaths as a sequence of attention maps. Sequence information but no participant identity are retained in this representation.

Scanpaths and their representations are also commonly used in visualizations. They are for example useful for exemplifying data in journals, for checking data quality, and to see what the fixation algorithm did, and have often been used for manual analysis. Scanpath visualizations can be used as elicitation in retrospective speech.

This chapter also outlined the principles for scanpath comparison, which include choosing a suitable scanpath representation, simplifying the scanpath, and aligning scanpaths with each other before calculating a similarity score. When selecting between 'scanpath comparison measures', of which there are many, be certain that you use a representation that retains the information you want in the comparison.

Finally, a number of open issues that researchers dealing with scanpaths should be aware of concerning scanpaths were discussed.