Chapter 4
Eye-Gaze

Abstract This chapter considers the principles of eye movements and how eye-gaze recording and analysis techniques can be used to study and understand the consequences of design decisions. First we look at the different ways in which gaze naturally reacts to different RSVP-like presentations and describe four distinct gaze behaviours sufficient to characterise the significant aspects of eye movements observed during RSVP sequences: (1) visual search, (2) steady gaze, (3) nystagmus, and (4) visual pursuit. The chapter concludes with a discussion of ways of representing and presenting eye-gaze information to best effect.

Keywords Eye-gaze recording • Saccades • Fixations • Visual search • Nystagmus • Visual pursuit • Fovea • Representing eye-gaze data

In the course of the last chapter we saw that users often tend to prefer 'static' RSVPs, where each image exists in only one location, rather than 'moving' RSVPs in which a given image can move around the display. We also saw that the search for a target image tended to be more successful with static presentation modes than moving ones. One obvious question to ask is "Why?"—and, equally, "what can an understanding of eye-gaze contribute to RSVP design?"

4.1 Why is Eye-Gaze Important in RSVP Design?

To seek an answer, Cooper et al. (2006) and others have recorded the eye-gaze of subjects given the task of identifying the presence or absence of a target image within a displayed collection of images. How gaze is detected and recorded is discussed later in this chapter when we know what it is we need to discuss: suffice it to say for now that eye-gaze is typically 'on the move', controlled by muscles around the eyeball, often in very characteristic ways. It is well known that a person's gaze adapts not only to the visual environment, but also to the activities and tasks they are performing (e.g. Yarbus 1967, for a classic demonstration).

R. Spence and M. Witkowski, *Rapid Serial Visual Presentation*, SpringerBriefs in Computer Science, DOI: 10.1007/978-1-4471-5085-5_4,

In normal image viewing conditions, gaze will typically be directed to one location on a display screen for about 300 ms (this is termed a *fixation*) and will then move to another location extremely rapidly, for example in about 20 ms (that movement is termed a *saccade*). Various RSVP modes, however, give rise to a wide range of different gaze behaviours, which in turn can have an impact on how the mode is viewed, how long the images must remain visible, how big the images must be, how effective the mode is, and the degree to which the user will enjoy using it.

Users are generally unaware of their own eye movements, but an understanding of gaze behaviour can give the interaction designer an indication of what will be noticed by the user and what might be missed, as well as some insight into the degree to which a user might prefer using a particular mode of presentation.

4.2 Gaze Strategies

In this section we consider four distinct eye-gaze strategies that people use when looking at different RSVP modes: visual search, steady gaze, nystagmus and visual pursuit. When viewing any particular RSVP mode users will immediately adopt one of these gaze strategies, or a combination of them, in characteristic ways for the duration of the presentation.[1] Later in Chap. 5 we examine ten different RSVP modes and describe and classify each in terms of these four types of gaze behaviour.

4.2.1 Visual Search

An example of gaze behaviour typical of a person viewing a static picture is provided in Fig. 4.1 which shows a sequence of fixations (coloured cyan and denoted as 'F') connected by rapid saccades (coloured yellow). It is seen that gaze movement is neither regular nor uniform: at times the gaze location moves right across the picture while, at other times, it moves to a nearby feature. Also, certain features of the picture—human faces, for example—receive more of the user's attention than the background architectural structures and the table.

Saccades are very rapid movements of the eyeball, attaining an effective angular rotation of up to 600°/second (Becker 1991). As pointed out in the discussion of saccadic blindness in Chap. 2, the viewer is momentarily effectively "blind" during these movements, although, as with blinking, there is no awareness of this in normal vision. Saccades are "pre-determined": gaze jumps automatically to a place determined by the brain's visual system during the preceding fixation(s).

[1] Tatler and Wade (2003) give a description and history of various types of eye movements; Findlay and Gilchrist (2003) discuss their broader context in human vision.

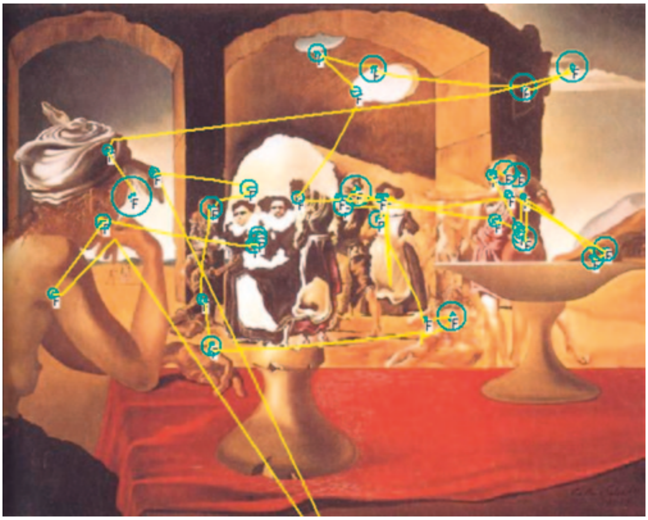

Fig. 4.1 Typical pattern of fixations and saccades when viewing a picture

There is a limit to how far gaze can be shifted during a saccade and if a movement of more than about 15º is required, the head will also rotate automatically.

The software used to represent the fixation and saccadic information in Fig. 4.1 has identified each fixation with an ‘F’ and a cyan point and circle, the diameter of the latter being proportional to the duration of a fixation. Fixation durations vary considerably. In the 12.5 s recording there are 38 fixations of average length 274 ms, accounting for about 83 % of the total time. For comparison with the other three gaze strategies to be introduced, this visual search strategy is summarised diagrammatically in Fig. 4.2a.

Note that certain commonplace activities, such as reading, also give rise to distinctive gaze patterns. Reading, for instance, is characterised by an ordered sequence of fixations along each line of text, though generally not equating to recognised word boundaries (e.g. McConkie 1983; Rayner 1998). A variety of different gaze strategies relevant to RSVP are shown in Fig. 4.2 and are described in the sections that follow.

4.2.2 Steady Gaze

If images are presented at a typical RSVP rate of 10 per second there is no time for gaze to move around an image in a characteristic fixation + saccade pattern; rather, there is a tendency for gaze movements to cease, often becoming centralised in a relevant location for extended periods of time. We call this behaviour steady gaze, summarised diagrammatically in Fig. 4.2b.

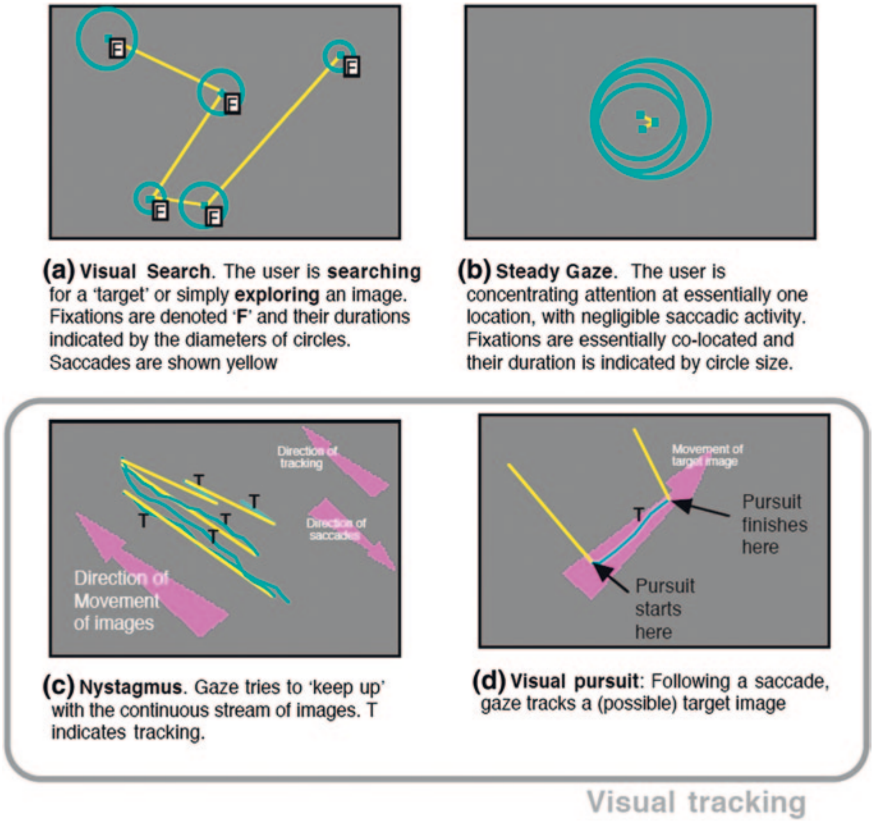

Fig. 4.2 Diagrammatic representations of the four gaze strategies. **a** Visual search. **b** Steady gaze. **c** Nystagmus. **d** Visual pursuit

Steady gaze can be distinguished from the intentional act of "staring" at something, in that the user adopts the steady gaze pattern without any voluntary thought or intention. It has been suggested that the observed improvements in text RSVP are largely due to the absence of saccades due to the single viewing location (Rubin and Turano 1992), although this is unlikely to account for all the improvements reported.

4.2.3 Nystagmus Gaze Tracking

Not all image presentations are static: in many of the RSVP modes we have encountered images follow defined paths at a range of speeds. Accordingly, human vision adapts to accommodate to these circumstances. Of particular relevance to gaze behaviour are adaptations to the movement of either images on a display or physical movements by a person—the phenomena called *nystagmus* and *visual pursuit* are two variants of what is generally termed ‘visual tracking’.

Nystagmus can perhaps best be explained by reference to a common experience, that of looking out of the window of a fast moving train and attempting to ‘keep up’ with the view rather than examine some feature within that view. The user is trying to stabilise the larger part of the visual field. Similarly, when observing rapidly moving images in an RSVP mode the eye will also adopt these characteristic nystagmus[2] movements. When viewed as angular movement with time, these movements have a distinctive repetitive “sawtooth” motion, a rapid ramp of movement followed by an even more rapid (saccadic) return. An illustrative diagrammatic representation of nystagmus is shown in Fig. 4.2c, and will be discussed in detail later in the context of moving RSVP mode designs (Sect. 5.2).

4.2.4 Visual Pursuit

While nystagmus is a response to generalised rapid movement in the field of vision, gaze response to specific moving objects is different. In these circumstances gaze may be seen to follow an object, much as a photographer might swing his camera to follow a moving object, keeping it sharp at the expense of a blurred background. We refer to this gaze behaviour as *visual pursuit*.

A diagrammatic representation of a visual pursuit event in response to a moving image can be seen in Fig. 4.2d. Visual pursuit will be discussed in detail in the context of multiple entry/exit RSVP modes in Sect. 5.3. There are limits to the speed of movement that can be accommodated by these forms of active gaze tracking and we consider this later. The gaze analysis software used can isolate either type of tracking event and shows them as an elongated cyan trace, marked with a ‘T’ to distinguish them from fixations (‘F’).

4.3 Structure of the Eye

Two questions immediately arise from the evidence presented in Fig. 4.1 above. First, why does gaze have to be directed between a number of locations in a scene? Why can’t the user simply take one ‘look’ at a scene and immediately know it in all its detail?

The answer lies in the make-up of the retina (Fig. 4.3, right) at the back of a user’s eye (Fig. 4.3, left). A relatively small and approximately circular area called the *fovea* is made up of a high density of cones (which provide colour sensitivity)

---

[2] These entirely normal gaze movements should not be confused with the medical condition also called *nystagmus*, where the eyes make continuous uncontrolled flicking movements and which has highly detrimental effects on vision: (http://en.wikipedia.org/wiki/Nystagmus).

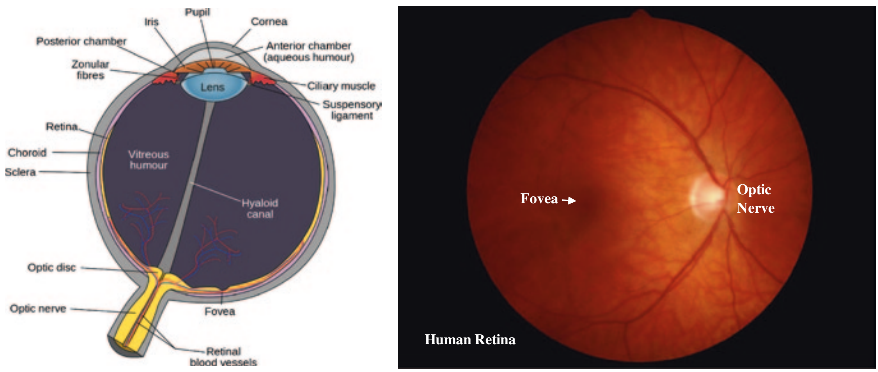

Fig. 4.3 (*left*) Anatomy of the human eye, (*right*) human retina

but a low density of rods, which do not respond to colours. The high density of cones provides extremely high acuity and sensitivity to colour within the visual angle subtended by the fovea (about one degree). The effect (and purpose) of fixations is to move specific parts of the visual scene to be within this high acuity foveal area, where detailed visual processing is possible.

Movement of the eye within its orbit is controlled by six muscles. These muscles are capable of the rapid and controlled movements of the eye, up to 600°/sec for saccades and around 100°/sec for tracking events. Equally they hold the eye stationary during fixations.

Away from the central fovea rod cells predominate. They are far more sensitive to light than cones and they are more numerous in the retina away from the fovea, but are not sensitive to colour (Fig. 4.4). A sparser covering of cones in this peripheral region maintains some colour perception. The ability to recognise detail drops off rapidly away from the central fovea.[3] As a general principle, objects must appear larger or be more distinctive to be accurately identified towards the edge of the visual field.

The effective size of the fovea is illustrated in Fig. 4.5. The angle subtended by the foveal region of the retina is around one degree, so at a distance of one metre the area of a distant display seen in great detail is about 1–1.5 cms in diameter: if you extend your arm, the width of the thumb nail is a reasonable indicator of the foveal angle. The area immediately around the fovea and subtending around 5° (roughly the width of a hand at arm's length) is called the *extrafovea* or *parafoveal region*: outside that we talk about *peripheral vision*. The display of

[3] It's true—try this perceptual party trick. Seat yourself in a normally lit room and stare continuously at a fixed point on the opposite wall. Ask a friend to move a suitable object, say a playing card, gradually towards your fixed line of sight from a few metres away. You will be surprised how close the card has to be to your sightline before you can reliably call colour, suit or rank. Of course, if you cheat and glace at the card—even for an instant—recognition is easy. See also Anstis (1974).

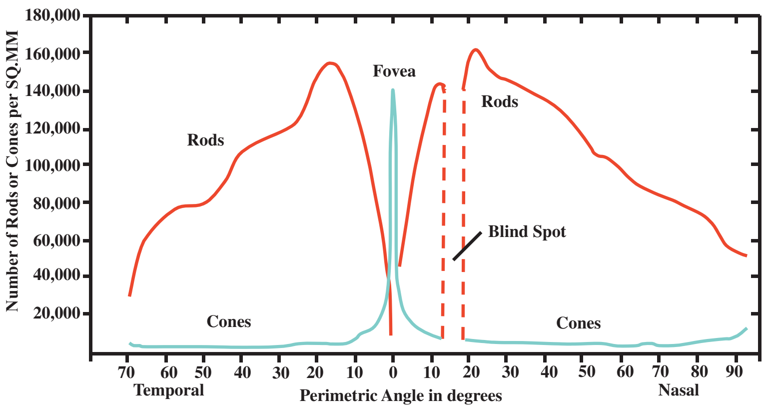

Fig. 4.4 Distribution of rods and cone cells across the fovea (adapted from Osterberg 1935)

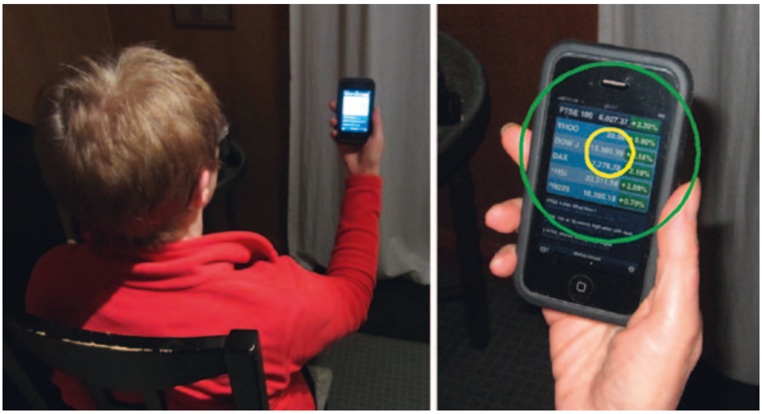

Fig. 4.5 Viewing a mobile device—approximate foveal (*yellow*) and parafovea (*green*) areas

a mobile or PDA held at arm's length falls mostly within the foveal region and completely within the parafovea, as illustrated in Fig. 4.5 respectively as yellow and green circles.

If we could only 'notice' items in the foveal region the number of car accidents would be much higher than it is. While the detail of a static image in the peripheral region cannot be deciphered in anything like the detail associated with foveal vision, *movement* in the peripheral region can, for example, easily be noticed. So can the change in colour of a traffic light. Nevertheless, the foveal angle is sufficiently small as to require the gaze behaviour illustrated in Fig. 4.1. Indeed, it may be useful to point out that, as far as the eye is concerned, the sequence of images projected onto the retina in the example of Fig. 4.1 could be considered to be a Rapid Serial Visual Presentation, especially since no rapid image movement is detected during a saccade. The question we now need to ask is what the user is actually *doing*, perceptually and cognitively, during fixations and saccades, and to explore the notions of salience and attention that would seem to lie at the heart of the task that a user is undertaking.

4.4 Where does the Eye Look Next?

Gaze control—determining what a person looks at—varies according to the context in which looking is taking place (e.g. Yarbus 1967). What the person attends to and their eyes fixate on is completely different when driving (road conditions, hazards, signposts and so on), as to when one is a passenger. If this were not so then car travel, and life in general, would be substantially more hazardous than it is. The immediate target of gaze control is subject to instant and conscious modification, over-riding the brain's autonomous choices. Nevertheless, there will be topics of continuing interest to individuals, which persist and may become evident under otherwise controlled test conditions.

There is a strong trade-off between angular distance from the fovea and image size when identifying images. As we have seen, detail is lost in the periphery of vision: an image easily identified when near the fovea may be completely unrecognisable in peripheral vision. The relationship between visual acuity and distance from the fovea is neatly encapsulated by Anstis (1974).[4] This trade-off is particularly significant for RSVP modes where small images must be accurately reported but which are located, or appear, away from the current point of gaze.[5] To be effective a mechanism must be provided to allow gaze to move to and fixate on an image, and sufficient time must be permitted to allow this to happen.

Fortunately, human vision is remarkably sensitive to general shapes and outlines, colourations, changes and movements in the visual periphery. Generally the appearance of a significant image will attract visual attention and the gaze point will move to it at the next or subsequent saccade. However, recall the problems of "cognitive blindness" (introduced in Sect. 2.5) caused by visual distraction or masking. These are of real concern in RSVP application design. If several aspects in the presentation change simultaneously, or even in quick succession, one of these changes may suppress recognition of the desired image—a manifestation of the "mud splash" effect described in Sect. 2.5.2.

There is an ordering or salience to items and images in the visual field, and this appears to determine what is attended to and which attracts gaze, and what is passed over. As the designer can rarely control the effective salience of images, care must

[4] Recall the game described in footnote 3.

[5] This effect, as we shall see, is particularly apparent in the Collage mode design investigated in Sect. 5.3.6.

be taken in an application design to ensure that items, which may not naturally be conspicuous, are not masked by potentially more noticeable and distracting images.

There is no one clear model that represents visual salience across all its properties, but several salience and eye-gaze models have been proposed. Itti et al. (1998) present a computer model of "bottom-up" (image data) driven salience, in which features (based on properties such as intensity, edge orientation, colour, etc.) are extracted from the image at multiple scales and salience assigned according to the application under consideration. Stark and Choi (1996) present a model that emulates the eye-gaze path of a simulated human observer using a Markov state model based approach. Witkowski and Randell (2007) present a model based on object and task precedence to account for rapid shifts in gaze strategy.

4.5 Detecting and Recording Gaze

A typical set-up[6] that detects eye-gaze is shown in Fig. 4.6 (left). An infra-red LED mounted in a camera lens directs its beam safely into the human eye. The camera, mounted directly below the display screen in Fig. 4.6 (centre picture), records the infra-red light reflected from both the retina and the surface of the cornea (shown in Fig. 4.6, right). A calculation based on the relative positions of the larger retinal reflection (through the pupil)[7] and the smaller corneal reflection from the surface of the eye can then determine to within an accuracy of about 2 mm the place on the display at which the user is looking.

Typically, the location of eye-gaze on the display is recorded at a rate of 50 or 60 per second (depending on the make of equipment used) and the system must be calibrated for each user before recordings can be made. Blinks—which typically

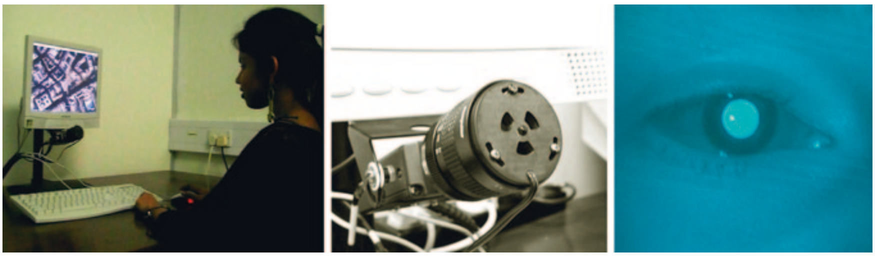

Fig. 4.6 (*left*) Eye-gaze recording set-up, (*centre*) gaze camera, (*right*) retinal and corneal reflections

[6] An early LC Technologies, Inc. system, see http://www.eyegaze.com/. Some of the data we present was captured on a Tobii system (http://www.tobii.com/).

[7] Similar to the "red-eye" effect often seen in flash photographs of faces.

last around 100 ms (equivalent to five or six gaze readings)—are also recorded. A full description of gaze recording technologies and techniques can be found in Duchowski (2003).

4.6 Representing and Analysing Gaze Recordings

The question arises as to how best to represent highly dynamic gaze data. The gaze recording equipment simply generates a long sequence of X-Y coordinates of gaze position on the display, once every 16.7 or 20 ms (60 or 50 times a second). The first task is to extract fixations. This is done automatically in software. By convention, a fixation is usually defined as a run of six or more gaze coordinate readings where the gaze location on the screen has not varied by more than a small distance between any of the readings.[8] We generally use a distance equal to a capture radius of 12 pixels, though individuals differ in the natural instability of their gaze (sometimes called eye "jitter"): increasing the capture radius can compensate for this. Once a single gaze reading falls outside the radius criterion the fixation is considered finished. This is usually due to the start of a saccade.

By changing the fixation detection algorithm to allow the fixation to continue while each successive reading (as opposed to *all* readings) remains within the current capture radius criterion, tracking events are effectively captured, as observed in both nystagmic behaviour and visual pursuit (Sect. 4.2.3). Again a single reading outside the capture radius signals the end of the tracking event. Tracking detection is an option in the analysis software, and is used extensively in analysing the moving and multiple entry/exit modes later.

Figure 4.7 (left) shows a short (3.25 s) extract from the gaze trace shown in Fig. 4.1, and a tabular form of the fixations extracted from the sequence is shown to the right. In this case the analysis software has been used to

Start	Duration	Samples	X mean	Y mean
3.917	0.267	16	400.4	277.9
4.234	0.366	19	344.3	306.2
4.650	0.184	11	329.6	443.5
4.884	0.316	19	343.2	513.1
5.517	0.450	27	704.5	468.3
6.000	0.100	6	664.3	472.7
6.150	0.300	18	593.6	322.7
6.467	0.133	8	605.8	288.6
6.650	0.217	13	768.2	324.6
6.917	0.200	12	804.8	349.1
7.167	0.367	22	802.5	256

Fig. 4.7 (*left*) Gaze track of saccades and fixations, detail, (*right*) tabulated fixation data

[8] Actually there are a number of different measures to define a fixation (e.g. Salvucci and Goldberg 2000), but in practice the differences between them are not significant for the type of analysis we are conducting here.

additionally show the fixation start times (in seconds from the beginning of the recording) and durations (seconds) on the trace. This form of representation is very useful for short sequences and details of particular segments of gaze behaviour, but quickly becomes cluttered as more of the gaze trace is included.

4.6.1 Temporal Representation

We use an XY-T plot to graph the X and Y position of the gaze point on the display as a function of time. Figure 4.8 shows this gaze representation method for the case of a visual search. The X (horizontal) coordinates on the display are shown in red and the Y (vertical) coordinates in blue. In these XY-T plots the top-left coordinates of the display (Y = 1, X = 1) are both shown at the top of the graph. Fixations appear as plateaux in both X and Y, and saccades as abrupt vertical changes. Any blinks would appear as short interruptions to the plot. Temporal representation of RSVP was pioneered by de Bruijn and Spence (2002).

Figure 4.9 shows an example of an XY-T plot for a steady gaze example (refer back to Fig. 4.2b and Sect. 4.2.2). It can be seen that the X and Y coordinates

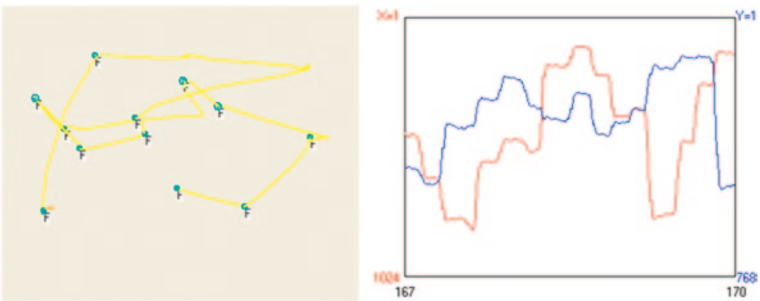

Fig. 4.8 XY-T plot representation showing a pattern of fixations and saccades

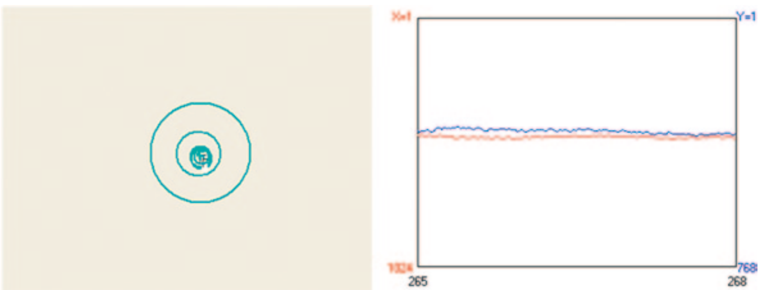

Fig. 4.9 Gaze track and XY-T plot for steady gaze

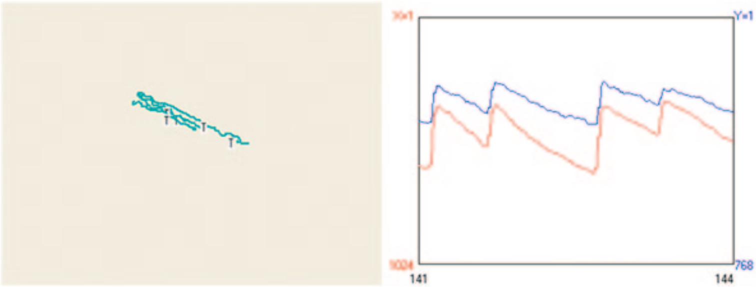

Fig. 4.10 Gaze track and corresponding XY-T plot for nystagmus

of the gaze location on the screen are stable. Slight movements in the gaze point cause the fixation algorithm to register several distinct fixations.

Temporal representation can also be valuable in clarifying the motion of the gaze point on the display in highly dynamic situations. Figure 4.10 shows an XY-T plot for a nystagmus mode sequence. Note the distinct “sawtooth” appearance of the plot. Tracking detection has been used in this example to highlight the smooth movement of the gaze point (marked with a ‘T’) as it follows the general direction of motion on the display. For clarity, the yellow saccadic movements are omitted from the gaze track image (compare with Fig. 4.2c).

Figure 4.11 shows an example of a single visual pursuit event during which gaze closely follows a specific moving target. Again tracking detection is enabled. The XY-T plot shows a distinct beginning and end to the tracking movement, starting and ending with a (yellow) saccade. Notice that the pursuit event is flanked by steady gaze periods, before and after. The sequence of events shown begins with a period of steady gaze (the first horizontal portion of the XY-T plot), which ends with a saccade (the abrupt vertical change) leading to the beginning of the tracking event (the downward slope). The tracking event terminates with another saccade leading to a second period of steady gaze. The small magenta segment embedded within the cyan tracking event is used to indicate a user response.

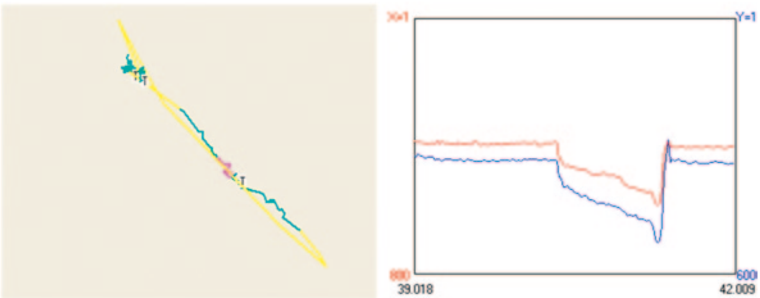

Fig. 4.11 Gaze track and corresponding XY-T plot for visual pursuit

4.6.2 Heatmap Representation

Another useful representation of gaze behaviour—the "cumulative heatmap"—is shown in Fig. 4.12. Heatmap representations are particularly useful for getting an overall idea of gaze activity across many trials or samples. Colours represent the density of gaze activity at locations on the display over an extended period of viewing or accumulated across several users.

Compare, for instance, the "raw" gaze track, mostly saccades shown in yellow, for a user viewing Volcano mode for 223 s (Fig. 4.12, left), with the equivalent cumulative heatmap shown in Fig. 4.12, right. Both show the same trial data of over 11,000 individual gaze point readings. In the heatmap the highest levels of gaze activity are shown in red, while blue represents an absence of activity, with intermediate levels of activity coloured according to the scale shown to the right. All the heatmaps shown are scaled and coloured in this way. When comparing like recordings, a heatmap with large red areas does not imply more activity overall, only that the gaze activity is more evenly distributed over the display.

4.6.3 Gaze Travel

One further useful measure of gaze activity is "gaze travel". In this measure the total accumulated distance travelled by the gaze point across the display is calculated as the sum of the Euclidian distances between successive gaze point recordings. As the distance travelled during saccades is much higher than during fixations, gaze travel represents a useful single measure of how extensive gaze movement was during a period of recording. We will usually present this distance in "screen pixels". A related measure to indicate gaze activity is that of gaze speed, expressed in screen pixels per second (pix/s). Where the geometry of the display and viewing distance are known, gaze travel can be expressed as effective angular travel or angular gaze speed of the eyeball (degrees or °/s).

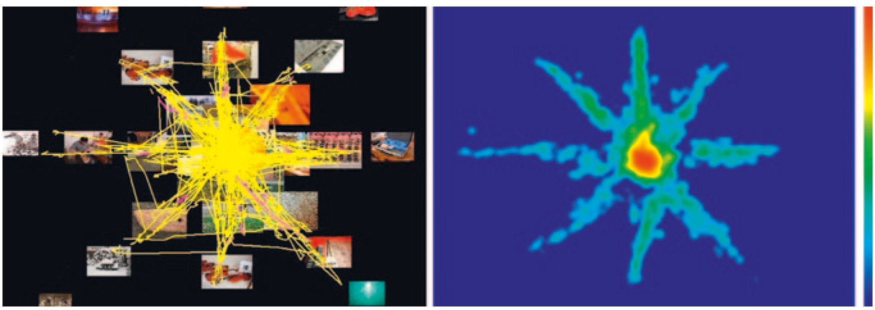

Fig. 4.12 Volcano mode: (*left*) gaze track and (*right*) equivalent heat map representation

Gaze travel measures are useful for comparing like recordings. For instance a presentation where static gaze predominates will exhibit a low gaze travel distance, while, for example, a nystagmic one will be characterised by very high travel distance.

In the next chapter we present an analysis of user gaze behaviour for ten of the RSVP modes described in earlier chapters: three "static" modes, three "moving" modes and four "multiple entry/exit" modes. We will see that gaze behaviours vary widely according to the RSVP mode and that the gaze strategy will adapt naturally according to the task the user has been set.

References

Anstis, S. M. (1974). A chart demonstrating variations in acuity with retinal position. *Vision Research, 14*, 589–592.

Becker, W. (1991). Saccades. In R. H. S. Carpenter (Ed.), *Eye movements* (Vol. 8, pp. 95–137), Vision and visual dysfunction Boca Raton: CRC Press.

Cooper, K., de Bruijn, O., Spence, R., & Witkowski, M. (2006). A comparison of static and moving presentation modes for image collections (pp. 381–388). *Proceedings of Advanced Visual Interfaces (AVI-2006).*

de Bruijn, O., & Spence, R. (2002). Patterns of eye gaze during rapid serial visual presentation (p. 11). *Proceedings of AVI-02.*

Duchowski, A. T. (2003). *Eye tracking methodology: Theory and practice*. New York: Springer.

Findlay, J.M., & Gilchrist, I.D. (2003). Active vision: The psychology of looking and seeing. Oxford: Oxford University Press.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Pattern Analysis and Machine Intelligence, 20*, 1273–1276.

McConkie, G. W. (1983). Eye movements and perception during reading. In K. Rayner (Ed.), *Eye movements in reading* (pp. 65–96). New York: Academic Press.

Osterberg, G. (1935). Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica, 13*(6), 11–103.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Rubin, G. S., & Turano, K. (1992). Reading without saccadic eye movements. *Vision Research, 32*(5), 895–902.

Salvucci, D.D., & Goldberg, J.H. (2000). Identifying fixations and saccades in eye-tracking protocols (pp. 71–78). *Proceedings of the Eye Tracking Research and Applications Symposium*. New York: ACM Press.

Stark, L. W., & Choi, Y. S. (1996). Experimental metaphysics: The scanpath as an epistemological mechanism. In W. H. Zangemeister, et al. (Eds.), *Visual attention and cognition* (pp. 3–69). Amsterdam: Elsevier.

Tatler, B. W., & Wade, N. J. (2003). On nystagmus, saccades and fixations. *Perception, 32*, 167–184.

Witkowski, M., & Randell, D. A. (2007). A model of modes of attention and inattention for artificial perception. *Bioinspiration and Biomimetics, 2*, S94–S115.

Yarbus, A.L. (1967). Eye movements and vision. New York: Plenum (Originally published in Russian 1962).