Rapid Serial Visual Presentation: A space-time trade-off in information presentation

Oscar de Bruijn
Imperial College of Science, Technology & Medicine
Department of Electrical and Electronic Engineering
Exhibition Road, London SW7 2BT
+44 2075946330
o.debruijn@ic.ac.uk

Robert Spence
Imperial College of Science, Technology & Medicine
Department of Electrical and Electronic Engineering
Exhibition Road, London SW7 2BT
+44 2075946259
r.spence@ic.ac.uk

ABSTRACT

Rapid Serial Visual Presentation, or RSVP, is the electronic equivalent of riffling a book in order to assess its content. RSVP allows space to be traded for time and has tremendous potential to support electronic information browsing and search particularly on small displays. However, before this potential can be realised, it is necessary to investigate the parameters involved in the successful application of RSVP in the user interface. The rapid display of images or text is well within the capabilities of current desktop computers and even of current or near future mobile devices. The limiting factor in the application of RSVP, therefore, has to be the limited capability of the user's visual system. Users' reading comprehension with RSVP of text has been studied extensively. The transfer of information with RSVP of images, however, has received relatively little attention. This paper examines some of the problems with applying RSVP for image browsing and search.

Keywords

Image browsing, rapid serial visual presentation, information navigation, visualisation, space-time trade-off.

1. INTRODUCTION

You go into a bookstore, and see what might be an interesting book. What do you do? You probably pick it up and riffle its pages. Seeing each page for a very brief time allows you to assess its contents. Riffling a book is just one of many forms of Rapid Serial Visual Presentation, or RSVP. Electronic RSVP is the electronic equivalent of riffling in which chunks of content are displayed rapidly, each for a brief period of time, in order to allow the user to browse or search through the contents. Evaluation of RSVP in video-on-demand browsing systems [5 and 10], a shopping service [29] and dynamic key frame presentation for video browsing [26] suggests that electronic RSVP can be applied successfully within these contexts. Some studies suggest that RSVP may even enhance the information processing capacity of the user relative to simultaneous presentations [14, 15, 16], but even in the worst cases, RSVP has been found to be at least as effective as simultaneous presentations [21, 27; although see 17].

Browsing and searching are important activities within the context of information navigation [e.g., 3, 4, 6 and 25], and RSVP may prove to be an important tool for making electronic information navigation more efficient [24]. Since RSVP is an example of a presentation technique in which space is traded for time, RSVP may be especially useful in situations in which screen real estate is limited. Such is necessarily the case, for example, in mobile devices such as mobile phones, handheld computers, Personal Digital Assistants (PDAs) and wearable computers. It seems likely that the number of applications designed for these devices is going to increase greatly over the next few years. In particular, the increase in services offered over wireless networks is likely to be very significant indeed [e.g., see 20]. There is, therefore, a tremendous potential for the application of RSVP for allowing users to access these services on small display screens. Before the potential of RSVP can be realised, however, it is important to investigate what the consequences of such a space-time trade-off might be for information transfer.

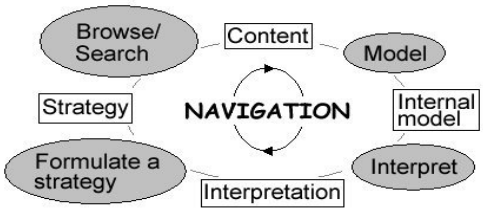

Figure 1. Spence's [25] framework for navigation; the grey ovals represent activities and the boxes represent the outcomes of these activities.

In the present paper, we examine the potential of applying RSVP in the user interface to allow users to navigate information spaces. This examination is structured around a framework for navigation, developed by one of the authors [25], that identifies a number of activities such as strategy formulation, browsing and searching, content modelling, and model interpretation. Traversal of these activities, almost certainly iterative, constitutes the navigational process (see Figure 1). According to this framework of navigation, RSVP should particularly be able to support the activities of browsing, content modelling and searching, as these are the activities whose success relies on the useful display (externalisation) of content. The question we are trying to answer is, therefore, whether, and under which circumstances, users can successfully engage in these activities when content is presented using RSVP. The answer to this question has to be obtained

empirically, and for this purpose we have developed a method for evaluating the effectiveness of RSVP. The results of an experiment investigating the effectiveness of RSVP for image searching are presented towards the end of this paper.

First, we need to find some indication of what the likely factors are that may affect the usability of RSVP for information navigation. When applying RSVP in the user interface, the interface designer has to make a number of choices of presentation parameters, often without knowing the consequences of these choices for the usability of the interface. A systematic examination of all the parameters, task by task, of the effects of overall rate, exposure duration per item, size of presentation, etc., is out of the question because of the dimensionality of parameter space. We therefore have to find a way of bootstrapping our investigations by somehow identifying the parameters that are most likely to affect the usability of RSVP. Since the rapid display of images and text is well within the capabilities of current desktop computer technology and even of current or near future mobile computer technology, the limiting factor in the application of RSVP has, therefore, to be the limited capability of the user's visual information processing system.

Indeed, whenever an application designer opts to present information temporally rather than spatially, it has to be taken into account that the human visual system's temporal resolution is low relative to its spatial resolution; this is a consequence of the way in which the visual world is projected onto the eye's retina. The eye makes constant small movements and frequent large movements (the latter called saccadic eye movements) causing the visual field to be projected onto different areas of the eye's retina after each eye movement. In order to build a stable representation of the visual world it is necessary to integrate all these different retinal images [e.g., 7]. Sometimes the temporal integration of visual information by the visual system can, as in the cinema, be exploited successfully to give the illusion of a moving coherent scene from the rapid visual presentation of individual picture frames when each frame represents a state intermediate between the preceding and following frames (Tse et al. [26] used this argument in favour of using RSVP for key frame video browsing). When there is no such relation between frames in the sequence, as is often the case in RSVP, the illusion of movement will break down. Nevertheless, when unrelated frames are presented in rapid succession, integration will still occur, albeit inappropriately, and the content of individual frames may not be fully resolved as a consequence (interference between sequentially presented items is often referred to as "masking" [e.g., 1]). In the next section we will explore what the consequences of masking may be for information browsing, modelling and searching.

2. BROWSING, CONTENT MODELLING AND SEARCHING

Depending on the number of items to be explored, each item can be displayed only for, say, 100 – 500 ms in order for the whole activity to be completed within an acceptably short length of time. At such fast rates of presentation, masking may seriously limit the amount of information processed for each item in the RSVP stream. However, it has repeatedly been shown that the critical rate at which performance in a task breaks down depends hugely on the nature of the task, as the information requirements may differ between tasks and different tasks may put different demands on the information processing system.

2.1 Searching

We first consider the task of searching. Searching implies that the user has some prior idea about what he/she is searching for. A typical situation demanding searching activity is one in which the user has formulated a goal and asks, "Is it here?" Thus, someone may be looking for a particular image, but cannot remember where it was stored. In this situation, the content is viewed until a match is found. When users are searching for a predefined target item, the visual system may employ attentional selection of the target item, allowing the target item to be processed more fully than the other items in the RSVP stream [22 and 23]. As a consequence of attentional selection of the target item, identification of the target item may occur at frame times too rapid for all presented items to be fully processed [2, 11 and 28].

The success of this attentional selection of target items depends on how specific a goal the user has formulated. For example, the user might be looking for an image of which the appearance is known. Or the user may be searching for images of which s/he only has a description by name (e.g., "a picture of a dog"). In the first case, attentional selection may take place based on a visual cue, such as a distinctive visual feature present in the target image, whereas, in the latter case, selection may take place based on a categorical cue. Although attentional selection can take place using visual or categorical cues, and even using negative categorical cues (e.g., "an image that does *not* belong to the category *transportation*"), it appears that selection based on visual cues can occur at higher presentation rates than selection based on categorical cues [e.g., 9 and 18]. However, whether higher presentation rates with visual cues can also be achieved in an applied context of information searching will need to be confirmed by future investigations.

Things become much more complicated, however, when the user is looking for more than one image, as there is a limit to the number of images that can be reliably reported after RSVP [2 and 28]. In addition, the identification of one target may interfere with the identification of subsequent targets. This phenomenon is referred to as the *attentional blink* and lasts several hundred milliseconds after presentation of the first target [22]. Thus, any target items presented within this interval subsequent to the first target may not be processed up to a level that allows these items to be reported. Obviously, this phenomenon may affect search performance in many ways. For example, in cases in which the user expects the target item to appear closely following an item that is used as a landmark, identification of the landmark item may interfere with detection of the target item. Again, whether or not this could happen will have to be verified by future research.

2.2 Browsing and Content Modelling

Next we try to establish the research questions concerning the usability of RSVP for browsing and content modelling. Browsing is defined as the registration of content. A typical situation demanding browsing activity is one in which the user asks, "What's there?" Thus, someone who has filed a number of items (e.g., image files) may come across a previously composed file whose appearance conveys little about its content, and may wish to explore that content. In this type of situation, the user is not searching for a particular item, but wishes to browse in order to develop an overview of the contents of the file (i.e., a model of the contents). The model of the contents can subsequently be interpreted and this may lead to the formulation of a new strategy

for further browsing or searching. Thus, for RSVP to support browsing it is necessary that the user be able to process and store enough information about each item to build a model of the contents.

An indication of the suitability of RSVP for content modelling may be gleaned from the ability of users to memorise information presented as part of an RSVP stream. The ability to memorise items presented in RSVP streams has been tested mostly using a recognition memory task in which participants are tested on their ability to separate previously presented items from items that have not been presented. Stimulus factors such as presentation rate and visual similarity of stimuli have all been shown to affect recognition memory for items in the RSVP stream [19]. Indeed, recognition memory suffers dramatically when the presentation duration for each frame is decreased. For example, Intraub [8] showed that recognition memory for sequentially presented pictures decreased from very good when the time between stimulus onsets (Stimulus Onset Asynchrony, SOA) was 5 seconds to very poor when SOA was only 110 milliseconds (Figure 2). This would suggest that RSVP does not support the modelling of content at rates (i.e., approximately 10 items per second) fast enough to allow a substantial number of items to be browsed.

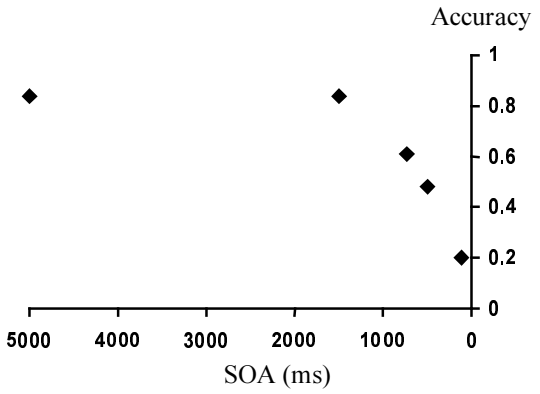

Figure 2. Accuracy in a recognition memory task with different lengths of time between stimulus onsets (SOA) obtained by Intraub [8].

There are, however, a number of indications that more information about individual items in an RSVP stream is being processed than is apparent from performance in a recognition memory task. In the recognition test, Intraub [8] also found that the ability to detect that images were mirror reversed in the recognition test decreased more rapidly than simple recognition memory, suggesting that enough information was stored to allow correct recognition to occur even though this information was not sufficient to distinguish between the original item and a mirror reversed version. Similarly, it may be the case that sufficient information about items in an RSVP stream is available to develop a model of the content even though individual items may not be recognised in a recognition memory test. Indeed, when reading passages of text, participants were often able to correctly describe the gist of a passage without necessarily being able to recall specific words [13]. Moreover, tests of implicit perception have shown that often more information about a briefly presented visual stimulus is available than can be reported by the observer [e.g., 12]. This suggests that it may be similarly possible to develop a model that describes a set of pictures in terms of subject matter even when these pictures are presented at RSVP rates that do not allow the observer to recall or recognise each individual picture subsequently. This suggestion is supported by Tse et al. [26], who found that observers were able to determine the gist of videos after observing an RSVP display of key frames at rates of up to 16 frames per second while, at that rate of presentation, their recall of objects in the frames was very poor indeed. However, it is important to verify these results of Masson and Tse et al., after which we need to determine whether models of the contents obtained from RSVP are sufficient to support further browsing or searching through the formulation of a browsing or searching strategy. This will be a focus point in our future investigations into the applicability of RSVP.

3. CONCLUSIONS

Rapid Serial Visual Presentation appears to offer a valuable technique to support information browsing and search. In applications such as PDAs, mobile devices and wearables, where display area is limited, it is especially valuable in view of the space-time trade-off it offers. Its use is limited by the human visual information processing system, so many issues, some of which are discussed in this paper, remain to be understood and resolved before the technique can be widely and robustly employed in practical applications. Simple experiments have already identified unexpected aspects of RSVP; more studies are needed both to characterise the RSVP phenomenon and to reconcile it with current knowledge regarding human perception and cognition.

4. ACKNOWLEDGEMENTS

This work was carried out as part of our contribution to the Living Memory Project funded under the European Network for Intelligent Information Interfaces (I^3).

5. REFERENCES

[1] Breitmeyer, B.G., & Ganz, L. Implications of sustained and transient channels for theories of visual masking, saccadic suppression, and information processing. Psychological Review, 83 (1976), 1-36.

[2] Broadbent, D.E., & Broadbent, M. H. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. Perception & Psychophysics, 42,105-113.

[3] Carmel, E. Crawford, S., & Chen, H. (1992). Browsing in hypertext: a cognitive study. IEEE Transactions on Systems, Man and Cybernetics, 22, 5, 865-884.

[4] Darken, R.P. (1997). A simple model of navigation in large-scale virtual spaces. ACM SIGCHI Bulletin, 29, 4, 44-49.

[5] Elliot, E. & Davenport, G. (1994). Video Streamer. CHI'94 Companion Proceedings, ACM, 65-66.

[6] Furnas, G.W. (1997). Effective view navigation. In CHI'97: Human Factors in Computing Systems (Atlanta, GA, USA, March 1997), ACM Press, 367-374.

[7] Henderson, J.M. (1994). Two representational systems in dynamic visual identification. Journal of Experimental Psychology: General, 123, 4, 410-426.

[8] Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. Journal of Experimental Psychology: Human Learning and Memory, 6, 1, 1-12.

[9] Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. Journal of Experimental Psychology: Human Perception and Performance, 7, 3, 604-610.

[10] Lam, K. & Spence, R. (1997). Image browsing: A space-time trade-off. Human-Computer Interaction: INTERACT'97, Chapman & Hall, 611-612.

[11] Lawrence, D.H. (1971). Two studies of visual search for word targets with controlled rates of presentation. Perception & Psychophysics, 35, 558-564.

[12] Marcel, T. (1983). Conscious and preconscious recognition of polysemous words: Locating the selective effects of prior verbal context. In R.S. Nickerson (Ed.), Attention and performance VIII. Hillsdale, N.J.: LEA.

[13] Masson, M.E.J. (1983). Conceptual processing of text during skimming and rapid sequential reading. Memory & Cognition, 11, 262-274.

[14] Matin, E., & Boff, K.R. (1988). Information transfer rate with serial and simultaneous visual display formats. Human Factors, 30, 2, 171-180.

[15] Matin, E., Boff, K.R., & Donovan, R. (1987). Raising control-display efficiency with rapid communication display technology. Proceedings of the Human Factors Society 31st Annual Meeting. Santa Monica, CA: The Human Factors Society, 258-262.

[16] Matin, E., Shao, K.C., & Boff, K.R. (1993). Saccadic overhead – Information-processing time with and without saccades. Perception & Psychophysics, 53, 4, 372-380.

[17] Payne, D.G., & Lang, V.A. (1991). Visual monitoring with spatially versus temporally distributed displays. Human Factors, 33, 4, 443-458.

[18] Potter, M.C. (1975). Meaning in visual search. Science, 187, 965-966.

[19] Potter, M.C., & Levy, E.I. (1969). Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology, 81, 10-15.

[20] Press, L. (1999). The post-PC era. Communications of the ACM, 42, 10, 21-24.

[21] Rahman, T., & Muter, P. (1999). Designing an interface to optimise reading with small display windows. Human Factors, 41, 1, 106-117.

[22] Raymond, J.E., Shapiro, K.L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? Journal of Experimental Psychology: Human Perception and Performance, 18, 849-860.

[23] Shapiro, K.L., & Raymond, J.E. (1994). Temporal allocation of visual attention: Inhibition or interference? In D. Dagenbach & T.H. Carr (Eds.), Inhibitory mechanisms in attention, memory and language. Boston: Academic Press, 151-188.

[24] Spence, R. (1998). A Content Explorer. Information Engineering Section Report 98/08, Imperial College, London, May 1998.

[25] Spence, R. (1999). A framework for navigation. International Journal of Human-Computer Studies, 51, 5, 919-945.

[26] Tse, T., Marchionini, G., Ding, W., Slaughter, L., & Komlodi, A. (1998). Dynamic key frame presentation techniques for augmented video browsing. Proceedings of Advanced Visual Interfaces, AVI'98 (L'Aquila, Italy, May 1998), ACM, 185-194.

[27] Uhlarik, J., Joseph, K.M., & Renfro, M.D. (1991). Effects of concurrent information load on serial and simultaneous visual display formats. Proceedings of the Human Factors Society 35th Annual Meeting. Santa Monica, CA: The Human Factors Society, 1527-1531.

[28] Weichselgartner, E., & Sperling, G. (1987). Dynamics of automatic and controlled visual attention. Science, 238, 778-780.

[29] Wittenburg, K., Ali-Ahmed, W., LaLiberte, D., & Lanning, T. (1998). Rapid-fire image previews for information navigation. Proceedings of Advanced Visual Interfaces AVI'98 (L'Aquila, Italy, May 1998), ACM, 76-82.