6 Areas of Interest

In the previous chapter, we explained how to process data samples into events such as fixations, saccades, and smooth pursuits (Chapter 5). This chapter defines and discusses areas of interests (AOIs) as a tool for the further analysis of eye-movement data. In simple terms, AOIs define regions in the stimulus that the researcher is interested in gathering data about: did the participant look where expected, and what were the properties of their eye movements in the area looked at? But more importantly, AOIs allow further events to be defined and detected: *dwells, transitions*, and *AOI hits*; these will be introduced in this chapter. In addition, segmenting stimulus space with AOIs allows us to transform and simplify the recorded data into representations such as *strings, transition matrices*, and *proportion over time graphs*.

Using AOIs in data analysis requires a number of issues to be addressed:

- First, we emphasise the important relationship between your hypothesis and what can be done in an AOI editor (Section 6.1, p. 188).
- In Section 6.2 (p. 188), we provide condensed hands-on advice for work with AOIs.
- We then define the three basic AOI events in Section 6.3 (p. 189): *AOI hits, dwells*, and *transitions*. We also define three related events: *returns, first skips*, and *total skips*.
- Five major representations of eye-movement data, based on the subdivision of space using AOIs, are described in Section 6.4 (p. 192): *AOI strings, dwell maps, transition matrices, Markov models*, and *proportion over time graphs*. Many of these come with several varieties.
- Section 6.5 (p. 206) describes the properties and usage of several types of AOIs: dynamic, distributed, gridded, and fuzzy AOIs, just to mention a few.
- It is commonly believed that AOIs are very simple to use. However, there are a number of challenging issues concerning the usage and analysis of AOIs that are discussed in Section 6.6 (p. 216). For example, can we use AOIs that are arbitrarily small, or is there a minimal allowed size?
- Finally, the summary in Section 6.7 (p. 229) draws together the most used AOI events and representations that will follow us through the remainder of the book.

Since AOIs have been repeatedly re-invented by different researchers, and because software developers want to contrast their products with those from other manufacturers, there is no standard terminology for AOI measures, and some of the measures are referred to by up to seven different names. Sometimes the same name is used about different measures (like 'gaze duration' which can be either 'dwell time' or 'total dwell time'). Even the AOIs themselves are known under different names, such as 'ROIs' (regions of interests), 'IAs' (interest areas), and 'Zones'. We use 'AOI' because it appears to be the most established term in eye-tracking research. The unclear naming situation causes unnecessary confusion in many research papers, and a one goal of this chapter is to propose a standardized and more logical vocabulary.

Fig. 6.1 Two AOIs drawn on a stimulus background (a webpage); one a polygon (named 'Yngves') and the other a rectangle (named 'USB stick'). The AOI editor is seen on the left as a number of tool buttons. AOI editors today give the user lots of freedom to position and edit the AOIs.

6.1 The AOI editor and your hypothesis

AOI events are defined *in relation to entities in the stimulus*. In this, they contrast with fixations and saccades which are calculated on the basis of data alone; the event detection algorithms of the previous chapter know nothing about the stimulus content.

AOIs are created using a tool for spatial segmentation sometimes called 'the AOI editor' or similar. AOI editors are usually supplied in the analysis software for the eye-tracker. AOIs are always drawn against the background of the stimulus. Figure 6.1 shows an AOI editor with two AOIs drawn on top of a stimulus.

The precise segmentation of the stimulus is crucial to your analysis, and will be discussed in detail on pages 216–224. Remember that the AOIs you draw are *part of your hypothesis*, because they decide which areas in space dwell and transition data should be calculated against. This has two important consequences:

- If you alter your AOIs, you alter your hypothesis.
- If you draw your AOIs after data recording, while inspecting the data, you are forming post-hoc hypotheses.

The freedom of AOI editors allowing you to alter AOIs however you like and at anytime, must therefore be used with great care in order not to undermine the validity of the study.

6.2 Hands-on advice for using AOIs

Before you apply AOIs to your stimuli, consider the following:

- There are many measures that use the events and representations of this chapter, and each and every one of them is very sensitive to how you divide your stimulus into AOIs.
- Let your *research hypothesis decide* what AOIs you put on the stimulus. If you edit or move your AOIs, you alter your hypothesis.
- Each AOI should cover an area with homogeneous semantics, and the semantics should be founded in the rationale behind your experimental design.
- If you are free to design your stimulus, do not put objects so close together that you cannot have a margin between AOIs.

- Overlapping A
  and then the ca
- Do not distribu
  link between th
  you employ.
- When using tra
  stimulus not co
  how transitions
- Be aware of me
  content of an A
  and the baselin
- The *minimal A*
  recorded data.
- Avoid arbitrary
  to the importan
  method (such a
- Manually codi
  limited section
- Inaccurate data
  exactly what c

6.3 The basi

AOI hits, dwells, a
be calculated they r
transitions are even
time to complex on
branches of eye-tra

The following s
they should be calc
the basic AOI even

6.3.1 The AOI

The most primitive
that its coordinate
measures, including
graphs. In the righ
indicate AOI hits
taken place. Some
upon during a mi
process the inform

The fixation-b
Figure 6.2 also sh
the graph of the rig

- Overlapping AOIs should not be used unless the hypothesis and stimulus demand it, and then the calculation of first fixations, dwells, and transitions must be reconsidered.
- Do not distribute a single AOI over many areas of your stimulus, unless there is a clear link between the semantics in those areas, your research hypothesis, and the measures you employ.
- When using transition measures, report what is known as 'whitespace' (parts of the stimulus not covered by any AOIs) as a proportion over the whole stimulus. Define how transitions are calculated with regard to whitespace.
- Be aware of measures that are scaling dependent with respect to, e.g. the size or the content of an AOI. However, this must be motivated by the semantics of the stimulus, and the baseline probability of looking towards each area.
- The *minimal AOI size* is limited by the precision and accuracy (pp. 33–41) of your recorded data.
- Avoid arbitrary AOI positioning; ensure your AOIs are as precise as possible in relation to the important elements of the stimulus. For complex real-life stimuli, use an external method (such as expert ratings) to decide if your division of stimulus space is suitable.
- Manually coding dwells and transitions from gaze-overlaid videos is not intractable for limited sections of data.
- Inaccurate data (offsets) can be repaired, but this should not be done unless you know exactly what calculations to make and the consequences for your data.

6.3 The basic AOI events

AOI hits, dwells, and transitions are events in the same sense as fixations and saccades, but to be calculated they require AOIs that connect the data to stimulus space. AOI hits, dwells, and transitions are events used in a very large number of measures; from basic ones like *dwell time* to complex ones like *entropy* and the *string edit measure*. They are used in virtually all branches of eye-tracking research, from human factors to reading.

The following subsections introduce AOI hits, dwells, and transition, and discuss how they should be calculated. We then present the *return*, and the two *skip* events, derived from the basic AOI events.

6.3.1 The AOI hit

The most primitive AOI event is the *AOI hit*, which states for a raw sample or a fixation that its coordinate value is inside the AOI. The sample-based AOI hit underlies all raw AOI measures, including those based on fairly complex representations like proportion over time graphs. In the right side of Figure 6.2, dark portions of the line along the path of samples indicate AOI hits on one of the two AOIs, while where the line is grey, no AOI hit has taken place. Sometimes an AOI is not considered to have been hit until it has been looked upon during a minimum amount of time, reflecting the minimum time it takes to cognitively process the information therein.

The fixation-based AOI hit is important in many of the counting measures with AOIs. Figure 6.2 also shows (on the left side) fixations that directly correspond to the raw data in the graph of the right side.
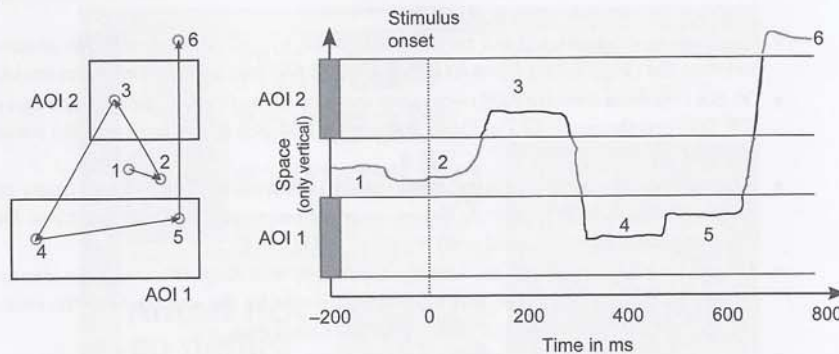
Fig. 6.2 Principle for the trials in which AOI data are calculated: stimulus space is divided into two AOIs, shown on the left side, with fixations from 1 to 6. Trial time is divided by a stimulus onset (at time 0), shown by a dashed line. The space–time diagram on the right side uses black lines to indicate the idealized path of raw data samples over the AOIs, corresponding to the fixations on the left side.

6.3.2 The dwell

The second AOI event is *the dwell*—often known as 'gaze' in reading and 'glance' in human factors (Green, 2002)—which is defined as one visit in an AOI, from entry to exit. Figure 6.2 (right) shows the raw data samples included within an AOI as black segments, and those samples outside AOIs as grey segments; a whole black segment equals a dwell. The dwell has its own duration, starting point, ending point, dispersion etc., as it is in several ways similar to a fixation, but a much larger entity both in space and time.

Dwell data represent sample data at a coarser level. If you only know that there is a dwell in the AOI, it can refer to any of the data samples; or in other words, you have lost information about the precise positions of samples. In return, you can categorize the AOI, giving its spatial extension a name that has a meaning to your experiment.

6.3.3 The transition

Another well-known event is *the transition*, also known as 'gaze shift', which is the movement from one AOI to another. For instance, any eye movement between text and graphics in a study of textbook reading counts as a transition. The notation for a transition between AOIs I and E is IE, in typewriter font. When AOIs have names with multiple letters, such as AOI RF and AOI LF, the transition is written RF LF. The predominant exception in the usage of the term transition comes from reading research, where researchers typically do not describe the movement from one word to the next as a transition but as a forward saccade.

Transitions are similar to saccades; they traverse spatial locations, they could have some sort of duration, an amplitude, and latency measures could be built from them. But transitions can be larger entities than just one saccade, since the transition can move from one AOI to another via fixations in parts not covered by AOIs (Figure 6.3). Indeed, any intermediate portion of the raw data sample in between AOI visits may be counted as part of the transition—the grey segments of the line in Figure 6.3. Note the two dubious cases, however: the term transition is usually reserved for the change in gaze allocation between one AOI and another, rather than exit and re-entry to the same AOI. Also, is it correct to characterize a movement as a transition even between two AOIs when several fixations have been made along the way?

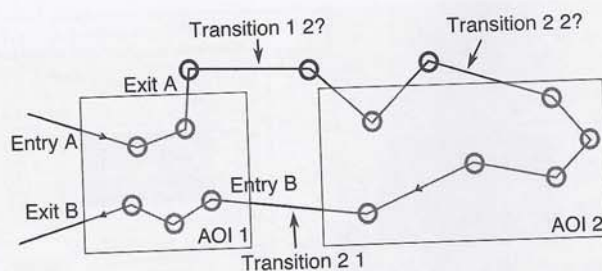A saccade within an AOI is not and should not be called a transition. Sometimes you may

Fig. 6.3 Two AOIs 1 and 2. There is clearly a transition 2 1, but should the movements marked as transitions 1 2 and 2 2 count as transitions? Transition 2 1 is an entry and a return to AOI 1.
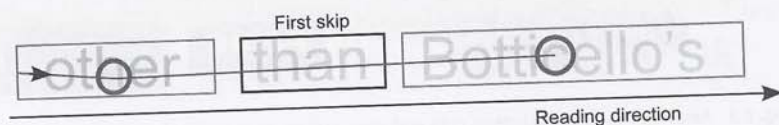

Fig. 6.4 An AOI (word) first-skipped: a later word was looked at without first looking at the skipped AOI.

see "within-AOI transitions" in transition matrices and other measures, but they confuse both the concept and the statistics. They are and should be called "within-AOI saccades".

Studying text–image integration, Stolk and Brok (1999) differentiate between *one-way* and *two-way transitions*. One-way transitions occur when a text has been finished and the graphics is then read, but there is no connected return. Two-way transitions back and forth between the two modalities were taken as indications of actual integration.

Obviously, the particular division of the stimulus space into AOIs is crucial to all transition calculations, and may in fact determine whether a result will be significant or not. It is therefore of the utmost importance to motivate the choice of AOIs based on hypothesis, task, and stimulus, before making the transition calculations to be presented as results.

6.3.4 The return

The *return*, also known as 'revisit', is a transition to an AOI already visited, exemplified in Figure 6.3, but the event exists in a version without predefined AOIs also. In research on radiology, it has been operationalized as the event when the eye strays further than 2.5° of visual angle—the approximate area of acute foveal vision—from the centre of any previous fixation and then comes back within that circle.

6.3.5 The AOI first skip

The *AOI first skip* event assumes that the AOIs are ordered, as in reading, and that more or less all AOIs are looked through. An AOI (that is, the word) is taken to be first-skipped if the eye of the reader lands on a later AOI (word) before landing on the word itself, as illustrated in Figure 6.4. The first-skip status of the AOI is not changed if the reader immediately regresses back to it. It was still skipped first, and will remain so.
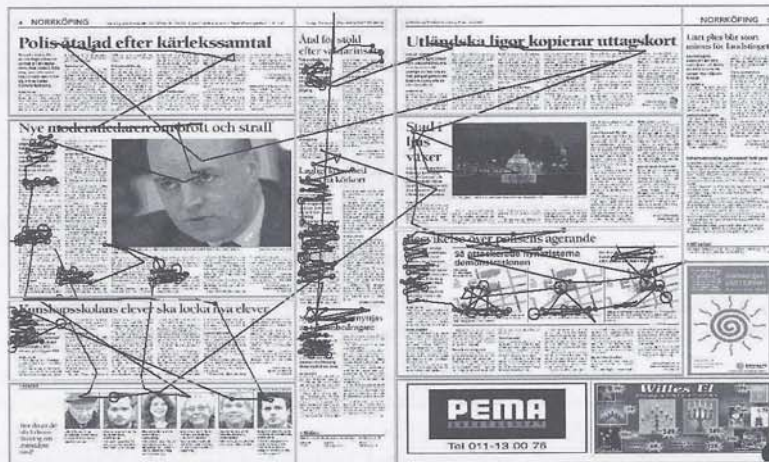
Fig. 6.5 Advertisements at the bottom right and short articles on the right totally skipped. Recorded with a 50 Hz head-mounted eye-tracker with head tracking.

6.3.6 The AOI total skip

The *total skip* status is given to an AOI that a participant does not look at for the entirety of a trial. While the AOI first skip is a dedicated reading event, the AOI total skip is a very general event, since it does not presuppose conventionalized order. For instance, in newspaper reading, an AOI total skip occurs for those AOIs (for instance advertisements) that were never looked at. Figure 6.5 shows an example of data where some parts of the newspaper spread have been totally skipped.

Figure 6.5 also illustrates that skipping may be too coarse a measure for semantic AOIs with a large coverage. Even the non-skipped AOIs are not very much read, which we can quantify with the more flexible reading depth measure (p. 390).

6.4 AOI-based representations of data

There are five AOI-based representations of eye-movement data that many measures make use of. The first is the dwell map, a gridded AOI with dwell time in the cells. Then there are the AOI strings, the most used sequence representation. The third is the transition matrix, which tells us how frequent transitions were between any combination of AOIs in our stimulus. The fourth are the Markov models, the probabilistic variety of the empirical transition matrices. The last are the proportion-over-time graphs and the other AOI over time representations.

6.4.1 Dwell maps

A *dwell map* is simply a list of all AOIs with dwell time (p. 386), as illustrated in Figure 6.6. If gridded AOIs (p. 212) are used, the dwell map can be superimposed onto the stimulus, with the dwell time value filled into each cell in the grid. Although calculated from dwell time data in the gridded AOIs, this simple representation can also be seen as a downsampled heat map, as will be evident in the next chapter, and is of great value in several position dispersion and similarity measures.

0.9	4.0	2.3	2.2	0.1	0.2		
1.4	1.4	4.5	9.4	9.8	3.3	1.3	
0.8	0.4	1.0	10.8	11.8	2.4	0.5	
0.4	0.9	1.7	2.9	2.4	0.8	0.1	
	0.2	0.3	0.9	1.5	0.7	0.1	
	0.2				0.2	0.8	
							0.3
							0.1

(a) Dwell map: a gridded AOI with dwell time filled into cells.
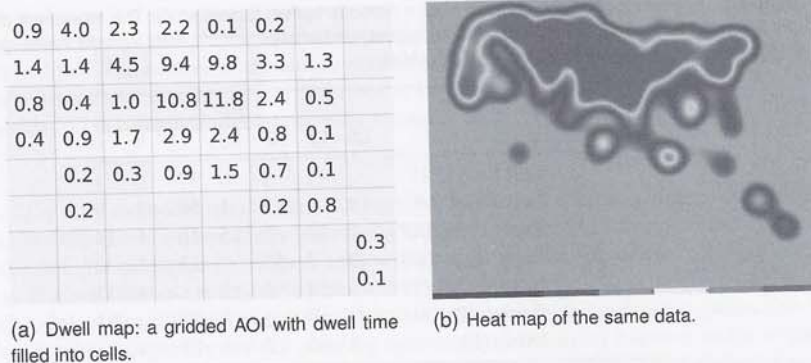
(b) Heat map of the same data.

Fig. 6.6 Dwell map versus heat map for the same data produced by BeGaze 2.4. The dwell map shows dwell times in seconds.

6.4.2 The AOI strings

The *AOI string* is a sequence of either fixation-based AOI hits or dwells in the order of occurrence. There are at least three different varieties:

1. In the string MMTCCHGM, each letter is a *fixation in an AOI* with the name of the letter. Each fixation is included, so we have cases of repetition in the string.
2. A *compressed* string consist of *dwells only*, which means that repetitions are removed when sequences of fixations within the same AOI are collapsed into a single dwell, to give MTCHGM.
3. A string with *first entries* only lists a dwell on the first entry into the AOI. This means that each AOI appears in strings only once, as in MTCHG, and that the maximal string length equals the number of AOIs in the stimulus.

We will write AOI strings as MTCHG when all AOIs have single letters as names, and as A6 C5 F0 I1 J1 K2 I3 or 1 3 4 in all other cases.

There are three applications of the fixation- and dwell-based AOI strings, according to Privitera (2006), each corresponding to a different timescale:

Full history analysis Using the string edit measure for calculating *scanpath similarity* (see p. 348).

Short history analysis Calculating *transition similarity*—a Markov model of transition matrices (p. 193).

No history analysis Calculating *locus similarity* as defined by the number of AOIs shared by both strings, independent of order, and as such could be a coarse pairwise position similarity measure (p. 370).

6.4.3 Transition matrices

A *transition matrix* is a full catalogue of all AOI sequences of length ℓ equal to the dimensionality of the matrix. A fictitious two-dimensional transition matrix is shown in Table 6.1, which shows the transitions between different parts of a machine control panel that the operator looks at. In this transition matrix, the AOIs are listed in rows and columns, and the number in each cell indicates how many times gaze has shifted from one AOI to another. For instance, after the operator looked at the left side (LS) of the dashboard, he often moved his gaze to the left front (LF) part (77 times) and only rarely to the right front (RF) part (3 times).

Table 6.1 A length-2 transition diagram from a fictitious human factors study. The movement direction Left Side to Left Front (LS LF) has scored the largest number of transitions. Dots indicate structural zeros, illegitimate cells representing saccades inside AOIs.

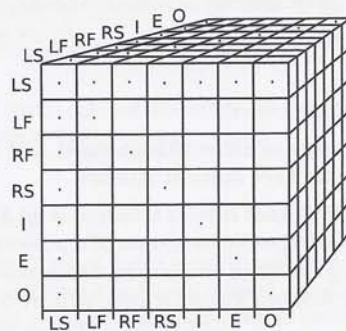|		To						
		LS	LF	RF	RS	I	E	O
	Left Side (LS)	.	77	3	0	17	0	1
	Left Front (LF)	18	.	14	1	56	2	9
	Right Front (RF)	1	52	.	15	16	1	14
From	Right Side (RS)	0	7	35	.	13	15	30
	Instruments (I)	3	54	2	1	.	4	37
	Engine (E)	0	9	0	3	27	.	61
	Other (O)	2	60	2	5	27	4	.

Fig. 6.7 Visualization of a 3D transition matrix for studying the frequency of AOI strings of length 3. Dots again mark structural zeros. The empty cells fill with the number of specific sequences of length three.

Note that the matrix ignores saccades within the same AOI, which is why the values on the diagonal of the table are all so-called structural zeros—they can never be larger than zero. Structural zeros are fundamentally and statistically different from cells that contain sampling zeros, such as transitions from left side (LS) to right side (RS), which were not observed during the study, but could have been. It is important that the software producing values for transition matrix cells clearly distinguishes between the many structural and the many real zeros. Note that some older software replace structural zeros with the number of within-AOI saccades.

Transition matrices can easily be extended to encompass longer sequences. For studying the prevalence of sequences of length 3, a three-dimensional (3D) transition matrix is formed. Continuing with the same fictitious example, Figure 6.7 shows the principle of a 3D transition matrix. Each cell corresponds to, not a transition, but a subsequence such as RS RF LF, a substring of the total AOI string produced by a participant.

There are three complications we need to observe when entering the higher dimensions. First, occurrences of strings shorter than 3 are counted twice or more. For instance, given the AOI string I RS RF LF I LF RF of one participant, the one and same RF LF transition of length and dimensionality 2 appears once in the two length-3 subsequences RS RF LF and once in RF LF I. This means that values in transition matrix cells cannot be straightforwardly compared between dimensions. Second, in the 3D transition matrix, there are structural zeros not only along the central diagonal, but also on the sides and along the corners. This is because

any cell that represents an immediate repetition, such as RF RF LF or IEE, but not IEI, is disqualified. The number of remaining valid cell values (N) in a transition matrix from n AOIs and with a dimensionality of ℓ is described in Equation (6.1). The authors have verified this calculation in computer simulations with n ranging from 2 to 12 and ℓ from 3 to 8.

$$N = n(n-1)^{\ell-1} \qquad (6.1)$$

The third methodological issue with transition matrices for longer strings is that the number of cells grows exponentially with string length ℓ. Unless enormous amounts of data are recorded, the vast majority of these cells will be empty, and many others will have low values. Not only do these empty cells make it more difficult to achieve statistical power, the zero-inflated data also contorts the statistical distribution. As explained in detail on pages 339–346, this exponential growth and the resulting sparseness of transition matrices as the sequence length increases can be dealt with in very different ways. First, by using probabilistic methods such as Markov chains, which convert the transition frequencies to probabilities and ignore probabilities at or close to zero. Alternatively, one can count only a limited number of the most frequent transitions, the 20 most frequent, for instance. Third, categorizing sequences (cells in the matrix) into a few meaningful groups is also a way of dealing with exponential growth.

The usefulness of transition matrices for strings longer than 2 can be disputed. For instance, Harris (1993) and Pieters, Rosbergen, and Wedel (1999) found no effects in higher-order Markov models, concluding that free-viewing is a reversible first-order Markov process. Using a 3D transition matrix to study scanpath sequences of multiple lengths in an air traffic control weather station, Ahlstrom and Friedman-Berg (2006) found that the most common sequences across all participants were single transitions of length 2, but very few longer sequences in the data were common over all participants. This means that a first-order Markov model—essentially a 2D transition diagram—governs the scanpath in these situations. In fact, we have found no studies reporting longer sequences than four AOIs even though, in theory, there are situations where longer sequences would be interesting to study.

It is usually taken for granted that a transition matrix is always a 'change of position' matrix, in that it counts transitions between AOI positions. However, Ponsoda *et al.* (1995) developed a *change of direction matrix*, based on their segmentation of saccadic direction in Figure 10.2(c) on page 302. They argue that a full 8×8 transition matrix would be inappropriate due to high standard errors in matrix cells when small data sets are used, and proceed with a 2×2 matrix, using only horizontal and vertical directions.

Usage of transition matrices

Transition matrices are flexible representations that have been used in many research fields. In human factors, Itoh, Hansen, and Nielsen (1998) constructed a model of ship navigation using a transition matrix in combination with dwell time analysis. Moray and Rotenberg (1989) found that instruments were fixated more frequently after a plant failure, but that dwell times were unchanged, and that operators tend to deal with multiple disturbances sequentially. Morrison, Marshall, Kelly, and Moore (1997) used a transition matrix to investigate whether different decision strategies can be visible in transition matrix results from interactions with military decision support displays.

Cook, Wiebe, and Carter (2008) investigated learning from displays with multiple representations of osmotic cell transportation. They showed that low prior knowledge students transitioned more frequently between macroscopic and molecular representations, interpreting this as evidence of a higher difficulty in coordinating the representations. The 'eye movement matrix' of Hyönä, Lorch Jr, and Rinck (2003) uses sentences as AOIs, and is one of the methods for studying global text reading. The likely but hypothetical examples given by

Hyönä *et al.* show that they can be useful in the study of inconsistent texts. Lastly, Holmqvist, Holsanova, Barthelson, and Lundqvist (2003) used transition matrices to compare internet newspapers.

The relation between dwell time and the number of transitions

Ellis and Stark (1986) point to the possibility that by chance alone, there are *more transitions between AOIs with a higher dwell time*, simply because the gaze is more often there. It is clear, however, that this correlation between dwell time and number of transitions is more likely with some stimuli and tasks than with others. In newspaper reading, for instance, the number of transitions a participant makes between two texts that he has read for 2 minutes each cannot be 24 times higher than the number of transitions between texts that he has read for 5 seconds each. The texts are separate and unrelated units, and there is no need to look back and forth between them to solve the task of understanding the news. But in human factors studies, in particular surveillance tasks where participants look at radar or instruments in a cockpit that all play a part in the task, it is much more likely that dwell time influences the number of transitions, simply because such tasks require many integrative transitions.

Ellis and Stark formally differentiate between three cases with different base probability of transitions:

Random Each AOI has the same dwell time, which is interpreted as the same probability of being fixated. The expectation is that transitions will be equal between all pairs of AOIs. This could possibly occur during task-free viewing of random scenes.

Stratified random All AOIs have different dwell times. The probability of transitions from/to an AOI is proportional to its dwell time. This is the case with the human factors task where information from all AOIs contributes to solving a single task.

Statistically dependent As in our newspaper example, the probabilities for transitions cannot be calculated from the dwell times in the AOIs. Such a situation can be modelled by a Markov chain with states for the AOIs and their dwell time, and transition probabilities for the transition frequences.

6.4.4 Markov models

Markov models are related to transitions matrices, but there are important differences. First of all, while a transition matrix is only a descriptive summary representation of collected data, the exact same numbers in the cells of a Markov model are assumed to be *the probabilities* for each transition. That is, the highest transition probability in a Markov model indicates the most probable sequence of two or more AOIs. In other words, Markov models can be used to examine the stochastic processes underlying observed transition sequences, and to explore the goodness of fit of a predicted model. Another difference between Markov models and transition diagrams is that Markov models can include the dwell time between transitions in the probability model.

Markov models exist at several levels, known as orders, which directly correspond to the dimensionality of transition matrices. The zero-order Markov model would be the dwell map for the set of AOIs. The first-order Markov model corresponds to the standard 2D transition matrix, that is, probabilities of movements between the cells of the zero-order Markov model. The second-order Markov models describe the probabilities of all triple-AOI sequences, i.e. all strings with a length of three AOIs. Higher-order Markov models can model even longer sequences of AOIs. In practice, however, Markov models higher than second order (three AOIs) are hardly ever used.

(a) AOI sequence chart with five AOIs and data from one participant.

(b) Scarf plots from fictitious five participants and the same five AOIs as in Figure (a). Participant 1 here has the same data as shown in Figure (a).
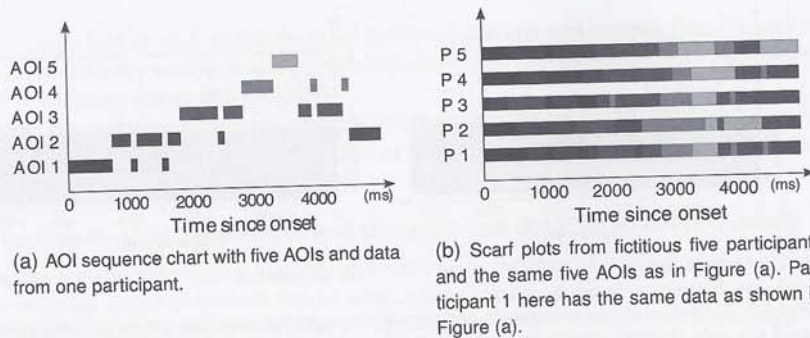
Fig. 6.8 Sequence chart and scarf plot.

Markov models have been used for very basic research on scanpath planning. For instance, recording data from airline pilots, Ellis and Stark (1986) compared statistically the empirical transition matrix to a first-order Markov model derived from zero-order probabilities (dwell times), noting the model fits the data better at the first-order level than if only comparing dwell times. Harris (1993) reanalysed the data from Buswell (1935), finding that they are readily modelled by a so-called stationary, reversible first-order Markov model. This result, replicated by Gordon and Moser (2007), Epelboim and Suppes (2001), and Pieters *et al.* (1999), can be interpreted as showing that the probability of fixating an object depends significantly on the object of the immediately preceding fixation, but not on the objects fixated further back in the scanpath.

Hidden Markov models include hidden states that may correspond to theoretical entities. Studying consumer brand awareness, Van Der Lans, Pieters, and Wedel (2008) use a model with two states, "localization" and "identification".

6.4.5   AOIs over time

There are many *AOI over time* representations, the best known of which are the *proportion over time graphs*. They represent time more accurately than transitions (p. 205). AOI over time representations use a time line which starts usually some time before the introduced and measured effect and stops at a point in time where the effect is likely to have dissipated. By observing and analysing the changes in the attended AOIs, the presence and actual parameters of the effect in question can be measured.

Each measure that can give values over successive points or windows in time along the sampled eye-movement data can produce value-over-time graphs. Figures 6.8 and 6.9 show how this is done with the binary measure AOI hit. Figure 6.8(a) shows an AOI *sequence chart* (also 'order versus time diagram') with five AOIs and data from one participant. The sequence chart shows the order and duration of dwells to each AOI. This participant starts looking at AOI 1, and then looks at AOI 2. After two short returns to AOI 1, the participant continues to AOI 3 and so on.

In Figure 6.8(b) we have collapsed the sequence charts of each participant to one line, and thus formed a *scarf plot*. The *scarf plot* is a condensed version of the AOI sequence chart, where the AOIs of each participant have been placed on a single line, so as to form what looks like a scarf, as in Figure 6.8(b), where each colour in the scarf refers to a unique AOI. With multiple participants, scarf plots allow visual comparisons over several participants.
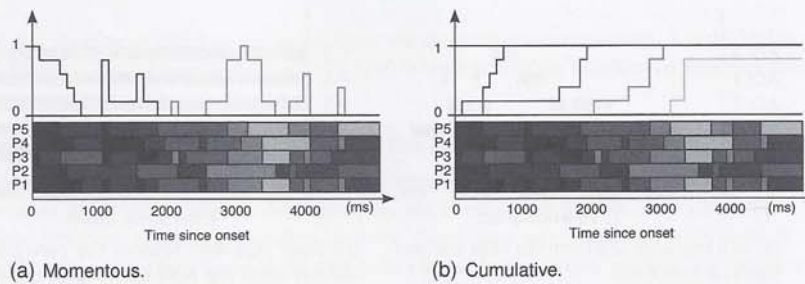
(a) Momentous.   (b) Cumulative.

Fig. 6.9 Illustration of how momentous and cumulative proportion over time graphs are calculated from scarf plot data. Bottom: Scarf plots for five participants. Top: The proportion over time graphs for two of the five AOIs.

Whereas the scarf plot is excellent for visualizing the AOI behaviour of a particular individual, it is less apt for identifying a particular AOI trend that is present, but scattered in many participants. For this case, it is appropriate to make a line graph showing the proportion of participants gazing at a particular AOI at a given point in time. Figure 6.9 plots the data from Figure 6.8 as proportion values over time.

Sampling frequency of data in the fictitious example in Figure 6.9(a) is however very low, only 10 Hz, and the number of participants in actual research is much larger than five. This makes the actual proportion over time graphs much smoother than the ones shown here, but the principle is exactly the same. Additional smoothing typically improves visual interpretation of the graph.

Proportion over time graphs have two varieties. The *momentous proportion over time graph* shows the differences in gaze behaviour for *one moment at a time*. It is very simple to calculate as the average over a scarf plot, as illustrated by Figure 6.9(a). An AOI having a large area under the curve in a momentous proportion over time graph also has a high average *dwell time*, and vice versa. This is because the proportion of samples that fall within an AOI during a time window will sum up to the total amount of dwell over all participants in the same time window.

For studies where it is important what proportion of participants have *so far seen* an AOI, the *cumulative proportion over time* diagrams can be a good tool. For each AOI, if it has been seen by a participant, then that AOI is marked as seen for the rest of that participant's data. The cumulative proportion over time graphs provide for latency measures closely related to the *entry time measure* (p. 437), expressing what percentage of participants have entered an AOI over time. What is known as 'hazard curves' and 'survival probability analysis' (Yang & McConkie, 2001; Hirose, Kennedy, & Tatler, 2010) plot the probability over time that an AOI will survive in the sense of not being hit by any fixations. As such, they are the mathematical inverse of cumulative graphs.

Proportion over time graphs were originally used with *fixed trial durations*. This means that exactly the same participants contribute at the beginning as in the end. When recording data with *variable trial durations*, this means that more participants contribute in the beginning of the graph than towards the end. In cases where the experimental design requires trials to be of different lengths for different participants, we need to choose either:

- For each raw data sample, count the number of participants that look at each AOI, relative to the total number of participants at the onset of the trial. Curves will generally decline as participants finish with trials and drop off.
- Count the proportion of participants, relative to the remaining number of participants

who look at each AOI at the point in time of that raw data sample. Data for each AOI will always sum to close to 100%, and curves will not decline, but they are based on a declining subset of participants.

A word of caution: these proportion-over-time graph only tells us *that* a proportion of participants look at, or have looked at, the AOI, not *what* proportion looks there. Participants may take turns.

The time-locking hypothesis, and the order and duration of the processes

Before we can enter into a discussion on the many varieties of AOIs over time, we need to ask what they can be used for. In what is now known as the visual world paradigm of psycholinguistics, a hypothesis soon emerged that points out that eye movements appear to be *time-locked* to the linguistic and cognitive processes being studied (Tanenhaus *et al.*, 1995; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Meyer, Sleiderink, & Levelt, 1998). Time-locked means that the development of a proportion over time graph coincides with the development of the linguistic processes of the task.

As an example, Allopenna, Magnuson, and Tanenhaus (1998) investigated the lexical activation of words in competition with cohort words and rhyming words. For instance, "beaker" will compete with words which begin with the same sound, such as "beetle", and also words which rhyme, such as "speaker". Allopenna *et al.* wanted to test competing computational models of lexical activation, but the precise theoretical applications are not important here. The important issue is that the models made predictions in terms of the activation of each word over time. These activation curves could be plotted against time, and compared to proportion over time curves from eye-tracking data both visually and statistically. So, to get proportion over time curves, Allopenna *et al.* set up an experiment with stimuli as in Figure 6.10(a). Participants were given instructions such as "Look at the cross. Pick up the beaker. Now put it above the square". The hypothesis is that about 200 ms after the sound "beaker" starts, the participants gazes will move to the most likely objects; namely those two that start with the "bee..." sound. After a few more milliseconds, the pronunciation of "beaker" has reached the "..ker" sound, and then gazes on the cohort "beetle" will drop. After data were recorded, the proportion over time curve in Figure 6.10(b) was calculated.

The recorded proportion over time curve turned out to be almost identical to the predicted activation curve of the model, which illustrates how the development of the proportion over time graphs can be time-locked to associated linguistic processes. If this result were to be a general principle of eye movement coupling to linguistic and cognitive processes, one that holds for a variety of tasks, proportion over time graphs would be a valuable tool for investigating many aspects of both language and general cognition.

Before we start to examine this approach more generally, there are some things we need to consider. First of all, the time-locking hypothesis is a close relative of the eye–mind hypothesis of page 378, claiming that processing of words during reading goes on for exactly as long as the duration of fixations, which is now known to be not fully correct. Moreover, in psycholinguistic research, where these graphs originate, participants always hear speech, the same speech, developing at the same speed for all participants. When psycholinguists use proportion over time curves and exploit the time-lock between linguistic processes and eye-movement processes, they use speech that is synchronized with particular moments in the trial, to guarantee that all participants are presented the same speech at the same time. If we were to use proportion over time curves for studies where no speech synchronization occurs, we may get very variable and noisy graphs as the eye-movement effects belonging to specific linguistic and cognitive processes are spread out over the trial instead of occurring at a distinct moment for all participants, resulting in a flatter curve with no obviously distin-
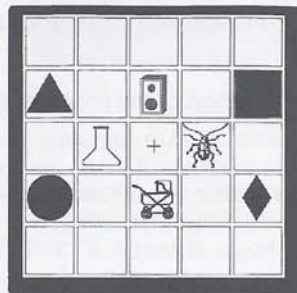
(a) Stimuli used while participant heard sentences such as "Pick up the beaker".
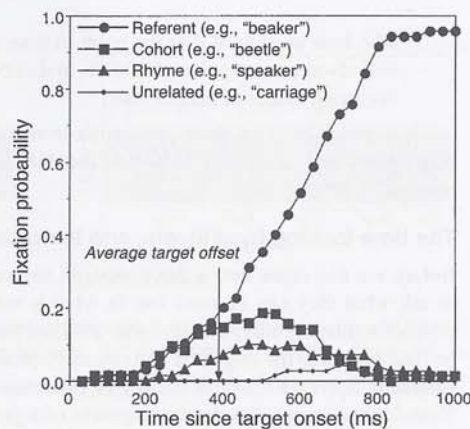
(b) Proportion over time graph from onset of target (e.g. "beaker"). The curve is almost identical to the predictions of the lexical activation model TRACE. Because they are so similar, we only show the data.

Fig. 6.10 An example of proportion over time graphs that could be time-locked to linguistic processes. Figures are reprinted from Allopenna *et al.* (1998) with kind permission from Elsevier Limited.

guishable peak. This risk increases if we cannot guarantee that the processing starts almost immediately when we present the synchronized part of the stimuli. In psycholinguistic research, proportion over time curves only stretch over fairly short time periods, from around 1 s (Allopenna *et al.*, 1998) up to around 2.4 s (Andersson, Henderson, & Ferreira, 2011) . In our example above, Allopenna *et al.* (1998) analyse data over a single up to a few seconds. What if we were to give participants tasks that range over 10 or 20 seconds with only an initial synchronization? Would we be able to see any common gaze pattern between different participants after more than the first second?

Lexical activation processes are not only fast, they are also very automated with little or no conscious deliberation involved. If we have a much more complex task, such as mathematical problem solving, is it not a risk that individual participants will each have their own pace and carry out the task using their own particular strategy? If so, if we calculate the averages taken at specific times over *all* participants, these averages will not reflect that some people are only at stage x_1 at time t_1 while others are at stage x_2. At any given time we will therefore have averages collapsed across participants at different stages, and the proportion over time graph will be very hard to interpret.

Taken from our mathematical problem solving task (p. 5), the proportion over time graphs in Figures 6.11 and 6.12 use data at 1250 Hz, and show the first 20,000 samples, or about 16 seconds. Participants were shown a mathematical task, the 'input AOI' in written text for 5 seconds, and then four alternatives for solutions appear, while the input remains. The graphs start at the onset of the four alternatives. Participants were either students of mathematics ($n = 21$) or students in the humanities ($n = 24$). The five curves do not sum up to 1 (i.e. 100%) because the AOIs cover only a portion of the monitor, and blinks cause further non-AOI time.

Visual inspection of proportion over time graphs produced from longer trials provides a valuable tool for explorative analysis of eye-movement data. It can be seen that at the onset of the four alternatives (time 0 in the graphs), no participants were looking at anything else
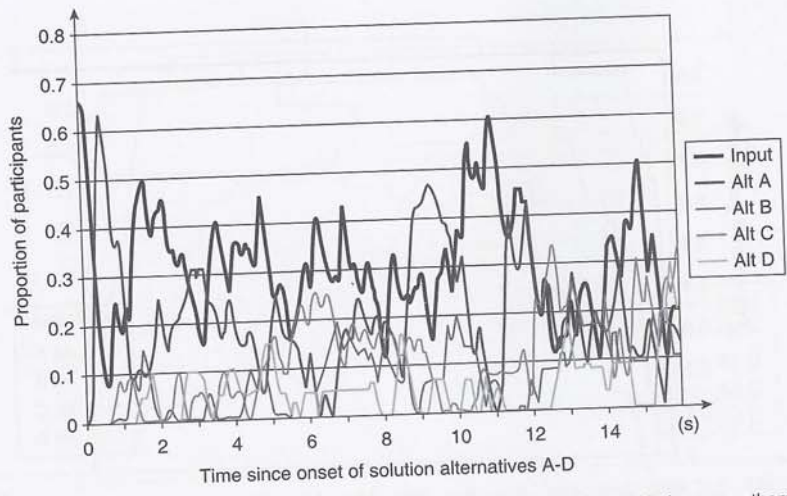
Fig. 6.11 Proportion over time graph of how *humanities* students look at the AOIs in one mathematical task. The AOI 'Alt A' is the correct alternative. The graphs have been smoothened by an averaging filter of 200 samples (160 ms).
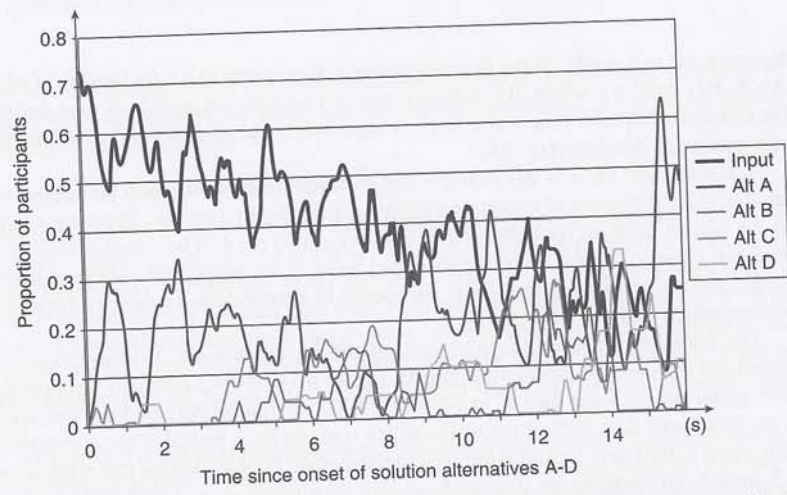

Fig. 6.12 Proportion over time graph of how *mathematics* students look at the AOIs in one mathematical task.

except the input AOI. At the onset, virtually all humanities students jump directly to the A alternative, while the mathematics students remain longer at the input with only few looks at the A alternative. This difference in visual behaviour between the groups could be interpreted as evidence of two different strategies to solve the task: the mathematicians appear to try to derive the solution from the input, while the humanities students seek help to understand the task by looking at the possible solutions. Granted, the mathematicians also look at the A al-
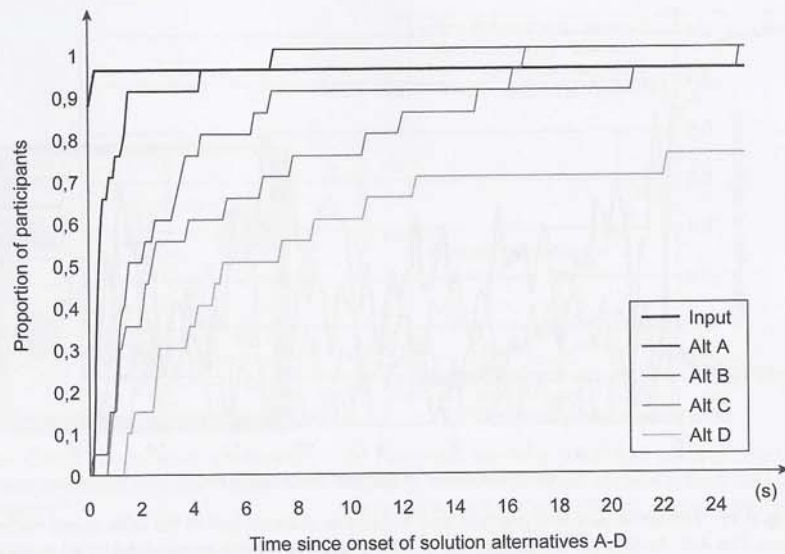
Fig. 6.13 Cumulative proportion over time graph of how humanities students look at the AOIs. Same data as in Figure 6.11, but not smoothened.

ternative, but to a much smaller degree (peak at 0.3 compared to 0.6 for looking at Alternative A). At this point we would like to have a way to calculate the periods in time during which the difference between proportion levels is significant, and when it is not. Such methods are only just being developed (p. 440).

It is much less clear in this example that participants in either group are in sync with one another in their process of comparing the input with the alternatives. There are some pulses of 1.8 second duration for both groups, but they would need to be substantiated by another analysis. Clearly, there is more variation and noise in the proportion over time graphs over longer periods of data where there is not speech to synchronize, because of variation in the precise timing between participants.

The cumulative proportion over time graphs are much easier to interpret over longer recordings than are the momentous graphs. Figure 6.14 shows a cumulative proportion over time graph for the same data as in Figure 6.12. It can immediately be seen that after about 2 seconds, more than 90% of the mathematics students have looked at the alternative A, but only about a third have looked at the second alternative. We can also see that while after about 3.5 seconds all mathematics students have seen the alternative A, the other three alternatives are seen only by 70–86% of them. Some 14% of them appear to have considered the A solution to be so obviously right that they did not even bother to look at any of the others. In contrast, the humanities students much earlier have much higher 'so far seen' proportions on the incorrect answers (Figure 6.14), in particular during the 2–10 second period after onset.

The time until 50% of the participants have seen an AOI—the T50 latency measure on page 438—can be read directly from cumulative proportion over time graphs. For the humanities students, T50 is 0.4 s for Alt 1, and 1.5 s, 2.4 s, and 5.2 s for the other alternatives. For the mathematics students, T50 latencies are much higher.

If we combine what we see in the momentous and the cumulative proportion graphs, we see plenty of evidence that the humanities students immediately go to the alternatives to seek
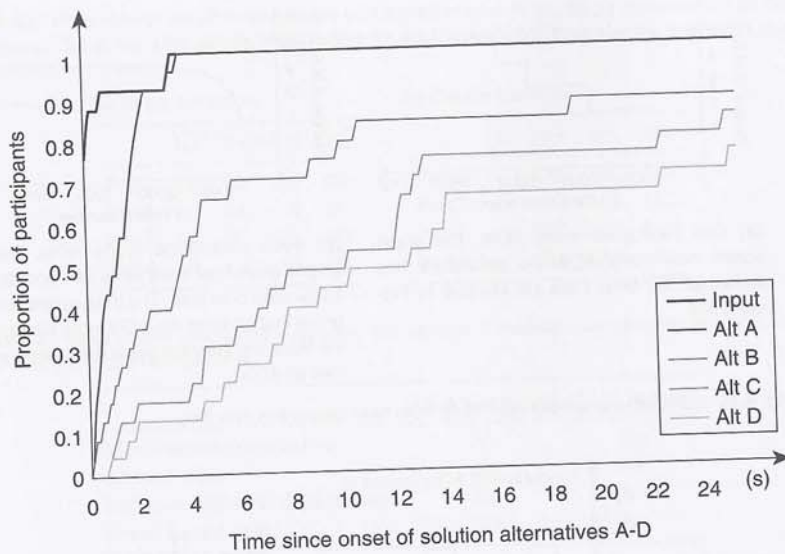
Fig. 6.14 Cumulative proportion over time graph of how mathematics students look at the AOIs. Same data as in Figure 6.12, but not smoothened.

clues for the solutions, while the mathematicians to a higher degree adopt a strategy of first thinking thoroughly about the input, and then directly selecting the correct alternative, not bothering about the incorrect ones.

Proportion of seen AOIs over time

Suppose that we were interested in finding out how many AOIs a single participant has seen over time. Figure 6.15(a) shows a graph of such data for a single participant. This participant progresses steadily through the AOIs, and after 3500 ms all AOIs have been seen. When using more participants, we can either stack each such participant graph on top of the others, to make a three-dimensional graph (not shown), or we can collapse the participant dimension by averaging their data. We would then arrive at a graph like Figure 6.15(b), in which one (fictitious) group of participants progresses through the AOIs much more rapidly than the other.

Both types of graphs in Figure 6.15 are cumulative, that is they sum up the number seen so far. Momentous graphs of seen AOIs over time are measures of dispersion, in that they show how spread out versus focused participants are over time. They differentiate from the spatial dispersion measures on pages 359–370 in that the underlying AOIs can be semantic units. A hypothetical measure based on these graphs would therefore reflect semantic dispersion, and would be particularly appropriate when using dynamic or distributed AOIs for which semantics can be very different from spatial extension. Note that the probability of a middle value in Figure 6.16 is always higher than the probability of a high or low value, and that if the number of participants is fewer than the number of AOIs, the graph can never reach the value 1.
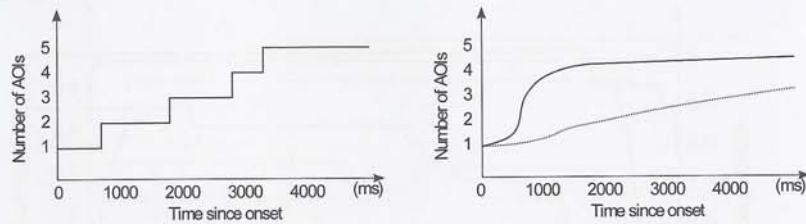
(a) One participant, many AOIs. The graph shows how many AOIs the participant has looked at over time. Data are identical to Figure 6.9(a).

(b) Many participants, many AOIs. The two graphs show how many AOIs participants have so far seen over time. The dashed line shows a group that explores the AOIs more slowly, while the black line shows a group that quickly examines all AOIs.

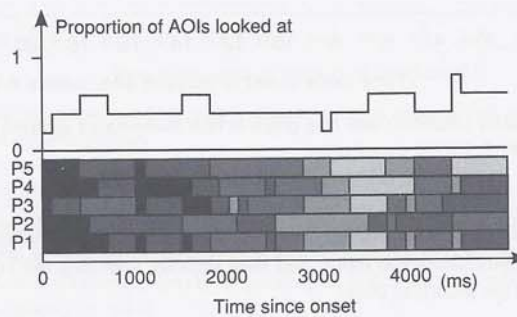**Fig. 6.15** Cumulative graphs of how many AOIs have been seen over time.

Fig. 6.16 The proportion of AOIs looked at by five participants is shown at the top, based on data in the five scarf plots below. At around 3000 ms, all participants look at the same AOI, so the proportion sinks to 20%. At around 4700 ms, they look at four different AOIs, and the proportion value is 80%. The same fictitious data as in Figure 6.9.

Proportion of transition sequences over transition time

The proportion of transition sequences over transition time is a useful representation if there are large differences in dwell time across participants, and you want to measure the development of transitions over time, accepting a shift between participants' data, in opposition to a strict time-locking hypothesis. For instance, in our example with momentous proportion over time graphs in Figures 6.11 and 6.12, participants' differing dwell times on AOIs may blur the pulse effects from being seen in the graph. When using transition time, the duration of dwells are ignored, and we can see sharper results.

To form the representation, align the lists of transitions for each participant, so you have everyone's first transition next to everyone else's first transition, and so on. Then simply count the number of transitions of each type at each time. In Table 6.2(a), AB means 'from AOI A to AOI B'. 100% of the first transitions were from AOI A to AOI B, and 75% of the second transitions went back to AOI A again. The measure can easily be developed for use with longer sequences of transitions. Table 6.2(b) shows how sequences of three consecutive AOIs are distributed across the participants and over transition time.

The transition over transition time data could be drawn in the same way as a normal

Table 6.2 Proportion of transition sequences over transition time. In (a), single transitions, in (b) double transitions. Out of four participants, we calculate for each transition in the order the proportion of each type of data.

(a) Single transitions

	1st	2nd	3rd	4th
Part1	AB	BA	AC	CD
Part2	AB	BA	AB	BC
Part3	AB	BC	CA	AD
Part4	AB	BA	AB	BC

(b) Double transitions

	1st	2nd	3rd	4th
Part1	-	ABA	BAC	ACD
Part2	-	ABA	BAB	ABC
Part3	-	ABC	BCA	CAD
Part4	-	ABA	BAB	ABC

Table 6.3 Types of time usage with measures, and whether it involves downsampling or a shift in time relative to the original sample time.

	Downsampling	Shift in time
Momentous/cumulative	*No*	*No*
Binned time	*Yes*	*No*
Left-, or rightward alignment	*No*	*Yes*
Event based time	*Yes*	*Yes*

proportion over time graph, although with discrete time. Since there are a total of 20 different two-AOI transitions with five AOIs, and 80 three-AOI transitions triples, the graph could be rather crowded, but in practice, only a few of the transitions and transition sequences actually appear in data.

In an explorative study, Goldberg, Stimson, Lewenstein, Scott, and Wichansky (2002) used a simpler version of the transitions over transition time diagram, which they call *navigation pattern*, on several tasks on a web portal page. Their study indicates that headers in a portlet were not usually visited before the body, by finding either the Header-Body-Customize or Header-Customize order of AOIs, in all cases but one.

6.4.6 Time and order

Time runs from the start of trial and onwards, at a sampling frequency decided by the specific eye-tracker. This basic, original time in the recordings is used for duration calculations in many eye-movement measures, including momentous and cumulative proportion over time graphs. In this section, we will look at three other organizations of time which are important for AOI events but also for the oculomotor event fixations, saccades, and smooth pursuit: binned time, left- and rightward alignment, and event based time. Each of these either downsample or shift data in time, as summarized in Table 6.3.

Binned time and ordinal positions

When data consist of fixations, dwells, transitions, or other events with variable duration, rather than raw data samples, we no longer have one point per time sample and cannot plot measure values over true time. One solution is then to use binned time and ordinal positions. We will use the example in Figure 6.17(a), which shows numbered fixations of two participants P1 and P2 laid out in true recording time. This is how we usually think about time.

Measuring how fixation durations develop over time requires grouping the fixations in bins of for instance 500 ms. Then all fixations starting during 1–500 ms after onset are grouped into a first bin, while saccades launched during the next bin 501–1000 ms are a

second group etc. Figure 6.17(b) presents the same data from P1 and P2 as binned average fixation durations. Binning means downsampling time, cutting it up in larger chunks, so this calculation comes at a loss of information, namely about precise start and end of fixations. Binning does not introduce a shift in time, however. Each bin in (b) represents a period of time from the original data in (a), which is the same for each participant.

The precise bin size is in effect an arbitrary setting that should be chosen as part of the experimental design or in relation to the data analysis, for example to ensure statistical independence between samples in successive bins. In literature, 500–1000 ms are commonly seen bin sizes. Occasionally, time bins may be much larger. For instance, Knoblich, Ohlsson, and Raney (2001) divided their trials into thirds: first, middle, and last bin.

In Figure 6.17(c), fixation durations are ordered by the fixation number; we compare the fixation durations between P1 and P2 for fixation number 6, for instance, regardless of the fact that the P2 fixation number 6 occurs much earlier than the P1 number 6 fixation.

Using ordinal positions to define the flow of time in measures has two consequences: first, time is *downsampled* into larger and variable chunks, in effect using fixation number as chronometric units, by which we lose information. Second, data are *displaced* between the two participants, so that although fixation 6 of one participant coincides with fixation 4 in true time, the comparison always uses fixation 6 of both participants, which is also a reduction in information. Whether this is desirable or not should be decided by the rationale of the experiment at hand, and in particular the expectations on the temporal flow of perceptual and cognitive processes, that is how much time they take, and in what order they occur, compared to fixations made.

The most common events that use chronometric units in measures are fixations, saccades, dwells, and transitions. Examples of such measures are the unique AOIs (p. 343), entry time (p. 437), and the ambient versus focal distinction (p. 266).

Leftward versus rightward alignment

When trials have different lengths, examining what happens just after onset of a stimulus event requires a *leftward alignment of trials*. Examining what happens just before the end of trials requires a *rightward alignment*. Figure 6.18 illustrates left versus right alignment for two participants with trials of different lengths. Aligning data does not downsample it any more than is already the case, but it shifts data in time between participants/conditions. In the example in Figure 6.19, Glaholt and Reingold (2009) calculate proportion over time curves with rightward alignment in a decision task between two images with variable trial duration. The two curves are calculated for a long time period of backward time, towards the decision.

6.5   Types of AOIs

The most basic type of AOI outlines a region in the stimulus that encapsulates what is 'interesting' from the perspective of your experimental design, and is used to quantify whether and how much someone looked at this particular region. Most of the time, the AOI is a single, static, and content-less region with well-defined borders. However, as we will see in this section, there are many varieties of the basic AOI type, and some of them violate one or many of these basic properties.

6.5.1   Whitespace

The areas of the stimulus *not* covered by AOIs can be referred to as *whitespace*. Whitespace has usually been regarded as wholly unproblematic: after all, if there is nothing there, why
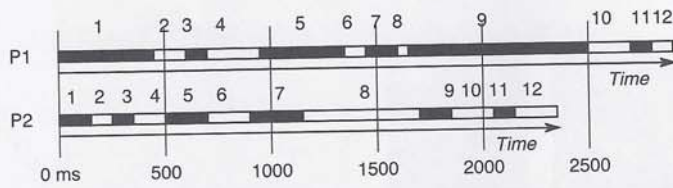
(a) Participants 1 and 2, with numbered fixations laid out in real time. For P1, executing 12 fixations takes 2.87 seconds, for P2 it takes 2.33 seconds.
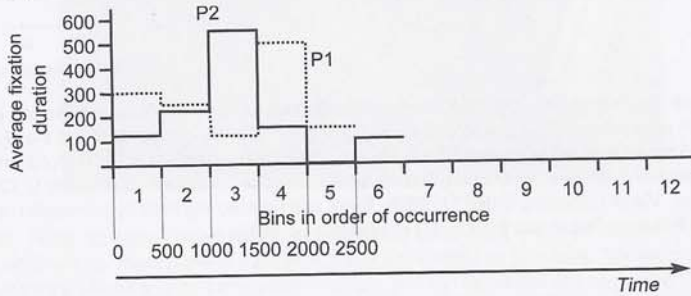
(b) Bins of size 500 ms have been applied to the two data in (a). For each bin, the graph shows the average fixation durations for fixations starting in the bin. Each bin represents a period of time in the original data in (a), which is the same for each participant.
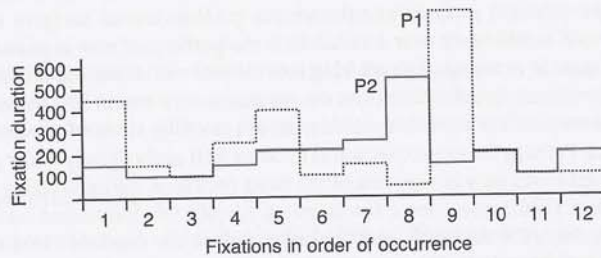
(c) Duration of fixations in the order of occurrence. Ordering data by events, such as fixations, efficiently shifts the original recording time. For instance, here we compare fixation number 6 of both participants, when in real time P1 produces his sixth fixation more than 400 ms later than P2.

Fig. 6.17 Real time data (a), binned time data (b), and data ordered by events (c). Three very different views of the same data.

Leftward	ACABCD AACB
Rightward	ACABCD AACB

Fig. 6.18 Leftward and rightward alignment of AOI strings over time. Leftward alignment is for studying behaviour just after onsets, and rightward alignment behaviour before offsets.
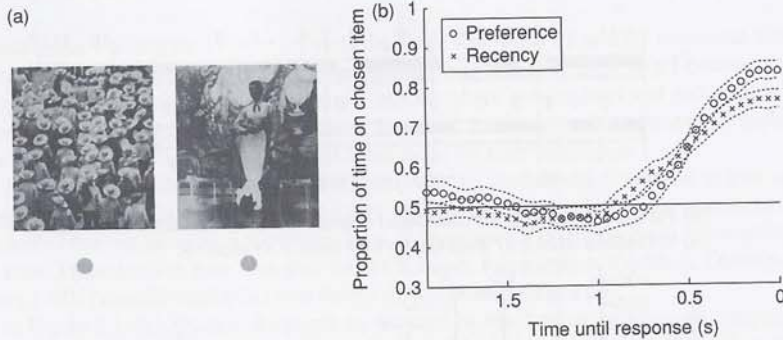
(a)

Fig. 6.19 (a) Two items in a decision task. When the decision is made, the user looks at a grey dot below the item he chose. (b) 'Gaze likelihood curves' plot the proportion of time spent on the chosen item, for each 50 ms time bin in the interval prior to the response. Dotted lines represent 95% confidence intervals about each time bin. The time course of gaze bias in visual decision tasks, Mackenzie G. Glaholt, Eyal M. Reingold, *Visual Cognition*, © Jan 11, 2009, Taylor and Francis, reprinted by permission of the publisher (Taylor & Francis Group, http://www.informaworld.com).

would anyone look at it? And if participants look there, why not just put in another AOI for the whitespace? Indeed, in most tasks participants will not look at whitespace, because the experimental task is set up to direct them to the objects selected as AOIs.

However, in longer experiments which require problem solving, looking at whitespace may indicate a mental process, afterthought, or perhaps mental imagery while solving the task, or simple indifference over a task which the participant was expected to do correctly. Moreover, as soon as we take eye tracking into the real world, outside the lab, the whitespace area becomes much larger, and affects the measures very much, in particular the order and transition measures. For instance, consider an eye-tracking study of product selection in the supermarket. Perhaps the researcher wants to do an AOI analysis only of the pasta shelves, but the participant looks only at very few of the pasta packages, before walking away over to the vegetable area (whitespace) for a few minutes. Should this movement count as a transition between the last AOI the participant looked at before the vegetable area and the first AOI looked at when returning? The percentage of whitespace is virtually never reported, but for many measures, in particular transition measures, it should be, so researchers can relate the reported results to it.

6.5.2 Planes

Planes are super-AOIs that exist in studies with multiple frames of reference, commonly due to combined eye and head tracking (supplied only by a few manufacturers). Technically, a plane is a two-dimensional surface in a three-dimensional space. A measurement procedure at the set-up ensures that the recording software knows exactly where each plane is located. The measurement system also knows the position and direction of the participant's head. This allows for online detection of AOI hits, in addition to saving data files with coordinates in each plane.

The keyboard and monitor depicted in Figure 6.20 are two planes, the picture behind is the third. Control rooms and aircraft cockpits are other examples of experimental settings that may require division of the stimulus space into a number of different planes. Some eye-tracker systems produce data where planes do not share coordinate systems, but each plane has its
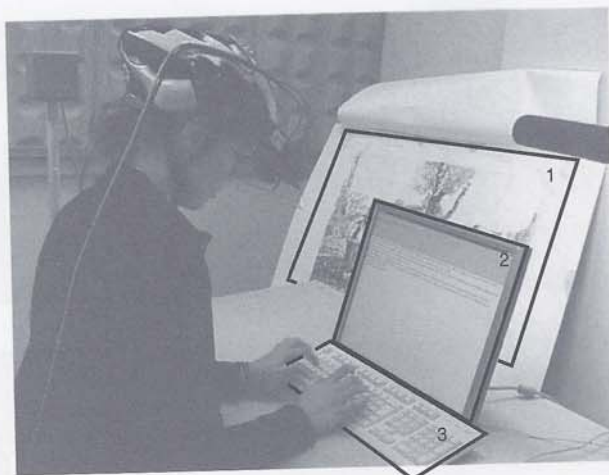
Fig. 6.20 Three 3D planes, numbered 1 for stimulus, 2 for monitor, and 3 for keyboard, in a study of reading while writing. Each plane is a super-AOI that can contain its own AOIs. The recording software knows where the planes and the participants head are, and can calculated AOI hits online.

own coordinate system (SMI head-mounted systems with Polhemus head-tracking, for instance), while others may have the same global coordinate system across planes (SmartEye). Also, planes need not be parallel, but can have any direction. A plane may contain many AOIs. Each AOI of course resides in one plane only, and all of the AOIs of that plane use the same coordinate system. Nevertheless, most of the AOI measures are applicable to planes as well. It makes perfect sense, for instance, to analyse dwell time on the monitor plane, or transition matrices between different planes in an air traffic controller's working space.

6.5.3 Dynamic AOIs

If your stimulus consists of animated stimuli or videos, the objects that you try to cover with normal, static AOIs will move away from under the AOI. You then need to use *dynamic AOIs*. These have recently been introduced in commercial software, and move in sync, following the underlying object. However, in current implementations, they require of the user not only specification of the AOI shape, size, and position, but also of how the AOI moves and changes form over time (Papenmeier & Huff, 2010). Figure 6.21 illustrates how a movie of a flying butterfly is tracked by a dynamic AOI throughout a short clip from a video sequence.

In the implementation of dynamic AOIs available in one of the commercial software packages, manual adjustments of AOI shape are required only for certain key frames, then the software automatically estimates the shape in the intermediate frames. This makes it easy to create AOIs following objects that move with constant direction and speed, whereas objects that move in a non-linear fashion require more manual work. Dynamic AOIs created this way are perhaps of most practical use when there are relatively few objects of interest in a stimulus, which is common to all participants.

Once the dynamic AOIs are in place it will be possible to use all AOI measures with the data. When dynamic AOIs overlap, however, a difficult prioritization is necessary (p. 221). Dynamic AOIs were first available in commercial analysis software in 2008, and have thus not yet been much used in research.
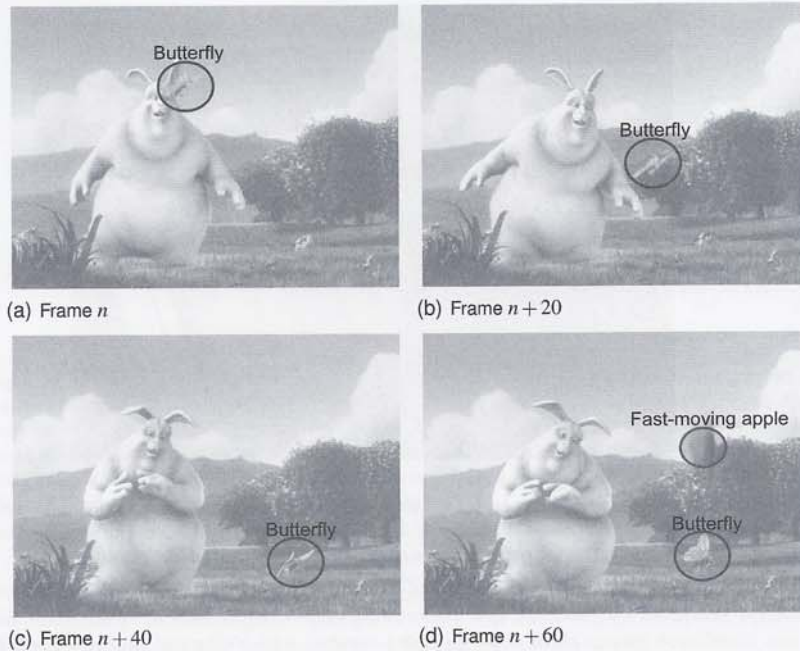
(a) Frame n

(b) Frame $n+20$

(c) Frame $n+40$

(d) Frame $n+60$

Fig. 6.21 Four frames of a film with 20 frames (833 ms) between them. A circular dynamic AOI tracks the flying butterfly, and the last frame also introduces another dynamic AOI to track the falling apple. Reproduced here with permission from the Blender foundation www.bigbuckbunny.org.

6.5.4 Distributed AOIs

The four instances of the man in Figure 6.22 are really only one single individual in four different locations according to narrative time when the picture is described. If we want to treat them as one in the statistical analysis, it can be seen as appropriate to use one single *distributed* AOI to cover all four instances of the man.

Holsanova (2008) notes that when participants describe Figure 6.22 and use phrases such as "it looks like early spring", there is no single, well-delimited item in the picture that corresponds to the concept "early spring". Spatially, the spring is spread out in all sorts of objects that provide evidence for the season (the dandelions, the birds in the tree, and the garden work). Distributed, non-connected AOIs are needed whenever the stimulus has semantics that are not spatially precise. "Early spring" is one such example, but also, "crucial information areas for solving mathematical tasks", as in Figure 6.23. Similarly, Morrison *et al.* (1997) collected several different and distinct areas into three categories: "situation awareness regions", "explanation-based reasoning regions", and "recognition-primed decision regions". Their analysis is based on this categorization rather than on the included physical regions. In fact, in experiments that have non-manipulated, natural stimuli, it is very common to have *a number of different AOI divisions* each pertaining to its own specific semantic level and subsequent analysis.

When using distributed AOIs, it is important to consider the interpretation of the measures used. The distributed AOI in Figure 6.23 seems semantically inconsistent with respect to first fixation durations, for instance, which can be expected to be lower for the ρ and θ parts of the AOI than for the $(\rho cos\theta, \rho sin\theta)$ part, simply because the latter is more complex to
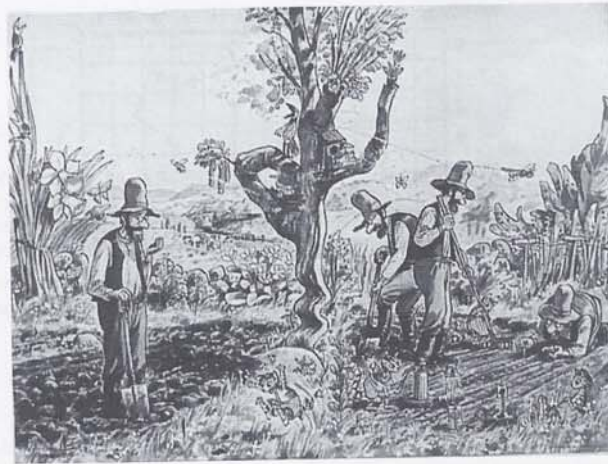
Fig. 6.22 Complex stimulus picture used by Holsanova (2008), Johansson, Holsanova, and Holmqvist (2005) and others. Reprinted from Nordqvist (1990), with kind permission from the author.
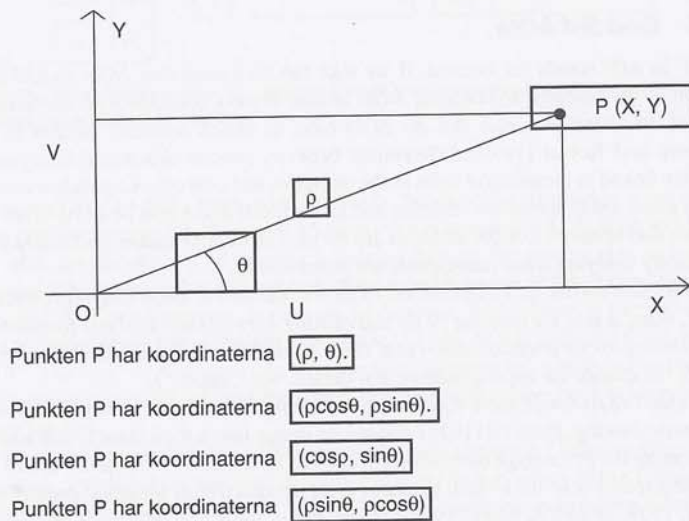

Punkten P har koordinaterna $(\rho, \theta).$

Punkten P har koordinaterna $(\rho\cos\theta, \rho\sin\theta).$

Punkten P har koordinaterna $(\cos\rho, \sin\theta)$

Punkten P har koordinaterna $(\rho\sin\theta, \rho\cos\theta)$

Fig. 6.23 A distributed AOI indicating crucial mathematical information as selected by experts. All seven AOIs are treated as one single AOI, only distributed in space.

understand. With a distributed AOI, we treat the first fixation in any of the seven part AOIs as the first fixation in the distributed AOI. If we compare first fixation durations between two groups, and the participants in one group look at the ρ first, while the participants in the other group look at $(\rho cos\theta, \rho sin\theta)$ first, then these durations will be different just because they land on areas where one requires deeper processing than the other. Moreover, ρ is closer to the centre of the display, where fixation is normally directed at the start of a trial, therefore any measure of latency to reach the AOI will obviously be contaminated because it is distributed
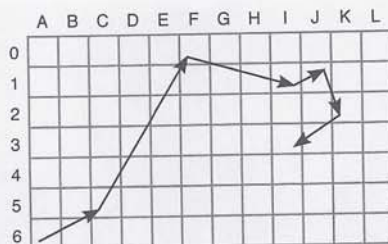
Fig. 6.24 Scanpath over gridded AOIs. A string representing this scanpath would be A6 C5 F0 I1 J1 K2 I3. On pages 273–278 and 348–353, such string representations will be used to quantify the similarity between scanpaths.

in space.

A distributed AOI should therefore have consistent semantics across all part AOIs. For instance, your research hypothesis should have a clear concept that generalizes over the different part AOIs, such as "crucial information", that can be motivated in relation to the dependent measures that you employ.

6.5.5 Gridded AOIs

The 'I' in AOI stands for interest. If we take the viewpoint that AOIs should be defined in relation to the research hypothesis, AOIs almost always coincide with the natural semantic units of the stimulus scene. But do AOIs have to match semantic entities in the picture? Goldberg and Kotval (1999) differentiate between *content-dependent analyses*, where the AOIs are linked to meaningful units in the stimulus, and *content-independent analyses*, which simply place a grid across the stimulus and let each cell in the grid be an AOI, see Figure 6.24. The fact that semantics of the stimulus are divided arbitrarily makes gridded AOIs unsuitable for directly studying *what* participants are interested in.

As we saw earlier (p. 192), gridded AOIs also define the dwell map representation of data. In fact, using a grid for creating AOIs and making a dwell time analysis for them, results in a crude version of an attention map (heat map) made from the same data. The larger the cells (AOIs), the cruder the approximation. For details, see Chapter 7.

Gridded AOIs found their way into eye-tracking research very early. In his seminal work on picture viewing, Buswell (1935) divided an image into a 4×4 matrix, and added a number representing the percentage of fixations in each AOI. Gridded AOIs can be useful for studying *how* participants scan the overall stimulus area, irrespective of semantic content (Goldberg & Kotval, 1999; Brandt & Stark, 1997). In particular, the string edit measures use the gridded AOIs representations, of the kind exemplified in Figure 6.24.

However, an inherent problem of gridded AOIs is how to choose the number of cells. Different cell sizes could yield very different results (Foulsham, 2008, p. 72). To avoid arbitrariness, studies using gridded AOIs in the analysis should employ several different cell sizes for the AOIs and show the same effect for each, which Pomplun, Ritter, and Velichkovsky (1996) do.

6.5.6 Fuzzy AOIs

All AOIs in use today have sharp borders. This means that a data sample or fixation is located either inside ('hit') or outside ('miss') the AOI. A fixation located directly outside the AOI border is therefore considered a 'miss' just as much as another fixation much further away.
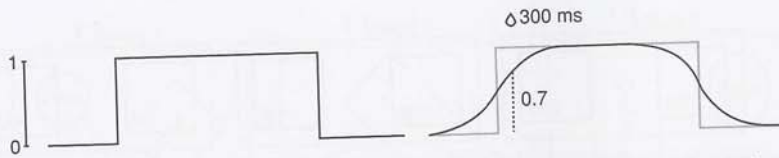
Fig. 6.25 AOIs can be seen as a function with infinitely sharp edges that defines the probability of 'hit' (left). If the edges are softened, we would have fuzzy AOIs (right), for which fixations share their duration with two bordering AOIs in proportion to the probability of being a hit, so that for instance a 300 ms fixation at height 0.7 adds 210 ms to the AOI and 90 ms to its neighbour.
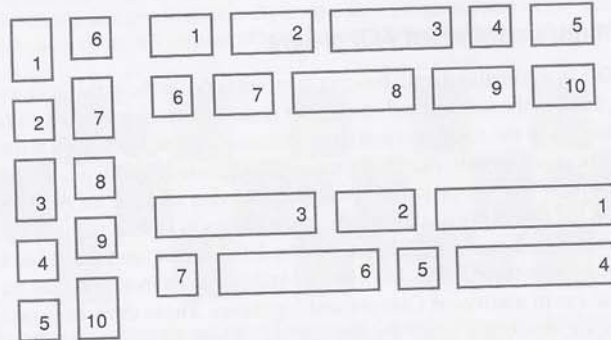

Fig. 6.26 The three most common conventionalized AOI orders, without which concepts like 'first skip' and 'regressions' would not be possible.

This motivates having AOIs with 'soft' or 'fuzzy' borders that, instead of a binary hit/miss decision, register partials hits. Figure 6.25 shows the principle for AOIs with soft borders. In a fuzzy AOI, the uncertainty of whether a fixation falls on the correct side or not is turned into a probability measure that assigns part of the fixation to one AOI and the other part to the neighbouring AOI or whitespace. In the case of duration, for instance, if the 300 ms fixation falls at a location where the probability of a hit is 0.7, the AOI is assigned a value of 210 ms $(300 \cdot 0.7)$. The level of fuzziness can be varied, for instance to correspond to the degree of imprecision and inaccuracy in your data, such that data with poorer quality have increasingly less sharp borders.

An alternative implementation of fuzzy AOIs is to append a Gaussian function around each fixation, and distribute the fixation duration to AOIs according to the volume covering each AOI (Buscher, Cutrell, & Morris, 2009). This approach emphasizes the other reason for smudging AOI borders; visual uptake may be distributed over a wide area, and this should be reflected in the measure values.

Although fuzzy AOIs are applicable to both dwell time and transition measures, the concept of fuzzy AOIs or the alternative of using a Gaussian attention deployment has not been thoroughly investigated in the literature, and validity experiments are needed to be able to confidently use fuzzy AOIs. For example, does attention constituted by a longer peripheral fixation equal a spot-on fixation, if they both result in the same weighted (as per above) dwell time?
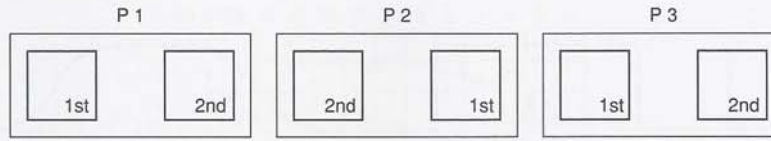
P 1   P 2   P 3

1st   2nd   2nd   1st   1st   2nd

Fig. 6.27 Three participants and the same stimulus, with two AOIs. Participants have later answered questions about what is their first and second choice of the two items shown in the AOIs. The hypothesis wants to compare entry and dwell times (AOI measures) for first versus second choice AOIs. This require same-position AOI to have different identities for different participants.

6.5.7 Stimulus-inherent AOI orders

Mostly, AOIs in a stimulus do not have an inherent order to them that tells us that one of them is the first, another the second, and so on. The great exception is the conventionalized reading order that has made the reading researchers define a large number of AOI measures that make little sense for other stimuli, like "between-word regression", which is a backward movement to another word. The regression in itself assumes that there is an AOI with a lower order number than the current one, and that the gaze moves to it. The numbering of AOIs follows the conventionalized reading order, be it from-left-to-right, and down (as in European languages), from-right-to-left, and down (as in Arabic), or from-top-to-bottom, and to the next right column (as in traditional Chinese and Japanese). These three are examplified in Figure 6.26. There are also other lesser known reading orders. In traditional Mongolian, words go from-top-to-bottom, and to the next left column, whilst in the ancient Boustrophedon system words follow a zig-zag-pattern, with alternating reading directions for every other line.

Few other stimuli than text, if any, have a conventionalized order of AOIs. Rather, a common research question is to investigate whether a particular type of stimulus has a conventionalized scanning pattern or not. Of course, a participant's reading order may be reflected in the order he looks at other objects other than text. For instance, Lam, Chau, and Wong (2007) show that participants scanning thumbnails on commercial web pages do so in an order reflecting their dominant reading direction.

6.5.8 Participant-specific AOI identities

AOI identities are not always independent of the participants looking at them. For instance, we may have a stimulus picture with two toys that the child participants are also asked to rate after the recording was made. Now we want to have an AOI that is "the best-liked toy", i.e. for each child that toy which was rated the highest. This AOI will cover different toys for different children, as in Figure 6.27. A similar situation appears if you have an internet study with different articles and ads, and you want to let the interests of the participants decide the identity of the AOI. To the extent that they handle *individual AOI identities*, modern analysis software solves this by letting the user name AOIs differently for different participants; the name then decides the AOI's identity during analysis. The analysis in Glaholt and Reingold (2009), exemplified on page 208 assigns AOI identity based on participants' choice between two AOIs; i.e. 'chosen' versus 'non-chosen'.

6.5.9 AOI identities across stimuli

The identity of an AOI can be just as much decided from the experimental design as it can from participant actions. In both cases, this overrides the basic definition of AOI by it spatial extension. For instance, in a study where stimuli are always a face and a cellular phone (we
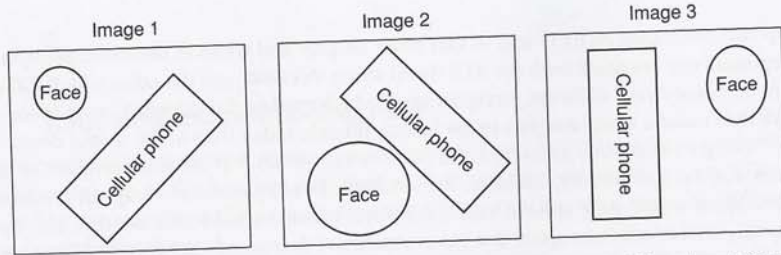
Image 1   Image 2   Image 3

Fig. 6.28 Three stimulus images with a face and a cellular phone AOI in each. Although each face looks different, has a different size and position in the image, from the perspective of our experimental design we can view them as a single AOI concept of faces instantiated in three trials.
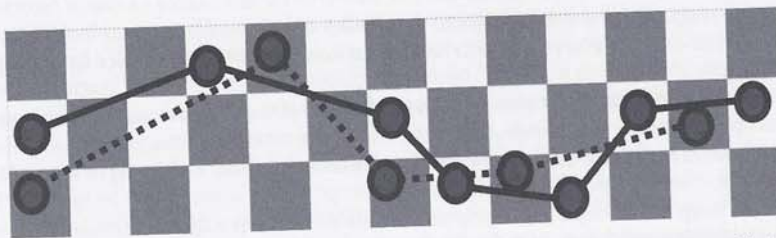

Fig. 6.29 These two scanpaths appear very similar when visualized in space, and many spatial similarity measures would give a high similarity score, but note that one participant scans only white areas, and the other only grey areas.

may be in the advertisement field now), but always different faces of different sizes at different positions, as in Figure 6.28. In her experimental design and result summary, our hypothetical researcher nevertheless considers all these different AOI as identical, and presents a single average entry time and dwell time value for the face AOI.

Letting a single AOI concept from the experimental design cover a number of AOIs from different stimulus images is then a way to increase generalizability and overall validity of the study, because if our researcher finds that entry times are significantly lower and dwell times higher for the cellular phone than for the face, then she can support that from a wide variety of combinations of sizes and positions. Again, software handle AOI identities across stimulus images by letting the name of an AOI decide its identity.

6.5.10 AOIs in the feature domain

AOIs can be defined in space and in terms of *features*, which take into account selected aspects of the content in the AOI. Feature analysis provides an important addition to all measures, increasing their usefulness manyfold by replacing space with a quantification of the semantics in the stimulus. As a simple example of how a feature space analysis works, Figure 6.29 shows two scanpaths over the same checkered stimulus. At first glance, the scanpaths are very similar: their spatial extension and form coincide very well, and would score high on many similarity measures. Looking more closely, however, you will find that one of the scanpaths hits only white and the other one only grey areas. If the grey–white difference is important in our study, the two scanpaths are not at all similar. In order to have measures that capture feature similarity, rather than spatial similarity, we should analyse data in a space spanned by the grey and white and not by the spatial x and y dimensions.

In Figure 6.29, the features are very simple, clearcut, and regular. It is not difficult to set

up two distributed AOIs G and W that cover all grey and white in the stimulus. We can then represent one scanpath with the AOI dwell string WWWWWWW and the other with GGGGG, which are obviously very different, irrespective of which method for scanpath comparison we use. We thus make a string analysis in the feature domain rather than in the spatial domain.

The grey and white areas have only two colour values, but the feature values of the AOIs may also be continuously variable. For instance, in a supermarket study, all products that a participant would look at have values for price, brand, carbohydrate content, etc, that serve equally well as a feature space that can complement the natural one. Instead of measuring, for instance, the saccadic amplitude in pixels or degrees of visual angle between successive fixations, we can now measure amplitudes along the scanpath in price or carbohydrate content: a saccade from pasta AOI number 1 to pasta AOI number 7 can have an amplitude of €−3.6 , calculated as price of AOI 7 minus price of AOI 1, or for that matter +1.6% of carbohydrates. The average of the absolute saccadic amplitudes of one participant may be €0.18 , which would indicate that every new pasta he looks at deviates very little in price from the previous pasta.

Feature-space analysis quantifies the semantics of the stimulus and replaces space with that quantification. As a consequence, the unit of the measure values are changed: instead of centimetres and visual degrees, we measure distances in colour or in price or in carbohydratic content.

As long as the objects carrying the feature have sharp edges, a feature-based analysis can be implemented with AOIs, and comparisons made using for instance semantic distances within a substitution matrix (p. 276). In real-life mammography images, however, there are no clear-cut borders where AOIs can be laid. Dempere-Marco et al. (2006) therefore used five previously developed *visual feature detectors* for mosaic attenuation, nodules, reticulation, ground glass, and bronchiectatis, which are theoretically important for understanding the search behaviour of radiologists. A visual feature detector is a small algorithm that takes a piece of the stimulus image and returns a value; as though saying for instance "this patch has 0.7 in mosaic attenuation, so the fixation should be given that value." Using five feature detectors, a fixation f_i that lands in a mammography image can thus be attributed a five-dimensional $(ma_i, n_i, r_i, gg_i, b_i)$ vector, where each dimension gets its value from a feature detector (ma for 'mosaic attenuation' etc.). Dempere-Marco et al. made all analyses in that five-dimensional feature space rather than in the two-dimensional spatial space.

More generally, Dempere-Marco, Hu, and Yang (2011) propose that a feature-based data analysis can be made in two steps:

1. Select a feature domain with dimensions such as price and carbohydratic content or the five-dimenional features of X-ray images. Feature domain selection crucially includes a mapping function from spatial positions in the stimulus images to feature values, which is trivial for the pasta case (feature value for price is read from the price tag), but less obvious for radiology studies.

2. Impose a measure on the feature domain. Dempere-Marco et al. focus on pairwise position similarity, but nothing prevents us from using any other of the many measures where space can be substituted for a feature domain, including the vast majority of measures in Chapters 10 and 11.

6.6 Challenging issues with AOIs

The precise location and shape of an AOI needs to be decided in close relation with the hypothesis, the composition of the stimulus, the quality of the recorded data, and the method

of analysis.

6.6.1   Choosing and positioning AOIs

How do you decide how your stimulus should be divided into AOIs? It is your hypothesis that decides what your AOIs should be, and there is therefore no point in using more AOIs than required by the hypothesis. This is easy for many simple constructed stimuli, but difficult for natural and cluttered scenes.

Semantic composition of stimuli

In many experimental settings, the stimuli are so constructed that the assignment of AOIs to parts of the stimulus is very straightforward; often there is a set number of objects and nothing else, except whitespace. The only issue then is how large to make the margins. With pure reading stimuli, individual morphemes, words, or sentences can each be given an AOI. Also, for many stimuli which were not originally conceived for use in eye-tracking experiments, but are nevertheless man made, AOIs can often easily be assigned to segments. For instance, an ad on a web page makes an obvious AOI, as well as shelf space for particular product brands in the supermarket. Such man-made pre-divided stimuli are *largely* unproblematic.

Properties of visual intake and recognition may complicate the AOI division however. If your stimulus is so simple, and your AOIs so close that your participants are able to take in one AOI in peripheral vision while looking at the other, it is dubious to contrast dwell times from the two areas and argue that visual intake is larger from one AOI than the other. Fortunately, the visual phenomenon known as 'crowding' (which originates from work by Loomis, 1978) tells us that, as peripheral information becomes more cluttered, it is very difficult to distinguish between different elements away from the current point of fixation. Therefore, for complex displays, AOIs which are close to each other may not cause a problem because crowding restricts focus to the fovea. In fact, in many studies, peripheral detection is manipulated by crowding the target with additional letters and characters.

If instead you use a naturalistic complex picture, like the stimulus image in Figure 6.22, where there is very much natural crowding, what is the proper division? Should the man to the left be represented by one AOI only, or should we make one AOI for his face, one for the soil he holds in his left hand, and one for the shovel? We could continue, and enclose each minimal semantic element within an AOI; one for each spider, butterfly, and the bird in the tree. However, we have to draw the line somewhere (literally!); every blade of grass and leaf on the tree should not be given AOI status. As pointed out earlier in this chapter, your hypothesis should guide your AOI divisions so that they are sensible and can provide answers to the empirical questions you are asking.

The stimulus can be divided into AOIs on several levels, according to your hypothesis. In Holsanova (2008, 2001), a crucial question was how speakers coordinated the spoken name "Pettson" with looking at the man, during free spoken descriptions of the picture. Then, each portion of the image in which the man is present should be given its own AOI, but with the same reference label. In another type of analysis, the transitions between the face, the hand with the soil, and the shovel were of main interest (as they may signify a sudden deeper understanding of the picture thematics). This requires an AOI analysis at a finer level of semantic composition.

Who should decide AOI positioning?

The exact positioning of AOIs is crucial, because it can determine whether you reveal a significant effect or not. However, who should make this important decision, and on what grounds? Usually, the researcher herself decides where to position the AOIs. This is an option

when interesting regions in the stimulus are easy to separate from each other, and there is no uncertainty of whether they qualify as AOIs, preferably using an unambiguous and exhaustive list of criteria, but otherwise it may open up for various degrees of subjectivity. Therefore we will discuss the following alternatives:

- Using *experts* to define the AOIs. This can be done before recording, but it is manual with the risk of being subjective. Nevertheless, using human experts improves the semantic link between your stimuli and their respective AOIs. To this extent the degree of objectivity may be greater than if you position AOIs yourself.
- Using *scene stimulus*-generated AOIs. Here, the positioning is algorithmic, and therefore more objective. The scene properties themselves plus algorithms define the areas.
- Using *attention maps* to define the AOIs. This is post hoc, but it is at least done by algorithms, although threshold settings are arbitrary. No semantics involved.
- Using *clustering algorithms*. Again post hoc, but also again algorithmic with arbitrary threshold settings. No semantics involved.

Expert-defined AOIs The manually defined AOI can be made somewhat more objective by having experts define it. Experts can be expected to have a very detailed knowledge of the semantics in the stimulus. When given the task to decide which are the most important areas in the stimulus for solving the task, mathematics professors selected AOIs as shown in Figure 6.23 above. Expert judgements of AOIs can be used in a variety of stimuli: air traffic control interfaces, nuclear plant control rooms, art, architectural facades, and medical education videos are just some examples. For mathematical stimuli, it is relatively difficult to point out which are the most important semantically coherent AOIs even for an expert, since mathematical problem solving is a process that uses many if not all parts in combination. It is easier for air traffic controller environments where the scene has been designed with clear functional distinctions between areas, and probably even in medical education videos, where one or two small areas on a patient's body can be decisive for a diagnosis.

Stimulus-generated AOIs The scene stimulus itself can be used to create AOIs. Mossfeldt and Tillander (2005) present several attempts to automatically identify AOIs using edge detection and colour segmentation. They conclude that using image processing to automatically find AOIs is very dependent on the specific stimulus. Edge detection may not work adequately for natural images (photographs), and colour will not always be an effective method of segregating the stimulus. Constructed images such as text, logos, and illustrations have clearer edges however, and also more uniform colour. Therefore these classes of stimuli are better suited for image processing segmentation of AOIs. When the stimulus is not an image per se, but a computer generated display perhaps, it is sometimes much easier to have the computer render the AOIs directly from the stimulus. For instance, when studying reading of long texts (hundreds or thousands of words), a lot of time can be saved if the stimulus software can be made to produce the many AOIs directly from the stimulus. You just take your formatted text, paste it into a window, and it can be processed to a stimulus image with AOIs for each item in the text, such as words, inter-punctuation, sentences and in some cases even graphics. Not only does it save time, but it also adds precision, since each AOI is positioned at the same height as the others, with the same margins. Figure 6.30 shows AOIs generated by such a system. The same approach is possible for some non-reading scene stimuli also, namely those where the stimulus software handles the objects of the scene and can automatically attribute AOIs to them. In an animated game, for instance, you may make a log of all the sizes and positions of the automated AOIs, and then use that log

Dunkin'	Donuts	Dunkin'	,	a	largely	Eastern	U.S.	coffee	chain	, has	ambitious	plans	to
expand	into	a	national	powerhouse	,	on	a	par	with	Starbucks	. But	Dunkin'	is	no	Starbucks
In	fact	, it	does	n't	want	to	be.

Fig. 6.30 Stimulus-generated AOIs for a short text.

to calculate AOI hits after recording or even online (Papenmeier & Huff, 2010; Holmberg, 2007). This could be the case for animations of various kinds and possibly for internet pages.

Attention maps for AOIs The attention maps described in Chapter 7 offer an alternative way to define AOIs by cutting off the top of the attention map, and letting the flat region(s) generated by the cut define the AOIs. As an alternative to cutting peaks in attention maps, Hooge and Camps (2009) hand-coded AOIs from heat map visualizations by manually drawing them at a specified colour (height in the attentional landscape, p. 233). This is an approximate, relatively objective, and not too time consuming method of obtaining AOIs that correspond to clusters. Page 248 provides more detail on the use of attention maps for AOI definitions.

Clustering for AOIs Clustering algorithms are used to divide an initial set of data samples or fixations into subsets that are similar in some sense, most often in term of spatial proximity. The strategies behind clustering differ somewhat between methods. A number of algorithms look only at the spatial proximity of the data points, such as Goldberg and Schryver (1995a), who present a clustering algorithm for finding fixation-like spatial clusters. Random samples are chosen for clusters, and iteratively the closest neighbour sample to each cluster is added to the cluster, resulting in what is known as a minimal spanning tree. Another related approach is the mean shift algorithm by Santella and DeCarlo (2004), which iteratively shifts points to higher-density areas in order to reach local maxima. However, the most common method by far to cluster data points is the k-means algorithm. It is related to Lloyds and the Linde-Buzo-Gray algorithms (Linde, Buzo, & Gray, 1980) used in vector quantization, and is straightforward to implement. In its basic form, it divides the image space into voronoi partitions based on the cluster centres. The user must define the number (k) of clusters to be found, and results may vary by each new run. Also, it does not provide actual AOIs, but only shows which points group well together; a convex hull could then be used to produce the AOIs from these points. A convex hull describes a selected set of raw samples or fixations by a minimal area that covers all points, as in Figure 6.31. This example illustrates that while it could perhaps be possible to use cluster-based AOI generation for image stimuli, it is considerably more difficult to find natural clusters when groups of raw samples are close to one another, as in the case of the text. In fact, these cluster-based AOIs have completely lost the connection to the natural semantics of the display, for both text and image. Using these AOIs for dwell time or transition analysis without substantive manual post-editing of AOIs may yield meaningless results, and manual editing would nullify the time saved by the algorithm. From a statistical point of view, using clustering techniques to create AOIs may violate assumptions of independence. The same points that are used to create an AOI are also used to calculate its contents. The result may be inflated values for AOIs that are large, positioned in relatively free areas and consequently capture more stray raw data samples. These values will be inflated compared to smaller AOIs located at areas with high competition from other AOIs, which will not capture stray raw data samples to the same degree.
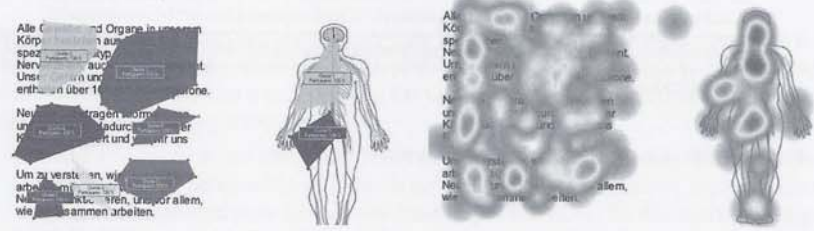
(a) Convex hulls around recorded fixations define AOIs as a result of clustering. Proportion of participants looking at each cluster are reported.

(b) Proportion-based heat map visualization on the same data.

Fig. 6.31 Clustered AOIs versus heat maps in the Tobii Studio analysis software. Hand-drawing AOIs around heat map centres may be less arbitrary and not take much longer.
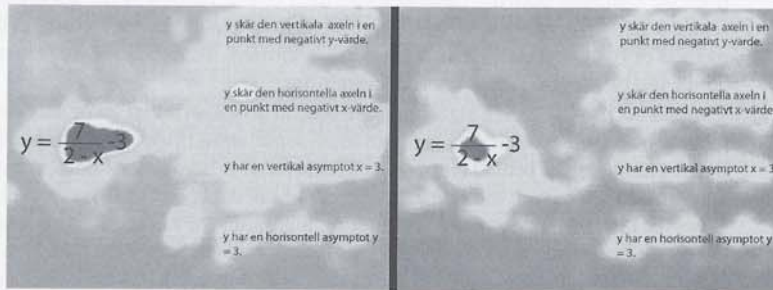

Fig. 6.32 Heat map visualizations of 21 students of mathematics (left) and 24 students of humanities (right) solving the same task. Same settings for kernel width and colour mapping in both heat maps.

Can we modify AOIs post hoc?

Ideally, your choices over precise levels of composition and divisions are matters which should be decided when developing your experimental design. But what if you prefer to record the data first, then look at the attention maps or scanpaths, before deciding where to put your AOIs?

For instance, Figure 6.32 shows the heat map visualizations of mathematics and humanities students when solving the same mathematical problem. The researchers are interested in finding which information the mathematics students use that the humanities students do not. Suppose that at the onset of the project, the researchers did not have a clear idea which parts they should choose for candidate AOIs. The heat maps show that the semantic item "-3" could provide the crucial difference, so it would be tempting to choose a fine-grained AOI analysis, and in particular put an AOI over "-3" and test the dwell time difference between mathematics and humanities students for that AOI. In doing this, we abandon the ideal of constructing the hypothesis before the data recording, and in effect enter into a semi-explorative research mode. This is not necessarily a bad thing: experiments often gain from being re-conceptualized on the path from original design to presentation of results. We must remember, though, that whenever a researcher alters AOIs to adapt better to her data, she is also altering her hypothesis and the whole story behind her study. This does not conform to

proper empirical practice.[21]

In reality, however, it is not uncommon that researchers do not position AOIs until early in the data analysis stage, long after they designed the hypothesis. Even if the hypothesis is very clear about the general position of AOIs, it is often inexact, if the stimulus is particularly complex for instance. For stimuli with small spatial distances between semantic units, it is advisable to get into the habit of drawing the AOI positions in the analysis software *before* data recording. Some post-hoc refinement of AOI positioning may be necessary afterwards, but at least you do not run the experiment blind to its analysis, and the important AOI-related data will be captured *with respect to your hypothesis*. This is particularly important if there is a substantial gap between the initial design stage and the analysis stage, during which many participants are recorded and important details about the study may be forgotten.

6.6.2 Overlapping AOIs

In general, AOIs should not overlap at any single level of composition, because of the danger that single AOI hits and transitions will be counted twice, rendering your statistics difficult if not impossible to interpret. Most statistical tests assume that the data are independent—such double occurrences invalidate this assumption. Counting twice also inflates the data from overlapping AOIs compared to non-overlapping ones.

Nevertheless, there are cases when AOIs indeed overlap. Figure 6.33 shows four such situations. In Figure 6.33(a), there are two distinct levels of composition. The smaller sub-AOIs are completely engulfed by the larger AOIs within which they reside. This would be the case if the stimulus were two documents with sections of text and images inside them. Here, counting dwell time on both levels is appropriate, because the larger documents are semantic owners of the smaller ones. In these cases, we can then just subtract the AOI dwell time of the 'important information' AOI from the 'other information' AOI to get a corrected dwell time for the latter. Transition counting is not as obvious. For instance, does a saccade from the small area into its larger owner area count as a transition from one AOI to the other, or only as a movement inside an AOI? It makes most sense if transitions are counted only at each level of composition separately.

Figure 6.33(b) shows the unfortunate case when two static AOIs overlap partially. Should a dwell in the overlap area be counted as belonging to neither, one, or both the AOIs? Deciding which saccades should count as transitions is even worse. If one AOI covers the other, being in front of it, then it would be clear that any data in the overlap area only belongs to the frontmost AOI. But what if the stimulus image consists of two semi-transparent and partially overlapping objects, as is often the case in advertisements and graphic design? How to quantify AOI measures for studies with such material must be decided depending on the particular experimental design.

Figure 6.33(c) represents an AOI from a drop-down program menu overlapping the underlying taskbar AOI. As it is not transparent, the menu dwell time is not shared with the taskbar AOI, and the saccades between them should be counted as actual transitions. However, the menu AOI has a limited duration which is decided by the clicks of the participant.

---

[21]The danger becomes apparent when we realize that the precise position of the AOI over the "-3" in Figure 6.32 could decide whether the researchers obtain a significant result. If the dwell time comparison between mathematics and humanities students yields a p-value of 0.074, it might be enough to move the border of this AOI just a little to be able to creep below the magic boundary of 0.05. Under pressure to produce results, a weak researcher might be tempted to argue to herself that there is really no objective and precise spatial border between the "-3" and the larger quotient that is more correct than any other spatial border next to it. If so, she may say, what harm is there then in moving the AOI border a pixel or two? That small distance is far below the precision and accuracy of the eye-tracker, anyway.
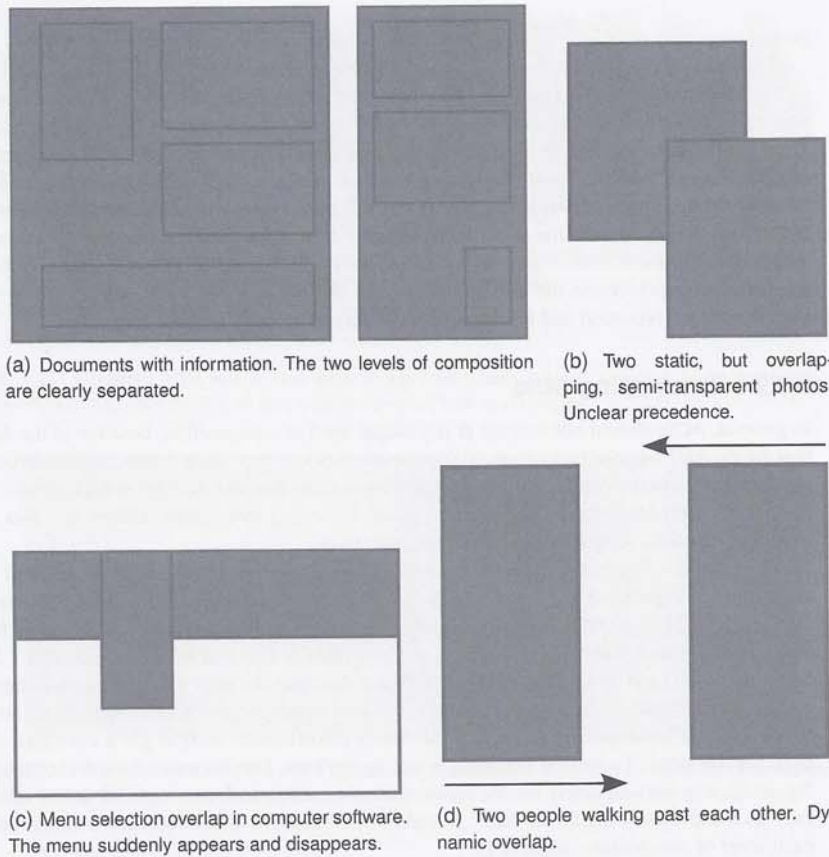
(a) Documents with information. The two levels of composition are clearly separated.

(b) Two static, but overlapping, semi-transparent photos. Unclear precedence.

(c) Menu selection overlap in computer software. The menu suddenly appears and disappears.

(d) Two people walking past each other. Dynamic overlap.

Fig. 6.33 Four different ways that AOIs can overlap.

The case in Figure 6.33(d) shows two dynamic AOIs moving towards an overlap situation. They could be two people walking towards each other, or, in a car-driving study, a pedestrian and a cyclist crossing the junction ahead. Such dynamic AOIs could occur at different depth planes according to the field of view of the observer. Is a hit on the pedestrian also a hit on the cyclist when they overlap? Should only the AOI closest to the driver be hit? This could easily be calculated if the scene is a 3D model, but it becomes more difficult if it is a real traffic scene, or just a video recording, and the AOIs are at different distances in depth. But even if we could calculate which the closest object is, seen from a visual intake perspective, how can we be certain that only one and not both objects are perceived and processed? The foremost object may not cover the more distant one completely, then, we have an overlap of AOIs that could allow both to be perceived. Even small children can easily recognize two objects in an overlap situation (Ghent, 1956), given time, but as Duncan (1984) shows, participants still tend to allocate attention only to *one* object at a time when looking at two objects which overlap.

There are at least five ways to deal with the potential problem of getting more than 100% of total dwell time in these dynamic overlaps. None of them is perfect.
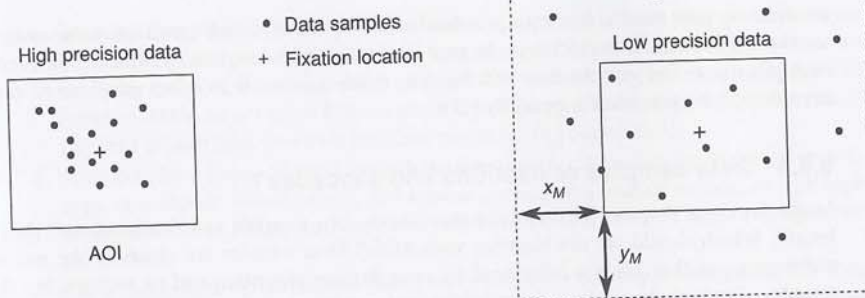
Fig. 6.34 Additional margins (x_M, y_M) have to be added to an AOI when the precision is low, otherwise some samples will miss the AOI resulting in, for example, shorter average and total dwell times. In this example, all data samples belong to the same fixation, which is located in the middle of the AOI.

1. Accept it, and modify the statistical tests. This alternative assumes that participants fully perceive both AOIs.
2. Simply divide the dwell time of each AOI by the total sum in all AOIs. This forces the excess to be spread equally onto all AOIs, whether they overlap or not.
3. While recording data, create an overlap duration matrix, which for each pair of dynamic AOIs tells us for how long that pair overlapped. Then take the excess dwell time (the part above 100%), and let each pair of AOIs pay for that excess in proportion to their share of the total overlap duration in the matrix. This is more fair to overlap per se, but ignores whether the AOIs have actually been looked at.
4. Calculate the distances between a data sample landing in the overlapping area and the AOI centres. Then assign the sample to the nearest AOI.
5. In an overlap situation, if we have a dwell on two transparent AOIs, we give each of them one half of the dwell time.

Alternative 5 has the advantage that none of the AOIs that were neither overlapping nor looked at need to contribute to the reduction in dwell time caused by normalization. From a visual intake perspective, we could argue that from eye tracking alone, we do not know which of the two transparent objects has been attended, nor how much attention has been allocated, so the equal distribution of dwell time is a fair probabilistic estimation. For transitions, we should only count saccades that move between the two AOIs at the time when the AOIs do not overlap.

6.6.3 Deciding the size of an AOI

The accuracy achieved in your measurements (pp. 41–43) is the major factor that decides the smallest element that can be given an AOI. In theory, it would be possible to use AOIs as small as 0.5° for participants and systems that give a high accuracy after calibration. In practice, however, this is rarely applicable since imprecision of the eye-tracker (pp. 33–41) requires additional margins to enclose all data samples, as illustrated in Figure 6.34. Taking both accuracy and precision into account, the practical minimal size of an AOI can be expected to be around 1–1.5° for high-end eye-trackers, because this is the size of the fovea and the best eye-trackers have the precision to accommodate such as size. Consequently, this is also a minimum margin to be added around objects of interest in your stimuli.

If your AOIs are smaller than the precision in your data, the results that you get from the AOI analysis will have shorter dwell times and a massive amounts of entries and transitions,

invalidating your results. Inaccuracy instead causes dwell time and transitions to be assigned to other AOIs than the correct one. In gaze interaction with low-cost eye-trackers, it is common practice to lowpass the data and this way make is possible to select small menu items, even though the precision is typically poor.

6.6.4 Data samples or fixations and saccades?

In the previous chapter, we have discussed both data samples and fixation-based data at length. Which should we use together with AOIs? Data samples are closer to the real eye movements, as they are not influenced by your fixation algorithm and its settings, but data samples also include artefacts from blinks and varying optic conditions during recording. Moreover, dwells could be dispersed throughout an AOI whereas fixations are typically evaluated solely based on their centre locations and durations. Taken together, this makes dwells and fixations different from each other in a number of aspects.

The dwell is comprised of all samples from entry to exit, regardless of whether they originate from a fixation or not. All samples belonging to a fixation, however, do not necessarily have to reside in the AOI, as long as the central location of the fixation does. In Figure 6.34, for instance, the fixation duration assigned to the AOI is the same regardless of whether the data have high or low precision (15 samples). However, both average and total dwell time calculated from the data samples differ significantly between the two cases.

When several consecutive fixations and data samples reside in the same AOI, total dwell time is per definition longer than the sum of fixation durations, and you can expect them to differ by about 20%. When investigating general viewing behaviour, it does not really matter which you use as long as you are systematic. However, if you are investigating a specific claim in the literature you should be aware of how duration has been previously calculated.

Some AOI hit and dwell-based measures require the use of data samples, like the proportion over time graphs. Running them through an event detection algorithm, as noted in one of the hands-on points on page 153, is then one possibility to exclude everything but, e.g. fixations, and then use data samples only from within the detected fixations.

For the *transition* event, artefacts from using data samples may be much larger, since a saccade may cross an AOI or be split by an intervening fixation, depending on the algorithmic definitions you choose.

6.6.5 Dealing with inaccurate data

Offsets in the data are major concerns for any AOI analysis, whether caused by *drift*[22] in the eye-tracker, droopy eyelids, a miscalibration in one corner, or something else. Before running an AOI analysis, always check whether data from any of your participants exhibits systematic offsets in parts of the image. You may have recorded 80 participants, for instance, and for 18 of them, data coordinates may be slightly shifted up or down, so that for a specific AOI, a large portion of the samples that rightfully belong to that AOI are in fact allocated to a neighbouring AOI. In reading studies, for instance, dwell times for one line of text may have been shifted to the line above or below it. The effect in your data analysis may be large enough to render another statistical result than the true one, and profoundly undermine the validity of your study. There are four ways to deal with this problem:

1. Before recording, when you construct your stimulus and your AOIs, be sure to add *a margin* to your AOIs, so that small offsets can be captured in the margin. Select margin

---

[22]System-inherent drift in the eye-tracking equipment causes increased inaccuracies and is not the same as drift of the eye during prolonged fixations.

sizes based on the expected precision and accuracy levels with your particular eye-tracker and participants. In some cases, competing AOIs are so close that you have no space for margins, as in Figure 6.1 on page 188; or there may be just a little space in between AOIs, as in Figure 6.30 on page 219. You need to be aware of this because it can lead to both false positives and false negatives in your statistics.

2. After data have been collected, consult the data quality ratings you made during recordings, or scanpath visualizations, and *remove participants with offset data or with high imprecision* from further analysis. You may have to record further data to compensate for the loss.

3. On a participant-by-participant basis, *move the AOIs* so they cover the correct data rather than the correct portion of the stimulus image, as in Figure 6.35(a). This takes time, and is difficult to do correctly, unless your stimulus material is text.

4. On a participant-by-participant basis, *shift the data* back so the offset is neutralized and data again covers the correct part of the image, where the AOI is, as in 6.35(b). The option to manually move fixations is currently implemented only in EyeLink software,[23] but automated "drift correction" and "offset repair" algorithms have been developed for reading data, for instance the iDict software by Hyrskykari (2006).

Alternatives 1 and 2 are the only fully satisfactory solutions, although alternative 2 adds a somewhat larger level of insecurity about your results. The two data repair alternatives are equivalent, but it is important to remember that such shifts of data or AOIs should be made only if it is *obvious* from scanpath visualizations how the repair should be made. For some stimuli, like text or newspaper reading, the scanpaths are so systematic in their alignment to the stimulus that any offset is immediately visible, and its size and direction easily calculable. For general scene images with a varying content, it is often much more difficult to correctly estimate the needed offset repair.

The problem with correcting data is of course that while increasing your chances of getting the right result, it also undermines the credibility of your conclusions, whether it is called "offset compensation" or "post-recording drift correction". If the reader of your paper knows that your data were not of sufficient quality to yield a significant result without corrupting them, then he will be in doubt about whether to believe what you report. Sometimes data corrections may increase the validity of your results, but we have no guarantee (other than the integrity of the researcher) that data were not shifted to make a non-significant result significant. Also, what about other measures; is saccadic amplitude affected when fixations are moved, for instance? The only sure solution is to remove data so poor that the resulting values for measures are not reliable.

6.6.6 Normalizing AOI measures to size, position, and content

When comparing AOI measures such as number of fixations and dwell time between AOIs (rather than between participants), you may sometimes feel that it would be fairer to the data to scale—or normalize—the dwell time value to the area, position, or content of the AOI.

---

[23] A pragmatic aspect of the EyeLink software/algorithms is the possibility of "performing drift correction on fixations" (SR Research, 2007, p. 25) by simply grabbing any fixation or group of fixations and pulling it to a new position. It is unclear from the manual whether saccadic amplitudes and velocities also change during these data editing operations, or only fixation positions. A tip is given that a whole line of fixations can be aligned to have the same vertical value while retaining their horizontal value; this is useful in reading research. The EyeLink manual states that when batch-moving fixations like this, more than a 30 pixel movement is not acceptable; however, for those users who want to move fixations more than this, the 30 pixel setting can easily be changed.
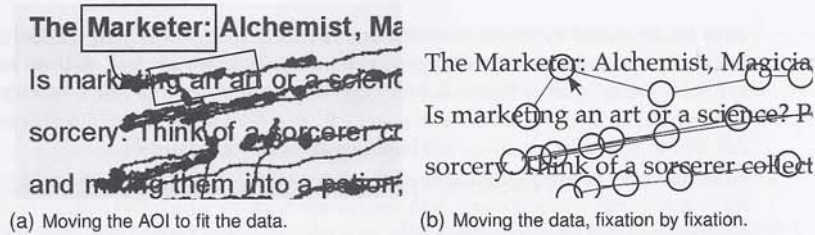
(a) Moving the AOI to fit the data.　　(b) Moving the data, fixation by fixation.

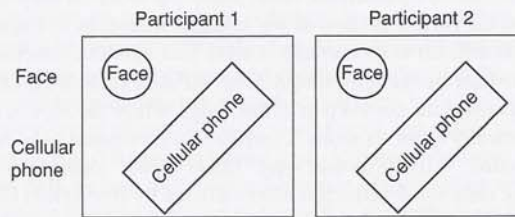**Fig. 6.35** Correcting offsets by moving (a) AOIs or (b) fixations.

Fig. 6.36 The same stimulus image, two participants and two AOIs. Normalization is only motivated when comparing between AOIs, not when comparing between participants.

For instance, Altmann and Kamide (2007) explicitly refrained from reporting statistical comparisons of the proportion of fixations between AOIs because of differences in their relative size.

Scaling is not necessary if you are only comparing between participants who looked at the same images (Figures 6.36), as the AOIs are kept constant across the comparison. However, when comparing between AOIs within one stimulus image, or between the same AOIs in different stimuli, sizes and positions differ across the comparison.

Note that the scaling factor—or function—is not easy to find. Scaling can be motivated for three different reasons, of which only one is easy to use.

Size If gaze can be expected to be *equally distributed* across the monitor—uncommon due to central bias—and two AOIs in your stimulus have *quite different sizes*, but only *small or unknown semantic differences* between them, we can then expect that gaze will be equally distributed across the stimulus, so that larger stimuli will have more data samples just because of their size. For instance, in their analysis of social stimuli, Birmingham, Bischof, and Kingstone (2009) normalize for AOI sizes by dividing proportion values by AOI area, and report large differences compared to non-scaled proportion values, in particular for the eye and head regions of people in social scenes. Using photographs of parks, Nordh (2010) also found that dwell time positively correlates with size, so that scaling by area is motivated. The reading depth measure for newspaper items on page 390 is one solution for scaling dwell time by AOI area.

Position If your stimulus is so constructed that there is one central AOI and four AOIs in the corners, you can expect that the central AOI will receive more attention simply because of the *central-bias effect* (Tatler, 2007). You may then need to scale down dwell time values on the central AOI, but it is currently unclear exactly what function to use for this.

Content Two AOIs may have very different contents, so that *in one case the task requires more dwell time* than in the other. An obvious example would be two AOIs of the same

size but containing different length text within them. The longer text would invariably have a higher dwell time, but not because it was more difficult to process. Two pictures with different content (number of faces, for instance) may also have different require-ments of gaze behaviour that can be taken as a baseline and used for normalization. In this case, scale by the unit causing the difference; for instance number or words or number of faces.

6.6.7 AOIs in gaze-overlaid videos

If data were recorded onto gaze-overlaid video only, as is typically the case with head-mounted eye-trackers without head tracking, the AOI hits must be coded manually, which may or may not be very time consuming. Once data have been coded, however, the same AOI measure may be used for gaze-overlaid data as for coordinate data. To be able to use AOI-based statistics is often reason enough to spend many weeks coding hours of video data. We will discuss current possibilities and some future directions.

Researchers are still lacking quick and robust methods for analysing gaze-overlaid scene video from head-mounted eye-trackers with no head tracking. Each of the methods below has serious limitations: they are either very slow, or they allow only for a small number of AOIs in a very limited space. Assume that in your study, you want to record eye movements from one hundred participants buying their groceries, and that they all go through the same supermarket. Such a study can be operationalized and conducted, but will have at least one thousand AOIs that you would want to code for. The ideal system would let me code AOIs using only *one* of the 100 video files, and then the computer would do the rest of the work for me. And, of course, the same coding method should also work for car-driving studies where all participants drive the same route.

The coding methods here give dwell times, that is dwell start, stop, and order of dwells. Such a coding renders possible the majority of measures defined for AOIs, that is the dwell- and transition-based measures, but not those that involve fixations. Fixations can be coded either from the gaze-overlaid video, (p. 175) or if contamination by smooth pursuit is accept-able, by using a velocity-based algorithm on gaze coordinates (Chapter 5).

Frame-by-frame coding from video Use a video player that allows you to play the gaze-overlaid video one frame at a time. When the overlaid gaze marker has reached an AOI, you start counting the number of frames until it leaves the AOI again. Such a procedure gives you the dwell time measure, in the unit of frame time. The AOI in itself is only implicit (absolute in stimulus coordinates, dynamic in head coordinates), since the coding is done on the basis of the actual semantic area rather than a geometric representation put on top of it. You can have as many AOIs as you want. This is a very general, but quite time-consuming method that has been used in studies of supermarket decision making (Vikström, 2006), newspaper reading (Garcia & Stark, 1988), cricket batsmen (Land & McLeod, 2000), gestures and face-to-face interaction (Gullberg & Holmqvist, 1999, 2006), and many other applied studies where it has been the only possible option.

Inter-coder reliability is virtually never reported, probably because the frame-by-frame coding method is considered very precise. If the study requires only limited sequences of the video to be analysed, such as the few minutes before a particular shelf in a supermarket, rather than the entire 30 minutes of shopping, it is fairly achievable to use such manual frame-by-frame coding. However, head-tracking or marker-based systems should be considered as a time-saving alternative, whenever possible.

Simulating gaze movement by hand motion Connect a graphic tablet to your computer, and draw the AOIs on a paper that you put over the tablet. Try to make the layout of the AOIs similar to that of the actual stimulus. For example, if you are studying an air traffic control station, and the communication radio used is to the right of the radar monitor in the scene, then put the radio AOI to the right of the radar AOI. Have the computer program learn where the areas are on the tablet, so when you hold the tablet pencil over an AOI, the computer program logs the time it stays there. If you hold the pencil over the radio AOI for five seconds, you get a 5 second dwell time mark in the data file. When the tablet and the computer are all set up, you play back the video at half speed, look at where the gaze marker goes, and move the tablet pencil to the corresponding AOIs on the tablet. A system like this was developed in the late 1990s at Halden research station in Norway, which Hauland (2002) used when coding 56 hours of analogue gaze-overlaid video data recorded from air traffic controllers. Hauland had a student do the same for inter-coder reliability (which was moderate at 70–75%). This is a fairly fast form of data coding, running at half recording speed, but it does not allow for more AOIs than there is room for on a tablet. Neither can you have more AOIs than it is possible to learn the positions of; at most 15 or so. This method of coding fits rather static stimuli, like control boards in aeroplanes, nuclear plants, and air traffic control operating rooms.

Dynamic AOIs in head-mounted videos As with any dynamic stimulus, dynamic AOIs can be used to code data from head-mounted videos. Put the dynamic AOIs on top of the areas in the gaze-overlaid scene video that you want for AOIs, and adjust for form, size, and motion changes. Since these AOIs are in the same coordinate system, again that of the scene video, as the coordinates in the data files, you can use the normal AOI inclusion algorithm to calculate the measures. This is a useful method if your participants make few and slow head movements, and if there are not too many AOIs in the scene. For car-driving studies, with AOIs for rear mirrors and internal controls, it works particularly well. For studies of consumers in supermarkets, the AOIs are too many, and the head movements too fast.

Computer vision solutions Use computer vision algorithms to calculate what parts of the video frames should be made into AOIs. If the contrast, patterns, and colour vary a lot between the different areas and the background, you stand a fair chance of succeeding. An extreme case to exemplify this would be if you are measuring how much students look out of a bright window compared to at the blackboard. Since the calculated AOIs will have the same coordinate system as the data file, namely that of the scene video, the calculated AOIs will fit the gaze data coordinates just as well as if we had shown the video on a monitor. Motion blur and other imperfections in the scene video can easily make the project difficult, and if the stimulus has only small visual differences between AOIs, it is not the way to go. An alternative solution involves markers. For instance, a simple and computationally tractable version of this is to attach black and white, or for that matter infrared markers with specific patterns onto the stimulus, and let the computer vision algorithm use the markers as corners of a coordinate system (essentially a plane) in which AOI can be defined. Marker-based systems typically have drawbacks, in particular a limited operative range. Markers may additionally look odd in many environments and distract natural viewing behaviour in a way that may make your study difficult to publish. An alternative could be to place the marker in the scene video—rather than on the stimulus—and have the software learn the image statistics around the marker and build a model from that.

Head tracking Knowing the position and direction of the participant's head in the measure

environment gives the same result as using markers: within a certain range, coordinate systems can be defined in the measure space, and AOIs set up in them. Head tracking has mostly been magnetic, requires a certain knowledge to calibrate and set-up, but can be combined with high-speed head-mounted eye-trackers.

6.7 Summary: events and representations from AOIs

A large number of events in, and representations of, eye-movement data are built upon what we call *areas of interest* (AOIs). They are regions in the stimulus that are interesting with respect to the experimental design, and are used to quantify whether and how much participants looked at the particular regions. There are different types of AOIs, and using them comes with a number of challenges.

Basic AOI events are calculated from raw data samples or events over AOIs:

- **AOI hit** with at least information about the fixation or raw data sample.
- **Dwell** events with at least AOI name, starting time and duration values, and information about the number of fixations.
- **Transition** events with at least starting time and duration, and names of exit and entry AOIs.

Each such event has its own values. A dwell, for instance, has a duration and a starting point. These values will later appear as measures, and as parts of measures. The following derived AOI events are often seen in the literature:

- **The return** event with information about the AOI name and the time.
- **The first skip** with AOI name and time.
- **The total skip** with AOI name.

Representations of eye-tracking data that draw on AOI division are:

- **The AOI string** A sequence of fixations or dwells in AOIs, such as MMTCCHGM, where each letter is a fixation in an AOI, and the order corresponds to the sequence in which the AOIs are looked at.
- **The dwell-map** The visualization of a gridded AOI in which each cell is given the value of the average or summed dwell time for data in it. As we will see in the next chapter, this representation of data is essentially a down-sampled attention map. Not only dwell time is used, gridded AOIs can be filled with a variety of measures that give rise to visualizations of *how early* different parts of the image have been looked at, for instance.
- **The transition matrix** A two- or higher-dimensional catalogue of the number of transitions or transition sequences of each kind.
- **The Markov model** A probabilistic model describing or modelling the data in a transition matrix.
- **The proportion over time graphs** An important representation for studying processes over time, with many varieties, including **sequence chart, scarf plot, cumulative proportion graph**, and **proportion of transition sequences over time**.

In the remainder of the book, we will very often refer to AOI events and representations defined in this chapter.

AOIs can be productively combined with fixation- and saccade measures to produce a range of other measures which take both aspects into account; for instance first fixation in an AOI, saccadic amplitude within an AOI, or total dwell time. In addition, AOIs allow for

substitution of the spatial dimensions with feature dimensions derived from the position. Remember, however, that the values in all these events and representations, as well as the validity of your conclusions, depend crucially on how you segment space with AOIs.