

A Comparison of Static and Moving Presentation Modes for Image Collections

Katy Cooper¹

Oscar de Bruijn²

Robert Spence³

Mark Witkowski⁴

^{1,3}Dept. Electrical and Electronic Eng.
Imperial College
Exhibition Road, London SW7 2BT
+44(0)20 7594 6259
r.spence@imperial.ac.uk

²Centre for HCI Design
School of Informatics
University of Manchester, UK
+44 (0)161 2005829

O.De-bruijn@manchester.ac.uk

⁴Department of Computing
Imperial College
180 Queen's Gate, London SW7 2AZ
+44(0)20 7594 8423

m.witkowski@imperial.ac.uk

ABSTRACT

In both professional and personal contexts, a common activity is the search for a target image among a collection of images. The presentation of that collection to a user can assume a wide variety of forms, and it would help interaction designers to be aware of the comparative properties of available presentation modes. A property of major interest is the percentage of correct identification of the presence or absence of the target image within the collection; another is users' acceptance of a presentation mode. Several modes of Rapid Serial Visual Presentation (RSVP) are compared for effectiveness in a number of image identification tasks, and with regard to user acceptance and stated preference.

Presentation modes have been classified as static or moving. For a selected representative group of three static and three moving modes, for three image presentation times and for three tasks of increasing complexity, we report experimental results which in most cases establish, with a high degree of statistical confidence, that – over the range of independent variables investigated - (a) static modes are more successful with regard to identification success than moving modes; (b) static modes are far more preferred than moving ones; (c) identification success generally increases with increase in presentation time per image; (d) for mixed and tile modes, identification success is relatively insensitive to image presentation time; and (e) success rate decreases with increase in task complexity except, notably, for slide-show and mixed modes. Evidence from eye-gaze records suggests that the eye-gaze strategy adopted by a subject exerts a very strong influence on both identification success and mode preference. Conclusions are drawn about guidance that can be offered to an interaction designer.

Categories and Subject Descriptors

H.5.2 User Interfaces, user-centered design

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AVI'06, May 23–26, 2006, Venezia, Italy.

Copyright 2006 ACM 1-59393-353-0/06/0005...\$5.00.

General Terms

Human Factors.

Keywords

Rapid Serial Visual Presentation (RSVP). User preference. Eye-gaze tracking

1. INTRODUCTION

A surprisingly large number of personal and professional tasks involve the identification, and often also the interpretation, of many images¹ within a collection. Especially since many of the tasks are undertaken instinctively and often regarded as commonplace, it is useful first to identify some examples in order to demonstrate the wide relevance of the study we report.

An online search for a birthday present, especially if its precise nature is initially unknown, can often involve the rapid (e.g., 10 per second) presentation/examination of available products, whether the presentation is achieved manually by riffling the pages of a catalogue, or by digital means. The search may instead be for a TV channel offering attractive viewing, and be facilitated by the rapid and concurrent presentation of frames from available channels. An entirely different scenario is one in which a person participating in a meeting must locate, on their laptop and as quickly as possible, a particular page being discussed (“round about the middle, with the greenish diagram bottom left”): the document will typically be rapidly paged. The selection of a photograph to show to a friend may involve the rapid ‘riffling’ of either a physical or digital collection of images. Similarly, a rapid sequential view of news items on a PDA can lead to the choice of a news channel of interest.

All these and many other applications are concerned with a collection of images sufficiently extensive that the problem of presenting those images to a user in an acceptably short time, and with an acceptable likelihood of identifying a target image, is a challenging one.

¹ The widespread nature of image identification and interpretation is a consequence of the meaning we attach to ‘image’; it need not be a conventional image such as a Picasso and can, for example, be an icon, a pattern or a page layout.

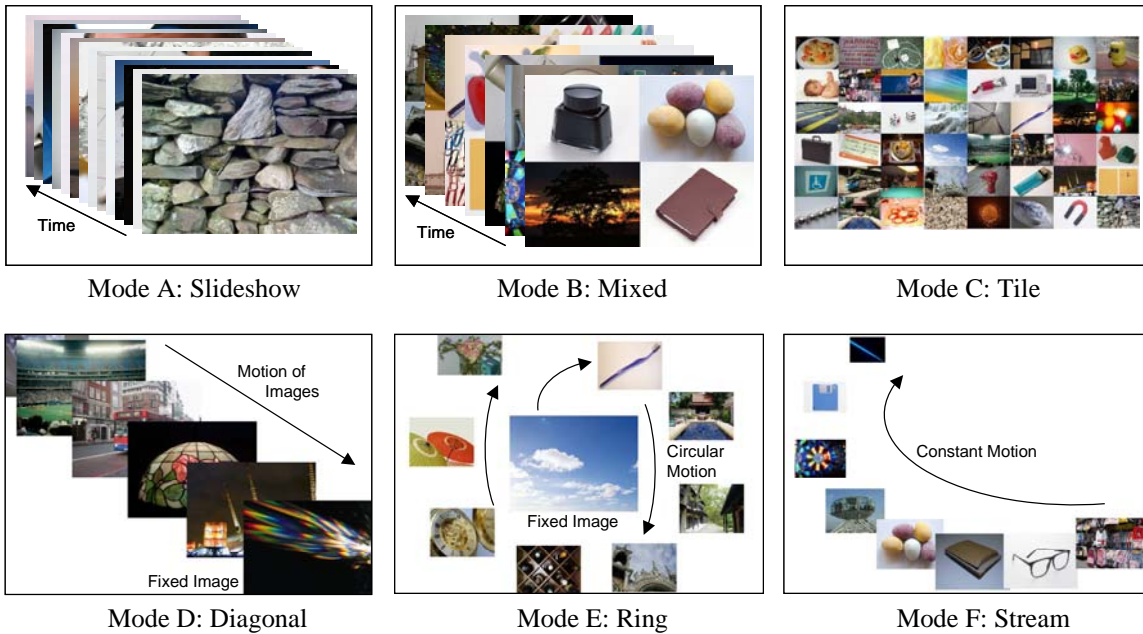


Figure 1: The Six RSVP Presentation Modes Used in the Experiments

2. IMAGE PRESENTATION MODES

About ten years ago considerable interest was first directed to the application potential of a presentation mode generally referred to as *Rapid Serial Visual Presentation (RSVP)* in which a collection of images is presented sequentially (Figure 1, Mode A) by showing each one separately, and in the same location, for a short period of time, typically 100 milliseconds (Spence, [14]). Its viability is subjectively confirmed by the familiar activity of quickly riffling the pages of a new book in order to gain some idea of its content, and is supported by many experimental investigations (see, for example, Coltheart [2]).

Many variants of RSVP have since been invented. As well as the modes shown in Figure 1, Wittenburg *et al.* ([18], [19]) described the ‘floating’ mode (Figure 2, left) in which images appear to start ‘a long way away’ and then ‘move towards’ the viewer, in a manner similar to that in which motorway signs appear to move towards you when driving.



Figure 2: Floating (left) and Collage (right) Mode RSVP

They also presented the ‘collage’ mode (Figure 2, right), based on the metaphor of images being thrown one after the other onto a table. Earlier, Spence ([13]) proposed a ‘carousel’ mode (Figure 3) in which images emerged from one side of a folder, traversed a circular path and then re-entered the folder.

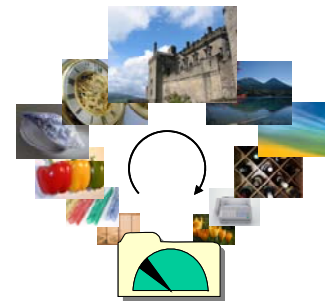


Figure 3: Carousel Mode RSVP

Manually controlled RSVP, combined with the bifocal concept to handle a large collection of video posters (Figure 4), was reported by Lam and Spence ([7]).



Figure 4: Manually Controlled RSVP

Also incorporating manual control of the display, Witkowski *et al.* ([17]) describe a kiosk based sales scenario in which potential customers could rapidly scroll through a large (~450) database of product images (e.g. wine bottles) as though presented on a shelf. The RSVP display was further augmented with search facilities to refine the contents of the shelf display, and an animated software ‘agent’ could be invoked to describe the product or provide assistance or ‘expertise’ (figure 5).



Figure 5: Agent Augmented “Shelf” Mode RSVP

With potential application to TV viewing, Wittenburg *et al.* ([20]) proposed the ‘diagonal’ or ‘timetunnel’ mode (Figure 1, mode D). The potential of RSVP to ameliorate the restriction of small display area led de Bruijn and Tong ([4]) to explore its use to allow news channels to be browsed rapidly on a mobile. A very different application of RSVP was investigated by Tse *et al.* ([15]) and Komlodi and Marchionini ([6]). The task addressed was that of discovering whether you want to watch a particular video or film, but without the need to start watching it at the normal speed of presentation. Tse and his colleagues arranged for the potential film viewer to see a rapidly presented sequence of ‘key frames’ so chosen that some gist of the film can thereby be gained. It was found that successful interpretation of gist could be maintained up to a key-frame presentation rate of 10 per second.

When listing the many presentation modes triggered by the RSVP concept, one involving no sequential behaviour at all, and presenting all images concurrently as in the ‘tile’ mode (Figure 1, mode C), must be borne in mind, at the very least for purposes of comparison.

Much earlier than most of the references just cited, cognitive psychologists conducted many experiments to elicit details of human behaviour under the condition of brief visual stimuli. They are summarised comprehensively by Coltheart ([1]).

We have briefly mentioned some of the many available methods by which a collection of images can be presented to a user to facilitate the identification of a target image. The main questions requiring answers are, “Which is the best?”, “Why is it the best?”, “For what tasks?” and “What do we actually *mean* by best?”.

3. BASIS FOR COMPARISON

To place reasonable bounds on our investigation we considered the single but common task of deciding whether a specified target image is contained within a viewed collection. The number of available presentation modes necessarily requires a realistic basis for their comparison. Recently we proposed (Fawcett *et al.*, [5]) a *resource* base in which one resource is available display area and the other is the total time available during which the presentation of the collection has to be completed. Consideration of typical scenarios strongly suggested that these two features were the principal – and challenging – constraints facing the interaction designer. These two limitations are represented by the ‘resource box’ of Figure 6 within which the collection of images must be located.

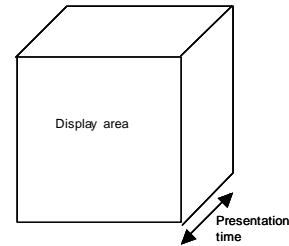


Figure 6: The Resource Box

The way in which the images are positioned in the resource box will influence the effectiveness of each mode. Effectiveness is, unsurprisingly, not a scalar metric. Whereas one principal property of a presentation mode is the efficiency with which it can support a task, another important measure is the degree to which a mode can be considered acceptable by the user. In the experiments to be described these two measures are investigated for a selection of presentation modes.

4. PREVIOUS WORK

4.1 Eye-gaze patterns

Early experimental investigations of RSVP-like presentations of images (de Bruijn & Spence, [3]) were motivated by a desire to see what role was played by eye-gaze, especially if the images constituting a collection did not remain fixed in the display area. The presentation modes examined were the floating, collage (Figure 2), carousel (Figure 3) and diagonal (Figure 1, mode D) modes. The task given to subjects was merely to look out for a previously seen target image and to say afterwards if they had seen that image (it was always present). The presentation modes differed in the way image position and image size changed during the presentation. The subjects’ eye-gaze behaviour was recorded.

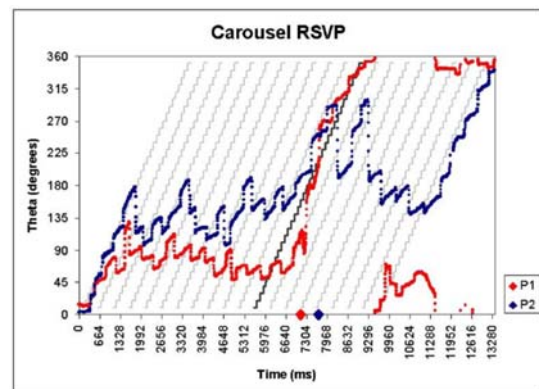


Figure 7 Eye-gaze Behaviour for Carousel Mode RSVP

Appropriate choice of a basis for the representation of eye-gaze patterns disclosed interesting behaviour. Figure 7 provides an example. Here, the angular position of both image and eye-gaze (for two subjects) along the circular path associated with the carousel mode is shown as a function of time. Tracking gaze on one image for around 1500 ms is followed by saccadic transition to a new image, which in turn is tracked for a similar time. It was concluded that eye-gaze tracking can provide a useful tool in the investigation of design issues underlying the application of RSVP. In particular the investigation revealed that participants were able to fixate and track all images until the target was detected. This

suggested that eye-gaze would not be a limiting factor in the effectiveness of these presentation modes.

The investigation helped to identify key questions, the answers to which could provide understanding relevant to the eventual application of image presentation modes. For example, if image presentation times took on different values, would the conclusions regarding eye-gaze still be valid? Moreover, could image presentation times be so short that eye-gaze limitations affect target recognition? Another question is – is there any relation between eye-gaze behaviour and preference for a particular mode? Furthermore, would the results generalise if we had more participants?

4.2 Static presentations and the role of eye-gaze

These and other questions stimulated a subsequent investigation (Fawcett *et al.* [5]). Here, static rather than moving presentation modes were investigated. In a static mode no movement of an image within the display area is involved, though images may be presented sequentially. The three modes investigated are shown in Figure 1. One is the slideshow mode (figure 1A, originally named the ‘keyhole mode’). Another is the ‘mixed’ mode in which each image appears for four times as long as in the slideshow mode but with each image one-quarter the size (1B). The third mode is the ‘tile’ mode (originally called the static mode) in which 64 images are presented concurrently, each 1/64th of the original size. The display area, and the total presentation time for the whole collection, constituted the resource base that was systematically explored. Three display areas were investigated, roughly approximating to mobile, PDA and monitor screen sizes. Four total presentation times were investigated, so chosen as to approximately identify the boundaries of successful recognition of the target image (1.6 to 6.4 seconds for 64 images). Preference, recognition accuracy and eye-gaze behaviour were recorded for 30 participants who were first shown a target image and then, after the presentation, said whether they thought the target image was present or absent. They were also asked to rate the difficulty they encountered.

For all three modes investigated it was found that, roughly, recognition error began to be significant when the total collection presentation time divided by the number of images in the collection fell substantially below 100 milliseconds. It was found that the mixed mode was not only characterised generally by lower recognition error rates, but also by a high preference among users. It was suggested that this result could be explained by eye-gaze behaviour. Following further investigation this proved to be the case: the radial dispersion of eye-gaze from the centre of the display area was similar for the slide-show and mixed modes, and substantially less than for the tile mode. For none of these modes was the dispersion of eye-gaze substantially affected by total presentation time over the range explored.

5. NEW RESEARCH

In the results we now report, and in contrast to our earlier focus on static modes, the investigation has been broadened to include and compare both static and moving modes. Indeed, to this end we hypothesised that:

(1) static modes are associated with a higher target recognition rate than are moving modes; and

(2) static modes are judged more acceptable than moving modes.

The six presentation modes investigated included the three static modes and the three moving modes shown in Figure 1A-F.

Although a common task in the real world is that of deciding whether an image collection contains a target defined by a previously viewed image, people may also want to search for an image based on a verbal description of either its exact or general content. We therefore included in our study three task conditions of expected increasing complexity. In *Task 1*, a participant was shown the precise target image prior to the presentation. In *Task 2* the unique target image is described in detail (e.g., “a cat”), while in a *Task 3*, the unique target image was described in general terms (e.g., “an animal”).

Three total collection presentation times (representing an average per image time ranging between 70 and 130 ms) were chosen to avoid ceiling and floor effects: in other words, high enough to ensure a satisfactorily high recognition rate at the slowest speed of presentation but allowing for the most difficult task results to be significantly more in error. In view of the fact that the recording of eye-gaze during image presentations had earlier provided useful insight into the role of eye-gaze behaviour we deemed it important to record this behaviour throughout the experimental work. The next section describes the experimental procedure followed.

6. EXPERIMENTAL DESIGN

This section describes the experimental procedure established to test each of the six presentation modes (figure 1), at each of three selected presentation rates (equivalent to 3.36, 4.8 and 6.24 seconds for the complete presentation of all 48 images) under three different target conditions (literal image, image description and image concept). Each of these 54 (6x3x3) permutations were also presented with the target image present and absent from the displayed set of images (108 options).

It was determined that the time taken to present all these options to an individual participant (>30 minutes) would be unacceptable. Consequently a set of four sequences was devised, each presenting 27 of the options to any participant. Statistical methods would then be applied to provide complete coverage of the test space; 40 participants providing 10 complete data sets. Each participant was shown three groups of nine presentation sequences. Each of these three groups corresponded to one of the three target tasks described earlier. The presentation mode, rate and target presence were selected such that each individual saw nine examples of each task type, nine at each speed and either 13 or 14 still and moving presentations, broadly equal for target present or absent. This division was made in order to maintain a fair and even split across all the groups.

In the literal task group the participant was shown a target image for two seconds, followed by a blank screen for three seconds, then the presentation sequence. In image description mode, a target was described in a direct manner (i.e. “a pair of trousers”) and indirectly in concept mode (“an item of clothing”). At the end of each image presentation sequence the participant was asked to report “present” when they had seen the target and “absent” if they had not. Participants were asked not to guess, but to report “pass” if they were unsure.

Images were drawn from a corpus of 200 photographs, each sequence of 48 images for the presentation modes selected at random (using publicly available generators at www.random.org) to minimise sequence learning by the participant. Examples of the images used may be seen in figure 1. Care was taken to ensure that the target image was not repeated for any individual participant (Raymond [10]). Care was also taken to ensure that the terms and descriptions used in the image description and image concept test were concise, clear and unambiguous. To this end each proposed description and image set was reviewed by ten individuals, not part of the test, to identify any potential problems, which were rectified and re-checked before the image sequence was finalised.

In each case the complete sequence was prepared as a video clip, to ensure accuracy and repeatability across each of the presentations to individual participants. Each clip lasted 450 seconds, including necessary calibration sequences, image presentation sequences, literal image presentations and sufficient inter-sequence time for the experimenter to describe the target image (for the image description and image concept modes) as well as record the responses made by the participant. The overall time to introduce the participant to the experiment, explain the procedure and show a brief example sequence so that the participant was fully prepared for the tasks requested of them and conduct the experiment was in the order of 15 minutes. Forty individuals drawn from the general student population (Electrical Engineering) were invited to participate.

Flash Macromedia was used to construct the individual image presentation (3.36, 4.8 and 6.24 second) clips and then Adobe Premier to assemble these clips and inter-sequence material into a single video stream sequence (450 seconds). Macromedia Flash provides the facilities required to accurately define the trajectories of images in the moving modes, allowing the diagonal/“Timetunnel” [20] mode to be emulated (figure 1D), and the new moving modes previously described (figure 1E-F) to be created. Adobe Premier was then used to create AVI encoded clips with sufficient compression to allow reliable playback without discernable loss of image quality. A video frame rate of 30 fps was selected to synchronise with the monitor display rate.

For the duration of each of the video sequence clips precise eye-gaze movements on the screen area were recorded using an LC Technologies eye tracking system (www.eyegaze.com). After suitable calibration, specific to the individual, this equipment returns the x and y screen coordinates of the screen pixel to which the participant’s gaze is directed (typically $\pm 2.5\text{mm}$). The equipment uses a modified CCD camera mounted directly below the screen to identify the relative positions of the retinal and corneal reflections of an (infra-red) light mounted co-axially with the optic axis of the camera. The angle of gaze is computed directly from the disparity between these two positions. A new reading is made 60 times a second and recorded in a data file for subsequent analysis. From this data, saccades (rapid ballistic eye movements) can be separated from fixations (periods of gaze stability). It is generally accepted that perception only occurs during fixations, and while the visual field is stable. Also of interest in the context of the current experiments are periods when the eye tracks a moving image across the screen. The equipment also records pupil size and blink rate. The analysis software used allows the total gaze path for any portion of the sequence to be identified and displayed, and a range of data extraction functions

to be performed, including, in this instance, the total effective screen distance of gaze movement during a presentation sequence (figures 15 - 18).

7. EXPERIMENTAL RESULTS

7.1 Efficiency of target identification

Figure 8 shows the target identification success rate achieved with all task types across all presentation modes. The overall proportion of trials in which participants responded correctly was 0.73 (SD = 0.27), which is approximately half way between perfect performance and purely guessing. This means that performance levels are unlikely to be artificially curtailed by floor- or ceiling-effects.

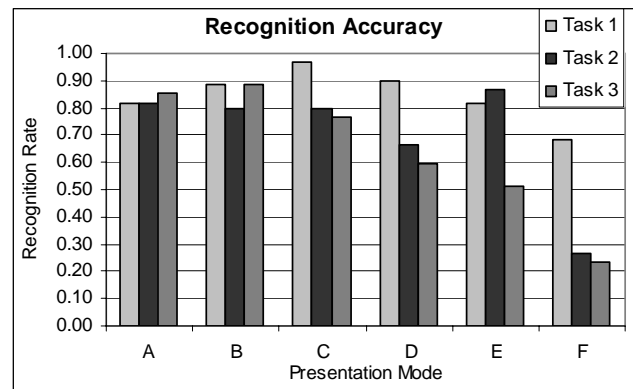


Figure 8: Effect of Mode and Task on Recognition Accuracy

Figure 9 summarises the performance partitioned between the static and moving modes, averaged over all presentation times and all tasks. The average success was 84% for static modes and 62% for moving modes. A Wilcoxon Signed Ranks Test for 27 matched pairs indicated that the number of successes was reliably higher in the static modes compared to the moving modes ($Z = 3.06, p < .01$). The bar chart of Figure 10 shows that every static mode performs better than every moving mode, again averaged over all presentation times and tasks.

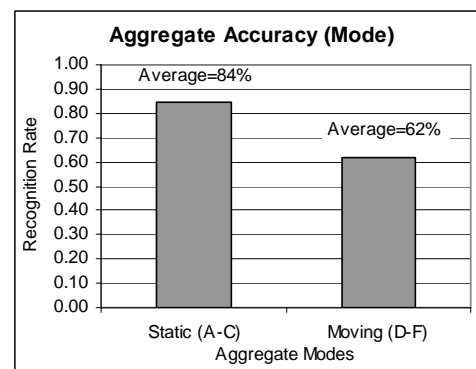


Figure 9: Aggregate Accuracy by Static and Moving Mode

Figure 11 shows the effect of presentation time on success for different presentation modes, in which each ‘bar’ represents 60 (3 tasks * 20 participants) experimental observations. Friedman’s three-way analysis of variance revealed that, over the ranges of independent variables investigated, the slideshow mode was very sensitive to presentation time ($\chi^2 = 25.4, p < .001$), as were the

three moving modes ($\chi^2 = 13.9, 35.9$ and 11.9 , all $p < .01$). However, mixed and tile modes were quite insensitive ($\chi^2 = 2.8$ and 1.0 , neither significant). Figure 12 summarises the effect of presentation rate across all factors.

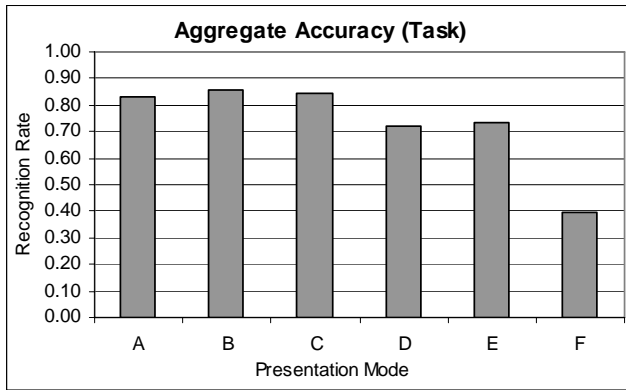


Figure 10: Aggregate Accuracy by Task

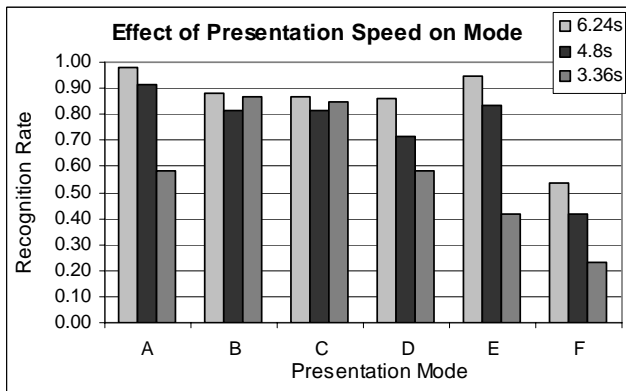


Figure 11: Effect of Presentation Speed on Mode

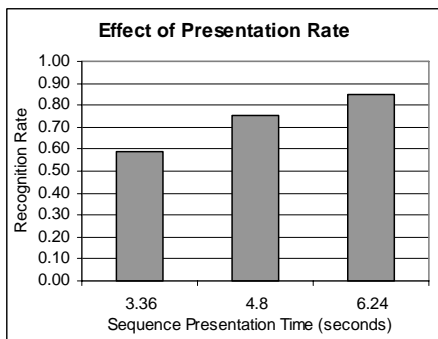


Figure 12: Effect of Presentation Rate

Figure 13 shows the effect of task on identification success. In Task 1 the subject is shown the target image; in Task 2 the target image is describe verbally and specifically (e.g., “a cat”) while in Task 3 the description is also verbal but general (e.g., “an animal”). The assumption that these tasks differ in some form of measurable cognitive complexity appears to be borne out by a decrease in success rate as we move from Task 1 via Task 2 to Task 3, averaged over presentation times and modes.

An obvious question is “Does the advantage of the static over moving modes of presentation generalize across different levels of

cognitive complexity?” The results presented in figure 13 appear to indicate that this is indeed the case, although the difference is perhaps less pronounced for Task 1 in which the participants have to detect a previously seen target image. In fact, it appears that when the cognitive complexity of the task increases in Tasks 2 and 3, the chance that the target was successfully identified decreases rapidly to chance levels in the moving modes but remains stable on the static modes.

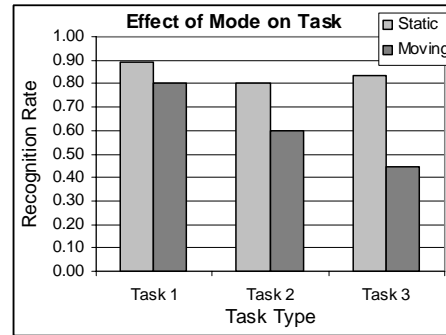


Figure 13: Effect of Mode on Task Success

This seemingly differential effect of task condition on success rate for static and moving images was confirmed by a significant Friedman two-way analysis of variance across task conditions for the moving ($\chi^2 = 7.1, p < .05$) but not for the static mode ($\chi^2 = 1.6, p > .05$). This may indicate that the level of detail in which the images are perceived in the static modes is much greater than in the moving modes, and sufficient for matching the target image with its abstract description. In the moving modes, the amount of detail perceived would be enough to identify a previously seen target image, perhaps on the basis of some gross visual features such as colours and large shapes, but not enough to identify a target image that was merely described in abstract terms.

7.2 User preference

Any use of a presentation mode in a publicly available application will almost certainly take account of user preferences. Justification for the proposed taxonomy (‘static’ and ‘moving’) was clearly provided by the evidence shown in Figure 14: mixed and tile modes accounted for over 50% of the preferred modes, and only 25% of subjects preferred a moving mode (though none preferred the stream mode).

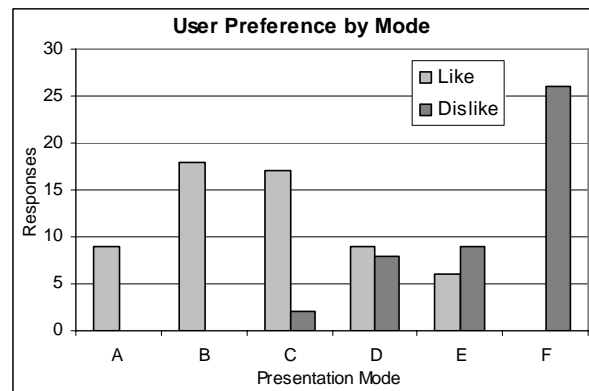


Figure 14: Favourite (left) and least favourite (right) Modes

8. EYE-GAZE BEHAVIOUR

Examination of the results shown in Figures 15 to 18 suggests that eye-gaze may play a significant role in the success of a presentation mode. For example, Figure 15 shows the eye-gaze trace for a subject who disliked the diagonal (moving) mode, whereas Figure 16 shows the trace for a subject who preferred it. Saccades and eye movements are represented as a green line (also refer to colour plate in proceedings); fixations are denoted by 'F'. The difference between these and similar traces is striking, and strongly suggests that eye-gaze strategies may underlie not only the preferences of subjects but also the recognition error performance. Notable in figure 16 is the concentration of gaze at the 'capture' location (bottom right) where, for a short while, an image is static. Similar evidence is provided, for example, by subjects who disliked and preferred the ring mode (Figures 17 and 18). Investigations continue to confirm this effect.

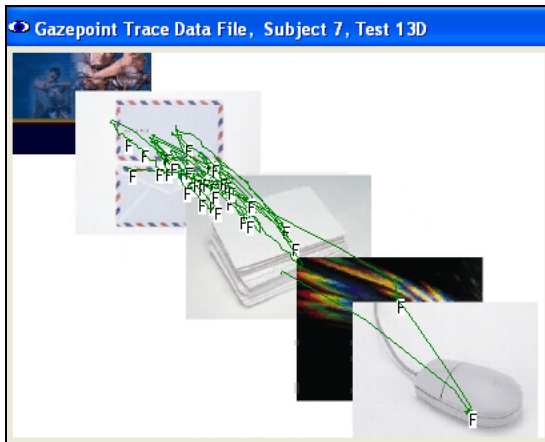


Figure 15: Eye-gaze trace of a participant who disliked the diagonal mode

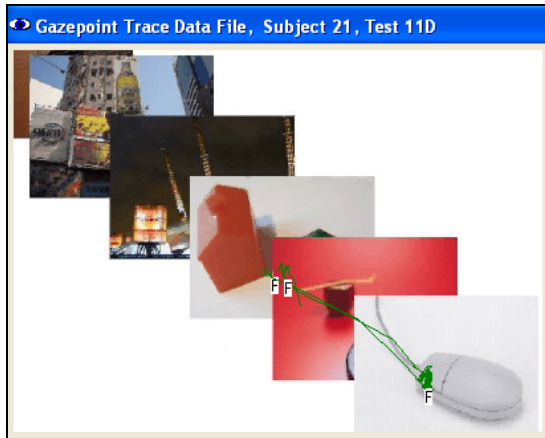


Figure 16: Eye-gaze trace of a participant who liked the diagonal mode

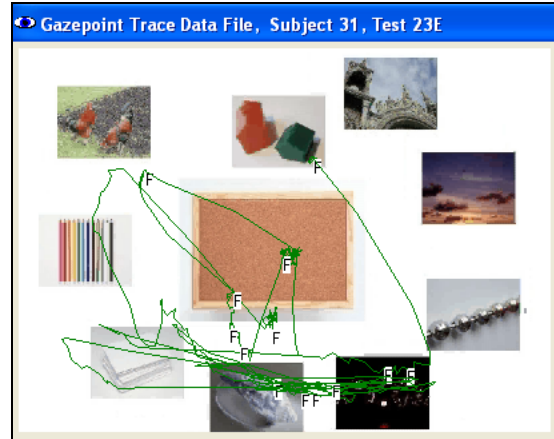


Figure 17: Eye-gaze trace of a participant who disliked the ring mode

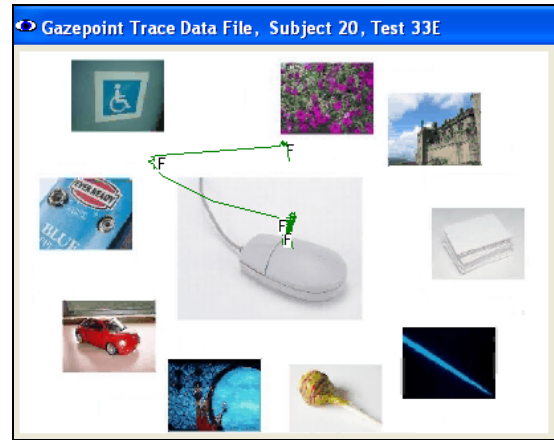


Figure 18: Eye-gaze trace of a participant who liked the ring mode

9. CONCLUSIONS

The proposed simple taxonomy (classifying presentation modes as 'static' or 'moving') has led to confirmation of the two hypotheses about these classes of presentation modes, in that static modes were found to be better than moving modes as far as recognition success and user preference were concerned. In general, our experiments identified a coupling between identification success, subject preference and eye-gaze behaviour.

Static presentation modes generally provide ample opportunity for eye-gaze behaviour to have low saccadic excursion, and seem to be correlated with user acceptance and high recognition success. In contrast, the likelihood that eye-gaze will roam widely appears to be associated with moving modes, with resulting degradation in both recognition success and user acceptance. There is, however, one exception, and that is the performance observed when a moving mode contains a 'capture' effect (Wittenburg *et al.*, [20]) as with the 'ring' and 'diagonal' modes. Provided eye-gaze is focused on the captured image (effectively treating the presentation as 'slide-show mode' in a restricted area of the display), recognition and acceptance are enhanced.

We conclude that the overall guidance provided for interaction designers who for some reason wish to employ a moving mode

should be to ensure that at some location each image is 'captured', even if only for 100 to 200msec. As our experimental results show, however, there is no assurance that the user will make use of this potential advantage. In selecting a moving mode indications are that the designer would be well advised to consider the effects of the user's gaze strategy, if possible, by direct measurement. In the event that gaze-tracking technology becomes commonplace on desktop PC, laptop and PDA devices, the individual's strategy could be monitored and the form of presentation mode adapted accordingly.

The effect of perceived cognitive effort has also been demonstrated, though not in any analytical manner that would allow predictions to be made regarding optimum design (Potter, [9]). We have examined just three tasks of assumed different cognitive effort, and many others remain to be investigated: the study of mobile-RSVP by de Bruijn and Tong ([4]), for example, suggested that the task of selecting an appropriate news channel on the basis of representations involving a simple image and brief text might well require a presentation time of around 500 msec for each channel. Further studies are required.

10. FUTURE RESEARCH

Accumulated evidence of the primary role of eye-gaze in the search for a single image in a collection points to the need for better understanding of human visual processing under the conditions we have described and in the context of the applications to which our investigations have ultimately been directed. It may well be the case that Change Blindness (Rensink, [11]; Varakin, [16]), Inattentional Blindness (Simons & Chabris, [12]) and models such as Conceptual Short-term Memory (Potter, [8], [9]) have considerable relevance. State-of-the-art eye-gaze detectors are such as to raise the possibility of using eye-gaze to direct navigation through an image collection. The investigation can also be extended to encompass other forms of task, such as the selection of news channels from rapidly presented summaries on a mobile, as explored by de Bruijn and Tong ([4]).

11. REFERENCES

- [1] Coltheart, V. (1999) Introduction: Perceiving and Remembering Brief Visual Stimuli, in Coltheart, V. (Ed.) *Fleeting Memories: Cognition of Brief Visual Stimuli*, Cambridge, MA. MIT Press, pp.1-12.
- [2] Coltheart, V. (1999) *Fleeting Memories: Cognition of Brief Visual Stimuli*, Cambridge, MA. MIT Press.
- [3] de Bruijn, O. and Spence, R. (2002) Patterns of Eye Gaze During Rapid Serial Visual Presentation, *ACM, Proc. Int. Conf. On Advanced Visual Interfaces*, pp.209-217.
- [4] de Bruijn, O. and Tong, C.H. (2003) M-RSVP: Mobile Web browsing on a PDA, in O'Neill, E., Palanque, P and Johnson, P. (Eds) *People and Computers – Designing for Society*, Springer, pp.297-311.
- [5] Fawcett, C., Craft, B., de Bruijn, O., Witkowski, M. and Spence, R. (2004) Image Presentation in Space and Time: Errors, Preferences and Eye-gaze Activity, *ACM, Proceedings of the Conference on Advanced Visual Interfaces, AVI2004*, pp.141-149.
- [6] Komlodi, A. and Marchionini G. (1998) Key frame preview techniques for video browsing, *ACM, Proceedings of Digital Libraries'98*, pp.118-125.
- [7] Lam, K. and Spence, R. (1997) Image Browsing –a space-time trade-off, *Proc. INTERACT'97*, pp.611-612.
- [8] Potter, M. (1993) Very short-term conceptual memory, *Memory and Cognition*, 21, pp.156-161.
- [9] Potter, M. (1999) Understanding Sentences and Scenes: The Role of Conceptual Short-Term Memory, in Coltheart, V. (ed.) (1999) *Fleeting Memories: Cognition of Brief Visual Stimuli*, Cambridge, MA, MIT Press, pp.13-46.
- [10] Raymond, J.E., Shapiro, K.L. and 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, pp.849-860.
- [11] Rensink, R.A. (2002) Change Detection, *Annual Review of Psychology*, 53, pp.245-277.
- [12] Simons, D.J. and Chabris, C.F. (1999) Gorillas in our midst: sustained inattention blindness for dynamic events, *Perception*, 28, 9, pp.1059-1074.
- [13] Spence, R. (1998) A Content Explorer, Imperial College, *Information Engineering Report 98/08*.
- [14] Spence, R. (2002) Rapid, Serial and Visual: a presentation technique with potential, *Information Visualization*, 1, 1, pp.13-19
- [15] Tse, T., Marchionini, G., Ding, W., Slaughter, L. and Komlodi, A. (1998) Dynamic key frame presentation techniques for augmented video browsing, *ACM, Proc. Conf. on Advanced Visual Interfaces*, pp.185-194.
- [16] Varakin, D.A., Levin, D.T. and Fidler, R. (2004) Unseen and Unaware: Implications of Recent Research on Failures of Visual Awareness for Human-Computer Interface design, *Human-Computer Interaction*, 19, pp.389-422.
- [17] Witkowski, M., Neville, B. and Pitt, J. (2003) Agent Mediated Retailing in the Connected Community, *Interacting with Computers*, 15, pp. 5-32.
- [18] Wittenburg, K., Ali-Ahmad, W., LaLiberte, D. and Lanning, T. (1998) Rapid-fire image previews for information navigation, *ACM, Proceedings of the Conference on Advanced Visual Interfaces (AVI-1998)*, pp.76-82.
- [19] Wittenburg, K., Chiyoda, C., Heinrichs, M. and Lanning, T. (2000) Browsing through rapid-fire imaging: requirements and industry initiatives, *SPIE, Proceedings of Electronic Imaging*, pp.48-56.
- [20] Wittenburg, K., Forlines, C., Lanning, T. Esenther, A., Harada, S. and Miyachi, T. (2003) Rapid Serial Visual Presentation Techniques for Consumer Digital Video Devices, *ACM, Proceedings Symposium on User Interface Software and Technology (UIST)*, pp.115-124