Chapter 6
Design

Abstract An interaction designer wishing to explore the potential of RSVP for a given application would expect to be provided with a concise set of guidelines requiring only a minimal understanding of the underpinning empirical evidence and theory. Both objective (e.g., image recognition) and subjective (e.g., fatigue) performances are of concern. Six questions typically asked by interaction designers concern those parameters that influence performance: they include the choice of visual style and the number of images simultaneously visible on the display. Following the presentation of design guidelines, examples of applications based on RSVP are provided.

keywords Interaction design • Design guidelines • Visual style • Presentation rate • Image size • Manual control • Simultaneous presentation of images • Applying RSVP

An interaction designer contemplating the use of RSVP in an application may wish to proceed directly to this chapter, but will consequently be unaware of the design considerations gradually accumulated in the earlier chapters. For this reason the present chapter will contain *brief* recollections of those considerations and the basis from which they were derived. The chapter then finishes with a discussion of a number of RSVP application designs as well as pointers to potentially useful directions of research.

6.1 User Tasks

An interaction designer minded to employ an RSVP mode in an application will have to consider not only the user tasks that require support but, in addition, user acceptance of the designed interface.

As pointed out in Chap. 1 it can be useful to consider two different tasks in which a user may be engaging, namely *exploration* and *search*. In some applications one or the other may dominate, but it is important to remember that even

R. Spence and M. Witkowski, *Rapid Serial Visual Presentation*, SpringerBriefs in Computer Science, DOI: 10.1007/978-1-4471-5085-5_6, © The Author(s) 2013

when the principal activity is search the user may, consciously or unconsciously, be forming a mental model of the information space encountered during that search. Indeed, in many situations, the visibility of context can be a most important consideration. Equally, to achieve their goal, a user may switch from exploration to search quite rapidly and frequently.

6.2 Issues in Design

For any task there will be two outcomes that the interaction designer will be striving for. One can be measured objectively, the other subjectively.

One is *task performance*. One example of measurable (objective) user performance is the accuracy of image identification in the course of undertaking an image recognition task. Another is the acquisition of the gist of a film. Yet a third is the time taken to locate a known image.

Subjective outcomes can include a user's perception of the fatigue they encounter in the execution of a task, their view as to whether the rate of presentation of images was too slow or too fast as well as their confidence regarding perceived task success. An outcome that is very difficult to predict is the extent to which an interface is visually attractive.

6.3 The Design Process

To develop guidance for the interaction designer we briefly reflect upon the design process that might be followed. In the very early, 'idea generation' stage of design the designer, with an intended application in mind, might well first sketch (Fig. 6.1a) the route to be followed by a collection of images, then tentatively indicate the approximate location of image frames/outlines on that route (Fig. 6.1b), perhaps also then adding (Fig. 6.1c) the—often fixed—location of appropriate text and other screen furniture associated with the application. Here, appropriate notation (see Chap. 3) indicates that image movement is continuous, and the image frame separation provides some indication of relative image speed at different points along the trajectory.

In the form shown in Fig. 6.1c the specification for the RSVP design is nevertheless incomplete. Lacking are two numerical parameters: one is the rate at which images appear (pace, $\boldsymbol{p}$, images per second) and the other is the transit time ($\boldsymbol{T}_{\mathbf{A}}$) for any given image—in other words, the time for which an image remains visible. Perhaps tentatively, the designer might annotate the sketch, as in Fig. 6.1d, with values for these parameters (their choice is discussed in the following section). However, one detail may be added at this point. The designer may feel that, notwithstanding the lack of supportive experimental evidence mentioned in Chap. 3, the application would benefit if, at one location, each image is 'captured'. If that

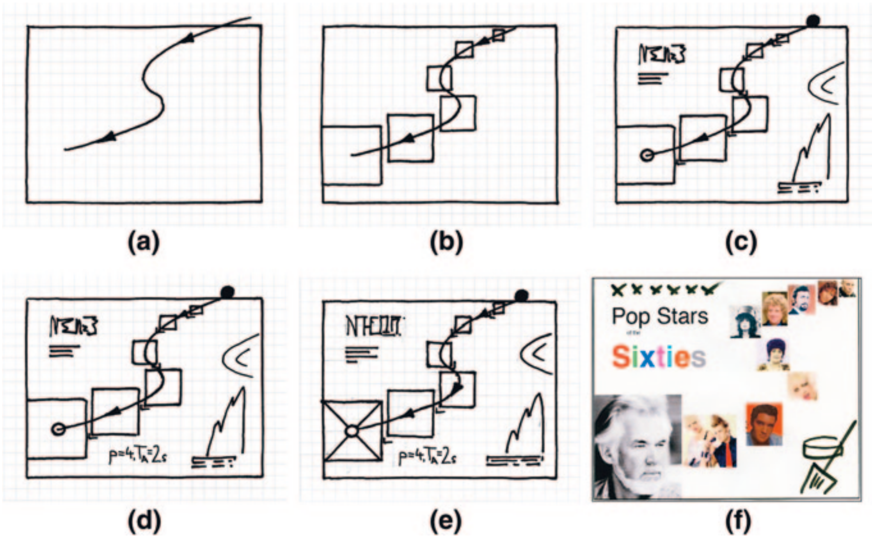

Fig. 6.1 A possible sequence of sketches created in the course of a first proposal for an RSVP application

is the case, the location and duration of the captured image might be as illustrated in Fig. 6.1e. The series of sketches might usefully lead to the creation of a paper prototype (Fig. 6.1f) to support formative evaluation by the designer as well as colleagues, client and representative users, and will often lead to design iterations. The approach illustrated in Fig. 6.1 can be taken in connection with any of the many modes introduced in this book.

6.4 Design Questions

The interaction designer will usually have five major questions in mind during the design activity:

- What is the most effective visual style (mode) for RSVP?
- How quickly can a collection of images be viewed?
- How large should the images be?
- Should presentation rate be fixed or user selected/controlled?
- How many images should simultaneously be visible on the screen?

Starting with the sort of questions a designer might typically ask and then proceeding to answers—as we do in this chapter[1]—may at first appear unusual. It can

[1] The approach adopted in this chapter is close to that described in Witkowski and Spence 2012.

be justified, however, in many ways. First, the questions listed above are arguably the principal factors in an RSVP design. Second, as in many other fields in which design typically proceeds with many factors considered in parallel, no step-by-step design methodology has emerged for image-based RSVP. Third, there are so many factors to jointly optimise, some of which are associated with aesthetic considerations, that flexibility in design decisions is paramount. In fact the guidelines that follow do not assume any sequence of design decisions.

We examine in turn each of the five questions listed above and establish corresponding design guidance, notwithstanding the fact that the topics referred to in those questions are not 'orthogonal' and that some iteration of design decisions may therefore well take place.

6.5 Design Answers

6.5.1 Visual Style

Especially for commercial applications a major consideration regarding visual design will be the degree to which an interface is found to be 'attention grabbing' and pleasurable. Such a property is difficult to define: even less does it lend itself to algorithmic description. Judgement of a design in this respect lies in the domain of an interaction designer's expertise, and is therefore not discussed further here. The interaction designer's expertise will undoubtedly benefit from the numerous books devoted to visual design (see, for example, Few 2009).

6.5.1.1 Static Mode

In some situations the answer to the question "What is the most effective visual style (i.e., mode) for RSVP?" may well be "a static mode", of which the Slide-show mode is the most commonly encountered in practice.

Common reactions to Slide-show mode point out the consequences of blinks (that can occur naturally at quite a rapid rate), and the fact that it can be quite stressful gazing at one point continuously for more than a few seconds, especially in the knowledge that each image is only visible for (say) 100 ms. For these reasons the Slide-show mode may only become of significant interest within a task such as that carried out by video editors when searching for an image in a collection of frames or a TV viewer engaging in fast-forward and rewind, especially where manual control of direction and speed are provided and when the visual display also represents context (Wittenburg et al. 2003b). At one extreme the provision of visual context (e.g., to help identify an approximate location) may justify a very high rate of presentation, while at the other the need to select exactly the right image may call for a very slow and manually controlled rate.

The Slide-show mode of RSVP may sometimes be relevant in circumstances closely related to the original ‘book riffling’ activity, as with the rapid presentation of key frames of a video in order to provide the gist of that video within a very short time. On TV channels, trailers of upcoming films are often presented at image rates not too far removed from ten per second.

6.5.1.2 Moving and Multiple Entry/Exit Modes

Moving RSVP modes are generally relevant to applications directed towards the consumer market, in sharp contrast to the professional market for which the (static) Slide-show mode may at times be appropriate. For consumer use heavy emphasis may well be placed upon the visual appeal of the application but, as explained above, it is not the function of this book to provide guidance in that respect. By contrast, firm guidance is provided immediately below that will apply to *any* RSVP mode involving image movement.

6.5.1.3 Gaze Travel

The discussion of eye-gaze behaviour in Chap. 5, and in particular the outcome of Corsato et al. (2008) experiment, strongly suggests that whatever trajectory is adopted by the interaction designer as the route to be taken by images, those images should *first* become capable of recognition at an essentially *single* location. That this criterion is satisfied for the Volcano mode shown in Fig. 6.2 (top) is suggested by the accompanying heatmap of gaze activity. It is not satisfied for the shot mode as demonstrated in Fig. 6.2 (lower).

6.5.1.4 Image Speed

The experimental evidence discussed in Chap. 5 suggests that image speed might usefully be kept comfortably below 1,000 pixels per second. A rough check of this requirement can easily be estimated from the sketch of Fig. 5.1d from knowledge of the total visibility time T_A of an image and the length of the image path.

6.5.1.5 Capture Frame

It is generally believed that the value of a moving RSVP mode can be enhanced if, at some fixed location, every image is captured for a short while, typically for 100–200 ms. As already mentioned, there is no experimental evidence to support this guidance: from what has been said in earlier chapters the interaction designer must judge, from knowledge of the application, whether a capture frame will enhance performance. A judgement in this respect might well take place during the first review of a prototype.

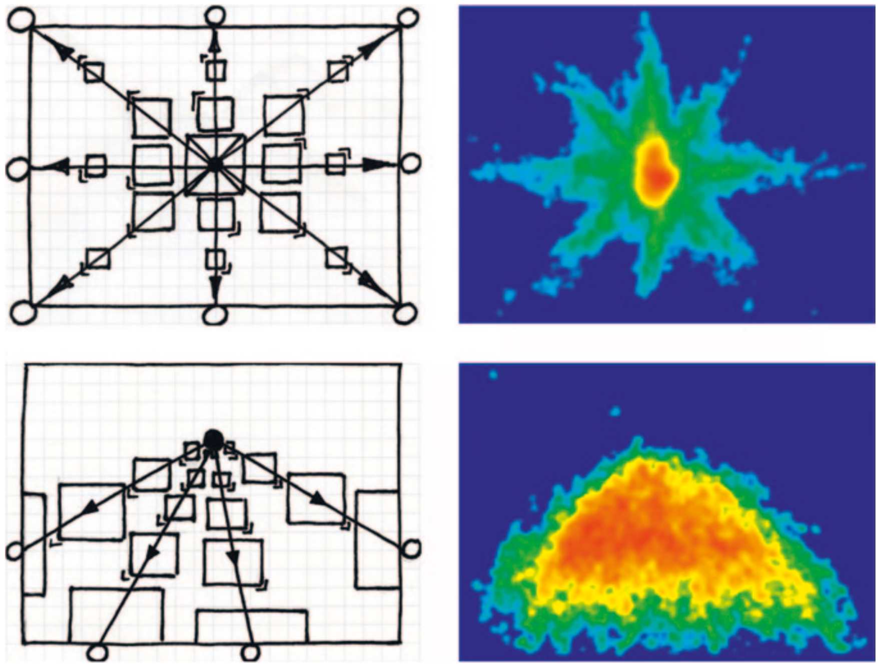

Fig. 6.2 Notational description and gaze heatmap for a volcano mode (*top*), notational description and gaze heatmap for a shot mode (*lower*)

6.5.2 Task Completion Time

"How quickly can a collection of images be viewed?" is a question motivated principally by the common need for a task to be completed within a reasonable time. Again, it is appropriate to answer this question separately for static and moving modes.

6.5.2.1 Static Mode

For Slide-show (static) mode the only parameter available for design is the pace, p—the rate at which new images appear on the display. Reference to Chap. 2 will confirm that a value of pace equal to ten per second should be regarded as a maximum, in which case the time taken for the collection of N images to be seen is at least N/10 s. However, as discussed above in Sect. 6.5.1.1, a task for which Slide-show mode is appropriate (for example, a video editor's exploration and search—both fast and slow—of a video sequence) would almost certainly benefit from an extensive range of pace under manual control.

6.5.2.2 Moving Modes

For moving modes a wider design flexibility is available, though a first consideration is simple: the higher the pace (p), the shorter in general will be the total image collection presentation time. An additional factor is the degree of overlap between successive images. The discussion of Chap. 3 showed that some degree of obscurement—even up to 50 % in some situations—might be tolerable for many types of image, and also identified the existence of a trade-off between pace and recognition task success. A decrease in image speed would increase overlap, but this disadvantage might be more than compensated by the additional time for which a given image is wholly or partially visible. The results discussed in Chap. 3 were derived for the Diagonal mode of RSVP, but would be expected to be generally applicable to other modes. In considering the use of overlap two considerations arise. First, to some extent, obscurement can be ameliorated by the existence of an unobscured capture frame. Second, if at some point in an overlapped sequence of images, each image 'rotates' to be visible in its entirety (as with CoverFlow—see Fig. 1.6) then, again, the overlap present elsewhere in the sequence may be less serious.

6.5.2.3 Multiple Entry/Exit Modes

All the considerations discussed above for moving modes apply to multiple entry/exit modes which, by their nature, offer additional potential for the confirmation of image recognition: there is more than one path that an image can follow before disappearing. This single feature may reduce considerably, or even eliminate, the need for image overlap.

6.5.3 Image Size

"How large should the images be?" is a difficult question to answer, especially if visual appeal is an important factor. Image features such as colour, distinctiveness, layout and previous exposure, as well as the manner in which a reduced size of image (a 'thumbnail') is generated from an original, all affect recognition. A factor to keep in mind is the fact that text, and especially its size, font and colour, constitutes an image: indeed, many products that might be advertised in floating mode, for example, may be instantly recognisable from familiar product name text (e.g., "GODIVA") rather than a product image (e.g., a box of chocolates). Experimental evidence relating recognition success to image *size* for representative images (including some containing meaningful text) is perhaps understandably hard to come by.

In one sense the answer will become obvious during the evaluation of a design, and necessary adjustments will be made before it is finalised. Limited experimental evidence available from the literature (Kaasten et al. 2002; Suh et al. 2003;

Burton et al. 1995) broadly suggests that an image resolution of about 300 × 300 pixels will lead to high confidence of recognition, while a resolution of 100 × 100 might lead to around 60 % correct recognition.

6.5.4 Manual Control

The execution of a task using RSVP, whether that task involves exploration or search or a combination of the two, may demand, for its most efficient execution, a wide range of image visibility durations. In early RSVP applications speed and direction control involved a mouse-click on one of a number of buttons, as illustrated in Fig. 6.3.

However, as Wittenburg et al. (2000) found, a user, on noticing an image of interest, would typically then have to search for and click the stop button, by which time the image would have disappeared. A much improved method of control was to use mouse-over in place of click. With the Floating mode example the speed control (Fig. 6.4) was analogue: placement of the mouse-cursor towards the tips of the arrows increases the speed of presentation and towards the tails decreases the speed. Stopping the presentation is straightforward—the mouse cursor is moved away from the arrows. Mouse control *per se* is not a requirement: Brinded et al. (2011) describe a system in which a rotating knob controls either position or speed.

Evaluations carried out by Wittenburg et al. (2000) also suggested that different speeds might be associated with different tasks, with 'perusal' typically involving

Fig. 6.3 Discrete speed control (after Wittenburg et al. 2000)

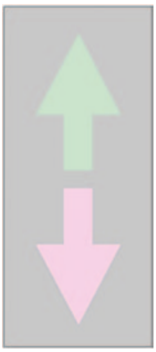

Fig. 6.4 Analogue speed control (after Wittenburg et al. 2000)

a pace of about one image per second, ‘scanning’ involving one to two images per second and fast movement to a known region reaching four images per second. These were approximate observations, but give some idea of speeds that might be made available in an application.

6.5.5 Simultaneous Visibility of Images

The question “How many images should simultaneously be visible on the display” could equivalently, though approximately, be rephrased as “how long should each image remain visible?”

A characteristic feature of RSVP is the opportunity to recognise an image satisfying a specific criterion *pre-attentively*, and therefore very rapidly and without conscious cognitive effort. If for any reason that recognition is in doubt, and needs to be confirmed or rejected, then *attentive* revisitation must take place. Obviously, if infinite time is available for such attentive examination, all doubt can be removed. However, by the very nature of RSVP, the constant arrival of new images and the departure of old ones, some of which may satisfy the criterion, is such that very little time can be devoted to attention as opposed to pre-attention. There is therefore a case to be made for the limited retention of each image somewhere on the display, perhaps for a period lasting between one and two seconds, especially if this retention takes the form of predictable movement along predictable routes as in the Volcano mode (see Fig. 5.21). The previously discussed heatmaps for Volcano and Floating modes placed such revisitation in evidence, indicating that in such circumstances gaze almost always follows an image trajectory.

6.5.6 Summary of Design Considerations

We have identified, above, a number of design considerations; in fact, so many that they may be difficult to keep in mind. What may therefore be helpful to the interaction designer is a ‘global’ view of the major design considerations such as the one sketched in Fig. 6.5. As explained, pace (***p***) and image speed (***s***) are major factors in a design, and it is possible to represent some of the principal “soft” boundaries influencing a design in (***p***, ***s***)-space. Together these boundaries define a central region that, at the very least, and tentatively, may be regarded as a “safe” location for the point (***p***, ***s***) representing two major parameters of a design. Intentionally, numerical values are not assigned to the s-axis because the position of the various soft boundaries will depend significantly upon the RSVP mode being considered. The extent to which the boundaries may be approached or even violated is of course at the discretion of the interaction designer.

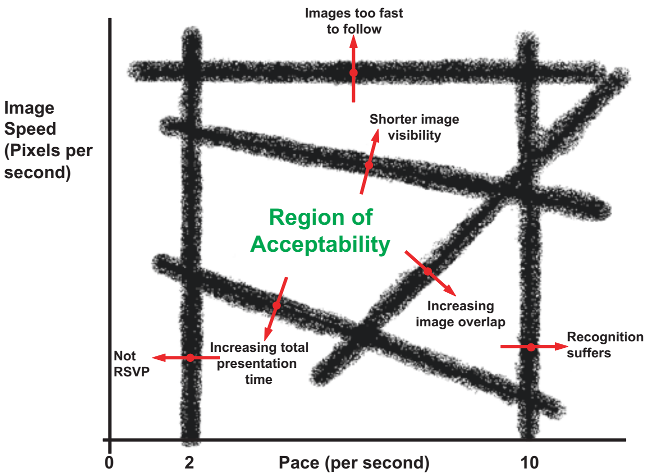

Fig. 6.5 A general summary of limitations affecting the design of an RSVP application

6.6 Evaluation

Evaluation of a design can occur at different stages in the development of an RSVP application.

A first evaluation might well occur at the stage at which a paper prototype has been constructed (e.g., Fig. 6.1f). Using this sketch, critical views can be elicited from professional colleagues, potential users and the client for whom the application is being developed, with the potential for useful design iteration. At this stage the satisfaction or otherwise of criteria regarding likely gaze travel and image speed can begin to be estimated.

When a first—and probably interactive—implementation is available an informal pilot experiment may be carried out to reveal any design weaknesses. Unless an experiment seeking statistical significance is planned, it would appear that 5–6 users would reveal most of the serious problems that need attention.[2] If eye-gaze detection is available this might provide a useful check on gaze behaviour. The many experiments described in this book identify ways in which both objective and subjective performance associated with a 'final' design can be observed.

6.7 Applying RSVP

The considerations that influence the design of an application exploiting RSVP can be numerous. To provide illustrations of this influence we briefly describe six case studies.

[2] See, for instance, http://www.measuringusability.com/blog/five-history.php.

6.7.1 Video Title Selection

Early in the commercial development of RSVP applications, Wittenburg et al. (2000) investigated three generic applications: video title selection; window shopping; and people directories. The applications consequently ranged in content from images *per se* (e.g., family photographs) to images constituting surrogate content (e.g., video) and images acting as adjuncts (as in a product catalogue): they are discussed respectively in this section, in Sect. 6.7.2 and then in Sect. 6.7.3. A major outcome was the design of improved manual speed and direction controls, as discussed earlier in Sect. 6.5.4.

The *SeeHere* interface for entertainment video title selection, already illustrated in simple form in Chap. 1, is illustrated more completely in Fig. 6.6. *SeeHere* is seen in title skimming mode: the user has narrowed down a working set of video titles—seen on the left—for further consideration, after which the dynamic (collage-mode) presentation proceeds at a pace determined by the positioning of a mouse over one of seven control buttons just below the image display. A click on an image or title leads to additional text describing the video or to an option to play a trailer.

6.7.2 Window Shopping

In an effort to introduce some elements of visual sensory engagement and enjoyment into the activity of window shopping Wittenburg et al. (2000) invented the floating mode of RSVP, introduced earlier and repeated in Fig. 6.7. Here the lesson learned about speed and direction control with the *SeeHere* interface led to mouseover

Fig. 6.6 The SeeHere interface for video title selection (courtesy of Kent Wittenburg)

control: placement of the cursor towards the tips of the arrows increases the speed of presentation, and towards the tails decreases the speed. Placement away from the arrows in any direction causes the presentation to stop.

6.7.3 People Finding

The third RSVP application reported by Wittenburg et al. (2000) addressed the problem of browsing an organisational directory, and is illustrated by the sketch of Fig. 6.8. A user can employ it for conventional look-up, but can also rapidly browse the collection of faces. As a consequence it is particularly relevant to a situation in which a user cannot remember a person's name, location or organisation but would recognise their face: it exploits a human user's cognitive ability for rapid facial recognition. The authors point out "Being able to attach faces to names would be a positive step in improving the basic social interactions necessary for developing effective business organisations".

Fig. 6.7 The floating mode of RSVP, supporting window shopping (courtesy of Kent Wittenburg)

Fig. 6.8 An organisational directory (courtesy of Kent Wittenburg)

6.7.4 Video Fast-Forward and Rewind

A task that is repeatedly daily in millions of homes is that of fast-forwarding or rewinding a video or recorded television programme. The reasons are varied: to skip a commercial, to briefly check the weather during a programme currently being watched or to refer back to a particular scene. Fast forwarding and rewind is also undertaken by professional video editors who have the same requirements for an interface that will enable them to perform their task of composing a video production.

In conventional multi-media systems browsing takes place one frame at a time, and has many disadvantages. A commercial application of RSVP to support fast-forwarding and rewind was developed by Dr. Kent Wittenburg (a pioneer of RSVP applications) and his colleagues (Divakaran et al. 2005). To satisfy the needs of users their novel interface designs exhibit one essential characteristic—they preserve *temporal context* by displaying past and future frames at reduced size. One of many interfaces they explored during the development process is illustrated in Fig. 6.9: it is overlaid on top of a conventional full-screen fast-forward/rewind interface.

The so-called 'fisheye' layout in Fig. 6.9 shows the current frame (where play-back will begin when selected) at full size in the centre: on each side, adjacent frames are both reduced in size and overlapped to an increasing degree. Even towards the end of the interface of Fig. 6.9 scene changes are immediately visible, and serve to support navigation. With the contextual information provided at the sides of the current frame consumers are better able to identify upcoming points of interest, scene boundaries and camera movement and, with familiar browsing

Fig. 6.9 Temporal context in video-editing mode (courtesy of Kent Wittenburg)

controls, are more quickly and accurately able to traverse to a desired location in the stream. A variant of the interface of Fig. 6.9 shows (Divakaran et al. 2005) enlarged versions of three images (current; 30 s 'ahead'; and 5 s behind) placed respectively in the centre, the extreme left and the extreme right with, in between, overlapped intervening images of slightly reduced size.

Evaluation of the interface of Fig. 6.9 as well as others embodying temporal context showed that subjects were significantly more accurate using such interfaces than when using a standard VCR interface. On average, for example, subjects resumed playback 25 % closer to an intended position (for example, the first frame after a set of commercials) when using the novel fast-forward scheme than when using the standard fast-forward (Wittenburg et al. 2003b). Of the many variants of interfaces designed to support fast-forward and rewind, and one which also displayed valuable temporal context, the 'V' design illustrated in Fig. 1.16 became a product.

6.7.5 Software Agents

In Sect. 1.4.13 we looked briefly at MAPPA (Witkowski et al. 2003), an e-commerce system incorporating both a manually controlled RSVP to inspect products, and a software agent to provide advice and guidance to users. The MAPPA system was evaluated extensively (Witkowski et al. 2001). Users were guided during a MAPPA session (lasting between 5 and 10 min) and encouraged to use the various features of the system, calling up the agent in both helper and advisor modes and using the manually controlled RSVP product display. The original analysis in Witkowski et al. (2001) was concerned primarily with how users interacted with the agent: here we have added some analysis of user interaction with the manually controlled RSVP based product display.

Only part of the overall time dedicated to product inspection was spent using the manual RSVP control. Figure 6.10 shows gaze behaviour during a single scan of the manually controlled product "shelf" lasting about 14 s in total. It may be seen to start with a single long fixation near the start of the controlling slider bar while the user positions the cursor over the control.[3] The user then draws the control tab from left to right at a somewhat uneven rate causing a corresponding

[3] Successive cursor positions are shown as small red 'x's here.

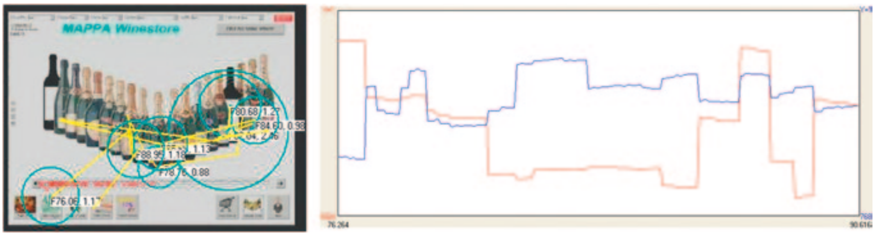

Fig. 6.10 Gaze behaviour during MAPPA RSVP activity

movement of the product display through about half of the 400 available products (i.e. an *average* pace of about 13.7 images/s).

It is seen from the corresponding XY-T plot of Fig. 6.10 that gaze behaviour is more analogous to Slide-show mode—long dwelling fixations tending to a steady gaze strategy (average 1.3 s, longest 2.46 s) at various locations on the product display. It is equally apparent that this user has managed the manual control so as to avoid both the rapid nystagmatic gaze scanning that is associated with Stream and Diagonal mode and the visual pursuit behaviours associated with the slower moving image streams of multiple entry/exit modes such as Volcano and Floating.

6.7.6 Search and Rescue as RSVP

The Search and Rescue study of Mardell et al. (2012), previously described in Sect. 1.4.14, serves as a useful illustration of the effects of changing from classic visual search (wide area, long time) to RSVP (small area, short time). Recall that Mardell et al. segmented a strip of aerial imagery into various degrees of segmentation and presented the segments at rates commensurate with their size to keep the overall time for the complete presentation constant. Table 6.1 summarises the parameters and resulting gaze travel speed results for their experiment. Embedded amongst the images were a small number of hard to spot human “targets”. Examples are shown in Fig. 6.11 for two different segmentation degrees; note that the targets that had to be identified and reported are to the same scale on the display.

Figure 6.12 illustrates the effects on gaze strategy as the presentation rate increases from slow to rapid. Figures (1) show the full size presentation commensurate with a classical saccade and fixation visual search pattern (left, heatmap, with the foveal area circled; right, a single example of a gaze track, with the target circled). As the area is successively reduced and the rate increased (shown in figures 2–6), a transition to a purely static gaze strategy is seen. The corresponding reduction in gaze speed (column 7, Table 6.1), confirms the change.

The fact that the extent of gaze travel decreases as the degree of segmentation increases is not surprising. By contrast, what is very surprising is the observation that target recognition is substantially the same (around 65 % in the example shown)

Table 6.1 Search and Rescue parameters and corresponding gaze speeds for different segmentation degrees (reprinted from Mardell et al. 2012)

Segmentation degree	Number of tiles	Tile visibility time—(ms)	Image size, pixels		Subtended visual angle on display		Average gaze speed pix/s
			Width	Height	Horizontal	Vertical	
1	15	3,878	1,024	768	22.62°	17.06°	1,108
2	60	970	512	384	11.42°	8.58°	739
3	135	431	342	256	7.64°	5.72°	456
4	240	242	256	192	5.72°	4.30°	430
5	375	155	205	154	4.59°	3.45°	394
6	540	108	171	128	3.83°	2.86°	415

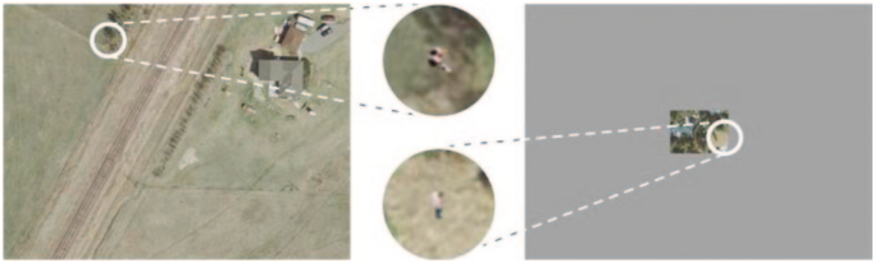

Fig. 6.11 Example targets embedded in the aerial imagery (courtesy of James Mardell)

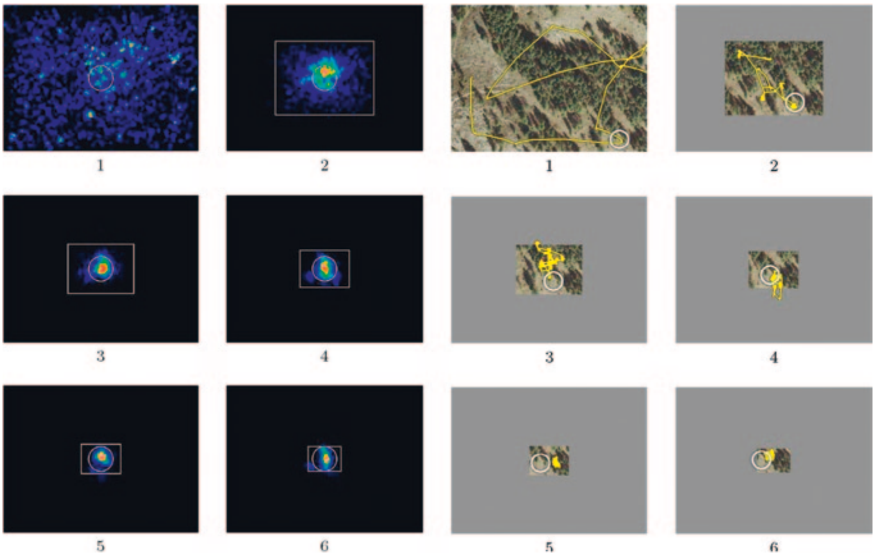

Fig. 6.12 Cumulative heatmaps for different segmentation levels (*left*) , gaze examples (*right*) (courtesy of James Mardell)

independently of the degree of segmentation. This result, together with the evidence shown in Fig. 6.12, suggests that the human visual processing system is remarkably robust in its recognition abilities across a wide range of presentation styles.

6.8 Potential Applications

The many modes and applications of RSVP discussed in this book inevitably lead to ideas concerning other possible fields of application, ideas that might be worth exploring. In this section we draw attention to potential applications that might benefit from further investigation.

6.8.1 Airline Baggage Inspection

The inspection of airline baggage to detect the presence of forbidden articles shares many of the challenges experienced in search and rescue; the consequence of a failure to detect a prohibited item could range from the serious to the catastrophic. Nevertheless the two applications differ in one major respect: with search and rescue the size of the target is so small that segmentation RSVP can safely be deployed, whereas the size of a target (e.g., knife, revolver) within a baggage can be substantial as a proportion to the whole, thereby introducing new challenges to the segmentation approach. Gale et al. (2000) investigated operator performance when identifying simulated Improvised Explosive Devices (IED) under rapid presentation conditions.

In airline baggage inspection the relevant images are mainly obtained by X-ray imaging (Fig. 6.13) and can be presented statically to a trained inspector. Currently, such a static image is examined for a few seconds by one or more trained staff,

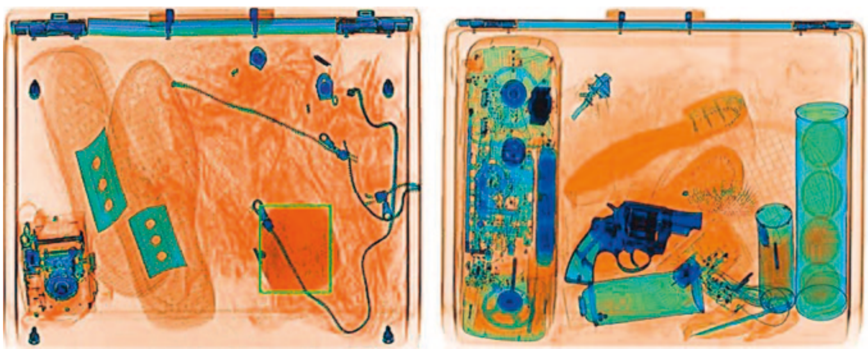

Fig. 6.13 X-ray images of airline baggage (courtesy of Smiths Detection [4])

[4] www.smithsdetection.com

but the potential of presenting the whole or part of such an image in Slide-show mode with a pace as low as 3 per second may be worth exploring. In this context it may be relevant to recall the findings of Kundel and Nodine (1975) who reported that trained radiologists were able to identify anomalies in lung X-rays with 70 % success when those X-rays were visible for only 200 ms each. Clearly, an issue is the use of more than one inspector. One factor in planning an approach to baggage inspection is acknowledgement of the very low prevalence of target occurrence, which has been shown to reduce identification success (Wolfe et al. 2005).

6.8.2 Blood Cells

The clinical investigation associated with many diseases involves the testing of a blood sample. In some situations a slide is prepared (Fig. 6.14a) containing a number of blood cells. It is possible, by using watershed segmentation (Roerdink and Meijster, 2001), to segment the individual cells (Fig. 6.14b) which can then (Fig. 6.14c) be treated as separate images. Clinical considerations may require the cells to be examined as a group in a single image, as in Fig. 6.14a and b. However, in some situations, as Forlines and Balakrishnan (2009) have pointed out, it may be appropriate to present the separate individual cell images to an investigator in Slide-show mode (as in Fig. 6.14c).

6.8.3 Crowd Screening

In certain security situations it may be necessary to examine the many faces within a large crowd to identify the presence or otherwise of specific individuals for whom facial images are already available and characterised by certain facial dimensions. One automatic technique available for this purpose is FaceAlert (www.facealert.com) and has been shown to be very effective. However, if there are circumstances where recognition must be carried out by a human being it may be worth exploring the potential advantage of presenting faces individually, for examination, in Slide-show mode.

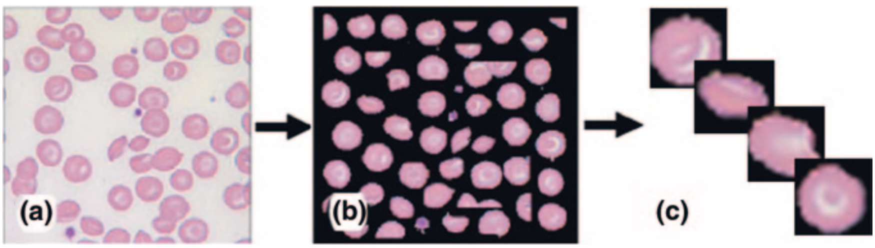

Fig. 6.14 **a** a blood cell slide, **b** the result of watershed segmentation, and **c** individual cells extracted for RSVP presentation (adapted from Forlines and Balakrishnan 2009)

6.9 Concluding Thoughts

From the earliest studies of pre-attentive processing, which established our surprising ability to perform certain tasks rapidly and without conscious cognitive effort, the technique of Rapid Serial Visual Presentation has developed along many lines. Advances in graphical processing, and computational power in general, have provided early innovators with the flexibility to explore a wide variety of modes (Chap. 1) in a range of practical applications. In parallel, advances in our understanding of the human visual processing system (Chap. 2) have helped to underpin those applications. Nevertheless, the range of possible modes (Chap. 3), as well as the manner of their detailed design, has emphasised the dilemma that the successful design of an RSVP application rests heavily upon the creative skill of the interaction designer. However, more recent studies of the eye-gaze behaviour associated with a number of modes (Chap. 4) has established empirical guidelines of value to the interaction designer. In particular, the need to be concerned with saccadic behaviour (Chap. 5) has emerged as a major guideline for the designer. It may not be an unjustifiable claim that the interaction designer contemplating the use of RSVP in an application is now provided with more useful guidance (Chapter 6) than was the case a decade ago.

RSVP is still being developed, even though some underlying empirical evidence dates back many decades. There are still many gaps in our knowledge, each of interest to researchers and designers in different disciplines. With regard to the underlying perceptual and cognitive processes there is still much to be learned, and some of the unexplained experimental results reported here will provide appropriate researchers with food for thought. There are also many unanswered questions in the mind of interaction designers who may consider the application of RSVP but need to be informed of user experience and, especially, acceptance of the many different features of an application exploiting RSVP. It will not be surprising if the creativity of those designers identifies useful new RSVP modes and, *inter alia*, new insights.

References

Brinded, T., Mardell, J., Witkowski, M., & Spence, R. (2011, July). The effects of image speed and overlap on image recognition (pp. 3–11). *Proceedings of 15th International Conference on Information Visualization* (*IV2011*), London.

Burton, C.A., Johnston, L. J., & Sonenberg, E. A. (1995). Case study an empirical investigation of thumbnail recognition (pp. 115–121). *IEEE, Proceedings Conference on Information Visualization* (*INFOVIS' 95*).

Corsato, S., Mosconi, M., & Porta, M. (2008, May). An eye tracking approach to image search activities using rsvp display techniques (pp. 416–420). *ACM, Proceedings of Workshop on Advanced Visual Interfaces*, Naples.

Divakaran, A., Forlines, C., Lanning, T., Shipman, S., & Wittenburg, K. (2005, January). Augmenting fast-forward and rewind for personal digital video recorders (pp. 43–44). *IEEE International Conference on Consumer Electronics* (*ICCE*), Digest of Technical Papers, (IEEE Xplore, TR2004-136).

Few, S. (2009). *Now you see it*. Oakland: Analytics Press.
Forlines, C., & Balakrishnan, R. (2009). Improving visual search with image segmentation (pp. 1093–1102). *ACM Procedings of Computer Human Interaction* (*CHI-09*), Boston, MA, USA.
Gale, G., Mugglestone, M. D., Purdey, K. J., & McClumpha, A. (2000). Is airport baggage inspection just another medical image? *Proceedings of SPIE, 3981*, 184.
Kaasten, S., Greenberg, S., & Edwards, C. (2002). How people recognize previously seen web pages from titles, URLs and thumbnails. In X. Faulkner, J. Finlay & F. Detienne (Eds.) People and computers XVI (pp. 247–265) (*Proceedings of Human Computer Interaction 2002*), BCS Conference series.
Kundel, H. L., & Nodine, C. F. (1975). Interpreting chest radiographs without visual search. *Radiology, 116*, 527–532.
Mardell, J., Witkowski, M., & Spence, R. (2012). An interface for visual inspection based on image segmentation (pp. 697–700). *Proceedings of Working Conference on Advanced Visual Interfaces* (*AVI-12*), Capri Island (Naples), Italy: ACM.
Roerdink, J., & Meijster, A. (2001). The Watershed transform: definitions, algorithms and parallelization strategies. *Fundamenta Informaticae, 41*, 187–228.
Suh, B., Ling, H., Bederson, B.B., & Jacobs, D.W. (2003). Automatic thumbnail cropping and its effectiveness (p. 9). *Proceedings of User Interface and Software Technology Conference* (*UIST-03*).
Witkowski, M., & Spence, R. (2012). Rapid serial visual presentation: An approach to design. *Information Visualization*, *11*(4), 301–318. Published online April 25, 2012, doi:10.1177/1473871612439643.
Witkowski, M., Arafa, Y. & de Bruijn, O. (2001, March). Evaluating user reaction to character agent mediated displays using eye-tracking technology (pp. 79–87). *Proceedings AISB'01 Symposium on Information Agents for Electronic Commerce*.
Witkowski, M., Neville, B., & Pitt, J. (2003). Agent mediated retailing in the connected local community. *Interacting with Computers, 15*, 5–32.
Wittenburg, K., Chiyoda, C., Heinrichs, M., & Lanning, T. (2000, January 26–28). Browsing through rapid-fire imaging: requirements and industry initiatives (pp. 48–56). *Proceedings of Electronic Imaging '2000: Internet Imaging*, San Jose, CA, USA.
Wittenburg, K., Forlines, C., Lanning, T., Esenther, A., Harada, S., & Miyachi, T. (2003b). Rapid serial visual presentation techniques for consumer digital video devices (pp. 115–124). *ACM, Proceedings UIST'03*.
Wolfe, J. M., Horowitz, T. S., & Kenner, N. (2005). Rare items often missed in visual searches. *Nature, 435*, 439–440.