Image Presentation in Space and Time: Errors, Preferences and Eye-gaze Activity

Bob Spence, Mark Witkowski, Catherine Fawcett
Intelligent and Interactive Systems Group
Dept. of Electrical and Electronic Engineering, Imperial College, London SW7 2BT, U.K.
r.spence@ic.ac.uk

Brock Craft
University College London
Interaction Centre (UCLIC)
Remax House, 31/32 Alfred Place, London, WC1E 7DP, U.K.
b.craft@ucl.ac.uk

Oscar de Bruijn
Interactive Systems Design
Computation Department
UMIST, P.O. Box 88, Manchester M60 1QD, U.K.
O.deBruijn@co.umist.ac.uk

ABSTRACT

Rapid Serial Visual Presentation (RSVP) is a technique that allows images to be presented sequentially in the time-domain, thereby offering an alternative to the conventional concurrent display of images in the space domain. Such an alternative offers potential advantages where display area is at a premium. However, notwithstanding the flexibility to employ either or both domains for presentation purposes, little is known about the alternatives suited to specific tasks undertaken by a user. As a consequence there is a pressing need to provide guidance for the interaction designer faced with these alternatives.

We investigated the task of identifying the presence or absence of a previously viewed image within a collection of images, a requirement of many real activities. In experiments with subjects, the collection of images was presented in three modes (1) 'slide show' RSVP mode; (2) concurrently and statically – 'static mode'; and (3) a 'mixed' mode. Each mode employed the same display area and the same total presentation time, together regarded as primary resources available to the interaction designer. For each presentation mode, the outcome identified error profiles and subject preferences. Eye-gaze studies detected distinctive differences between the three presentation modes.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Graphical User Interfaces (GUI), Interaction Styles, Screen Design.*

General Terms

Design, Experimentation, Human Factors.

Keywords

Rapid Serial Visual Presentation; Space-time Trade-off; Visual Interface Design; Visual Information Browsing; Eye-gaze Tracking.

1. IMAGES

Many tasks involve the perception and interpretation of images. A shopper presented with a new catalogue, for example, will often riffle through its pages to find products of potential interest [16]. Or a television viewer may examine a collection of 'stills' taken from a film to identify the 'gist' of the film sufficiently well to decide whether to view it [12, 15]. Or a large collection of family snapshots is browsed to select some to be shown to a friend [10]. A similar task is faced by a person in a meeting who wants to 'find their place' in a report being discussed, and is able to do so rapidly by recognition of the page layout, images and annotations.

Images intended for viewing and interpretation in graphical interfaces are often presented concurrently and statically. If insufficient space is available they are arranged within pages which are under the control of the user. Recently, however, a new method of presentation has evolved which, to some extent, can relieve the pressure on space. The new method is called Rapid Serial Visual Presentation (RSVP).

The rapid serial presentation of text (such as with scrolling marquees or stock tickers), often designated as RAPCOM (Rapid Communication), has been studied extensively, especially within military applications where fast interpretation is sought. [1]. More recently, however, the existence of many tasks requiring the interpretation of a large number of *images* has focused attention on the potential offered by their rapid presentation. In RSVP, images are typically presented sequentially, and usually at a high rate (e.g., 10 per second) appropriate to the human being's remarkable ability to recognise images displayed for as little as 100 milliseconds or even less [2, 9]. The physical equivalent of such 'slide-show' RSVP is the fast riffling of the pages of a book often undertaken to acquire, quite rapidly, some idea of its contents.

'Slide-show is not the only available RSVP mode [3, 11]. Others differ according to the 'route' taken by each image within the display area, the variation of image size during that transition and any image occlusion and/or temporary freezing that may occur. Already, RSVP techniques have been applied to a variety of applications including TV channel selection [15], Web browsing on a PDA [4] and online and kiosk based shopping [14]. There is, therefore, considerable interest in how best to exploit this new method of image presentation.

2. INTERACTION DESIGN

RSVP exploits the time domain for the presentation of a collection of images, offering an alternative to the conventional static and concurrent presentation of all images at the same time. The question therefore arises in the mind of an interaction designer considering the presentation of a collection of images to a user, as to which mode of presentation is best suited to a given user task. Indeed, the possibility of a mixture of the two modes must also be given consideration.

To elicit answers to these and other related questions we conducted experiments to compare the relative merits of the three presentation modes shown in Figure 1 (see Plate 1) for a task (see later) in which a user has to identify a familiar image within a collection. For simplicity of illustration we assume a collection of only 16 images. In 'slide-show' mode the sixteen images are presented, sequentially and rapidly, at the same location. If the images were to be presented at 100ms intervals – a not untypical rate – then a total presentation time of 1.6 seconds would be required. For comparison, in the 'static' mode presentation, all images are displayed concurrently for the same time (1.6 seconds) that is assigned to the entire slide-show presentation. In what we have termed the 'mixed-mode' presentation, four images appear simultaneously, so that four sets of four images are presented, again for 1.6 seconds to allow comparison with other modes.

The basis of our comparison of the three modes – a fixed total presentation time – is the premise that an interaction designer will be required to support a user task within limits on display area and total presentation time. Thus, our comparison of the three modes is resource-based: it is designed to inform an interaction designer charged with arranging a presentation within limits on display area commensurate with other uses of that area, and with a presentation sufficiently long to allow a task to be performed with acceptable accuracy yet sufficiently short so as not to impede some other task. Commensurate with this basis for comparison it was hoped to be able to characterise a 2-dimensional 'design space' associated with the two resources. One is display area. The other is 'time-per-image' T_i, a parameter which must be multiplied by the number of images in the collection to yield a total presentation time relevant to all modes.

To summarise, the principal motivation for our study was the need to inform an interaction designer concerned with the presentation of a collection of images by comparing the relative merits of the three presentation modes illustrated in Figure 1.

3. THE TASK

The task investigated was a simple but representative one. A subject is first required to look at a 'target' image. The subject is then told that they will be presented with a collection of images, which might or might not include the target image and is asked to indicate, during or after the presentation, whether or not they thought the target image was present. After each presentation the subject was asked to rate the difficulty of the task on a 5-point scale, and at the end of a complete session the subject was asked to indicate their order of preference for the three presentation modes.

4. EXPERIMENTAL DESIGN

For each of the three presentation modes illustrated in Figure 1 the experiment was carried out for 4 different total presentation times, for 3 display areas and, to avoid false positives, with the target image both present and absent. Thus, each subject performed the task (3x4x3x2) = 72 times. The need to explore four presentation times arose from our desire to establish a boundary in design space defining essentially error-free task completion.

The images were selected to be distinct and distinctive, and were in full colour. Figure 2 (see Plate 1) shows a (representative) selection of the images used in the eye-gaze experiments. Of the 200 images 76 were of human manufactured artefacts (for instance, common office or household items); 68 represented scenes of manmade vistas, such as interior or exterior views of buildings or groups of fabricated objects; 20 were of natural scenes or outdoor spaces and 12 were of human subjects. The remainder were of animal subjects, natural artefacts or were highly abstracted images. Figure 3 (see Plate 1) illustrates the mode/area combinations.

5. PROCEDURE

Presentations of the image collection were made with variations along three separate dimensions: presentation mode, display area and total presentation time. In the experiments the volunteer subjects were asked to memorise a target image and to say whether or not that image appeared in the collection of images presented immediately afterwards. As illustrated in Figure 1, three image presentation modes were tested:

- "Static" or thumbnail display, in which 64 images were presented in a composite 8x8 matrix for the total presentation time.
- "Mixed" display, in which the 64 images were presented in composite groups of four (in a two by two matrix), each group being presented for $1/16^{th}$ of the total presentation time.

- "Slide show" display, in which the 64 images were presented sequentially individually, each image being presented for 1/64th of the total presentation time.

Three display areas were used:

- Area 1: The individual or composite images were presented at the full screen size (approx. 32 x 24 cm, 640 x 480 pixel, 307k pixel resolution). This represents a normal size presentation on a computer terminal screen.
- Area 2: The individual or composite images were presented at quarter screen area (approx. 16 x 12 cm, 320 x 240 pixel 77k pixel resolution). This mode represents the presentation area and resolution of a handheld or PDA device. The individual and composite images were centred on the screen and unused screen area was set to black.
- Area 3: The individual or composite images were presented at one-sixteenth screen area (approx. 8 x 6 cm, 160 x 120 pixel, 19k pixel resolution). This mode represents the presentation area and resolution of a modern mobile telephone device.

To generalise the results to any display size at any distance from the subject it may be useful to note that the vertical angle subtended at the subject's eye for the three display areas was approximately 30, 15.2, and 7.6 degrees for areas 1 to 3 respectively.

Four total presentation times were used: 960ms, 2560ms, 4160ms and 5760ms. The relatively low display resolution was selected to allow a correspondingly high monitor scan rate to support the higher presentation rates. These times were determined empirically from trial studies to adequately bracket the onset of substantial error rates.

In each presentation instance the subjects were shown a selection of 64 images drawn at random from a sample set of 200 different (photographic) images. The composite images, and the sequence in which images would be presented, were randomised and

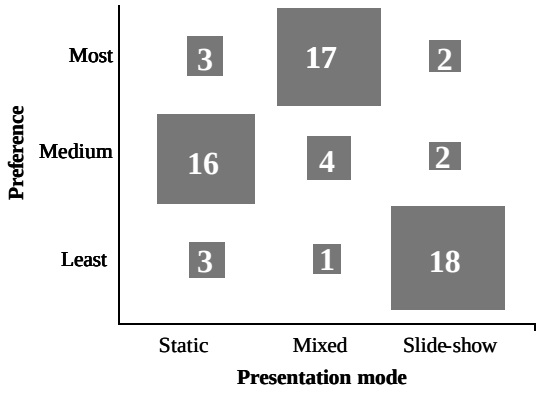

Figure 4. Number indicates how many subjects indicated a preference level for each mode.

constructed dynamically just prior to the presentation sequence. To complete the procedure each subject was shown a total of 72 presentations, one at each combination of the three modes, the three sizes and the four total presentation times where the target image was in the display set, and once in each of the combination of mode, size and time where the target image was not in the display set (i.e. 3 modes x 3 sizes x 4 times x 2 target conditions). Each of these combinations was shown to each subject once, in a random order.

In total 30 volunteer subjects were tested, 15 male and 15 female, all aged between 20 and 23. All were drawn from the Imperial College student population and were computer literate. All had normal or corrected to normal vision and each performed an online Ishihara test for colour blindness prior to the test to confirm normal colour vision. Prior to the experiment each was informed of the purpose of the experiment and consent was obtained. Subjects sat at a normal viewing distance from the screen (about 45 cm). Once started, the procedure was fully automated and the experimental conditions and subject responses were recorded to a file for later analysis. For each of the 72 presentations:

- Step 1: The target image to be identified in the sequence was

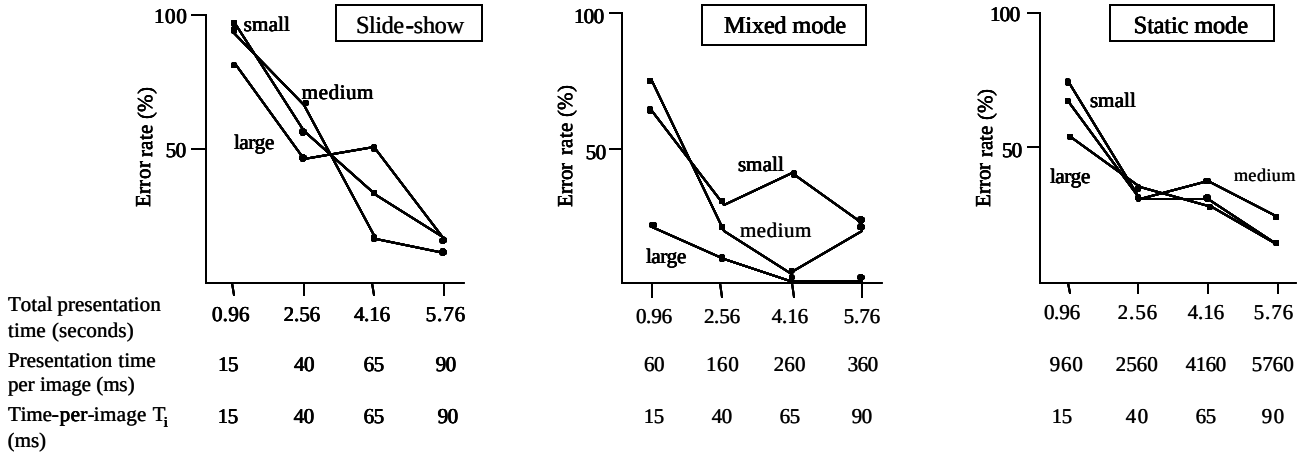

Figure 5. Error rate compared to total presentation time for slide-show, mixed, and static modes at various display areas.

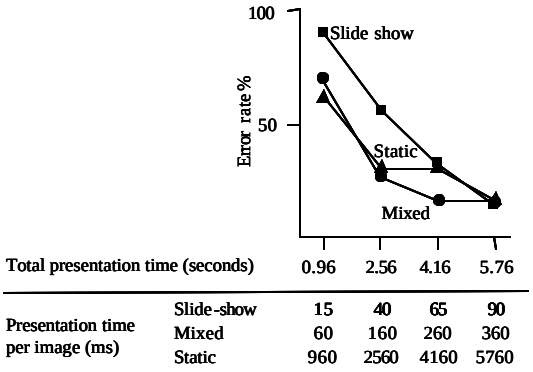

Figure 6. The percentage error rate for the three presentation modes, averaged over all three display areas and 30 subjects.

displayed for 10 seconds

- Step 2: After two seconds of blank screen the sequence of 64 images was presented in one of the 72 combinations described.
- Step 3: The subject was presented with a "Windows" style dialog with a radio button pair to indicate whether the target was, or was not deemed to be in the sequence; and a 5 way radio button selector for the subject to indicate how they found the speed of presentation, selecting between "Too fast", "Slightly fast", "Just right", "Slightly slow" or "Very slow". Once the selection was confirmed the sequence restarted at Step 1 until all 72 combinations had been presented. Each session lasted approximately 30 minutes.

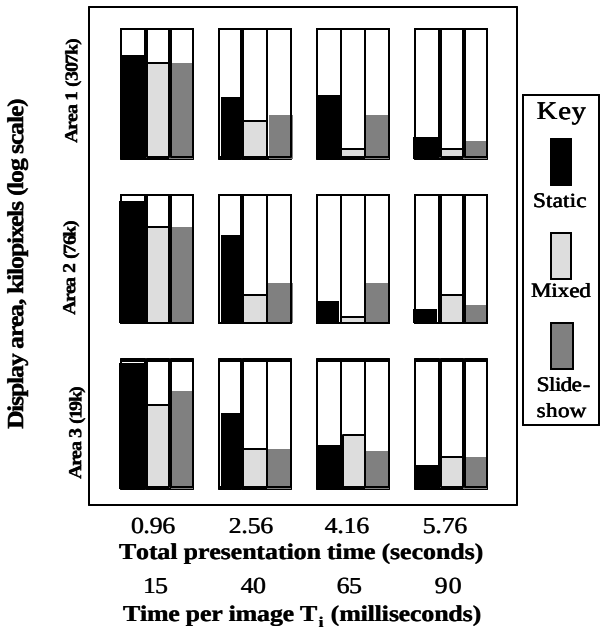

Figure 7. The percentage of errors for each presentation mode, as a function of area and total presentation time (30 subjects). A completely filled bar indicates 100% error rate.

6. EXPERIMENTAL RESULTS

6.1 User Preference

Remarkable consistency was shown in subjects' stated order of preference for the three presentation conditions: the mixed mode was most preferred, the static mode was second and the slide-show mode was least preferred (Figure 4).

6.2 Errors

Of particular interest was the effect of presentation time and display area on error rate. The results for the slide-show mode (Figure 5) cause no surprises: a largely monotonic increase in error rate as the image presentation time decreases, with the onset of unacceptable error occurring at the expected [2] image presentation time of about 100ms. The generally monotonic increase in error rate is also – again not surprisingly – a feature of the mixed and static modes.

If error rates are averaged over the three display areas, the result (Figure 6) suggests that the mixed mode, and to a lesser extent the static mode, leads to a lower error rate when image presentation times fall below 100ms. One suggested explanation for the mixed mode result, explored in more detail below, is that each image is visible for four times longer than in the slide-show mode, and that eye-gaze can effectively be focused at the centre of the display area rather than saccading to widely different regions.

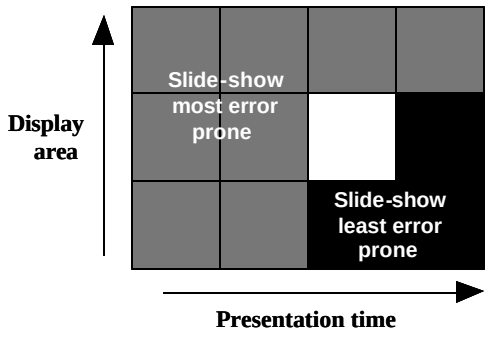

Figure 8. Areas of resource space where the slide-show mode is the least and most error-prone of the three modes.

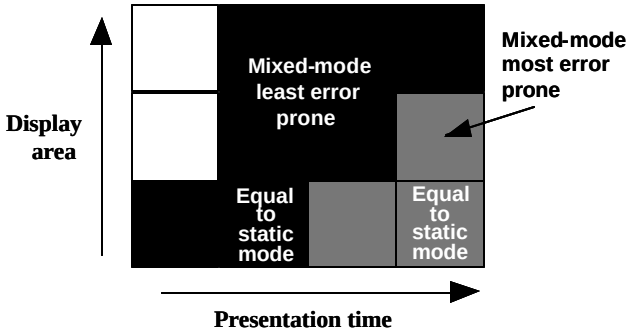

Figure 9. Areas of resource space where the mixed mode is the least and most error-prone of the three modes.

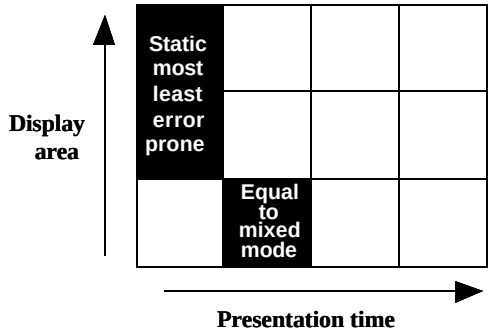

Figure 10. Areas of resource space where the static mode is the least and most error-prone of the three modes.

Figure 7 presents the error results in the resource plane, with axes representing the display area and the equivalent presentation time per image (= total presentation time/number of images). Trends and patterns, however, are more easily perceived from the discretised versions of Figure 7. Thus (Figure 8), at low presentation times, as well as for the large display area, the slide-show mode was found to be the most error-prone. This mode is, in fact, the least error-prone in only three locations in the resource plane, and then only marginally so. A similar though not identical pattern (Figure 9) is exhibited by the mixed mode: this mode is either the least error-prone or no more error-prone than the static mode in seven of the 12 locations observed and is the single most error-prone in only two locations. We also note (Figure 10) that in only two of the 12 explored locations is the static mode less error prone than any other mode. These observations of the results presented in Figure 7 must of course be interpreted with caution in view of the fact that only 30 subjects were tested. A rigorous statistical analysis of the results is reported in section 7.

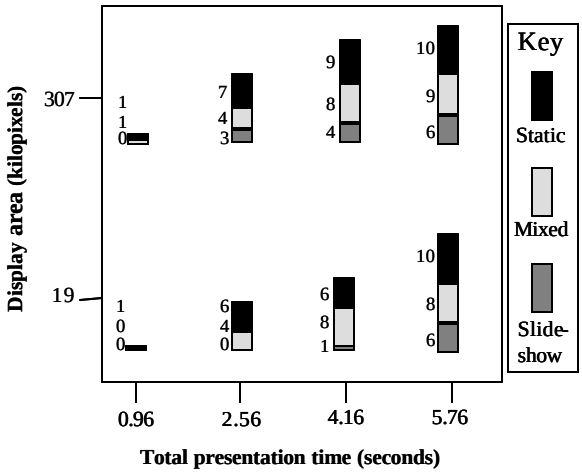

Figure 11. The acceptability of presentation time for the three presentation modes, as a function of total presentation time and display area. Column heights, and accompanying numbers, indicate the number of subjects expressing acceptability.

6.3 ACCEPTABILITY OF PRESENTATION TIME

After each presentation of the image collection the subject was asked to rate the speed of presentation on a 5-point scale: 'too fast', 'slightly fast', 'acceptable', 'slightly slow' and 'very slow'. For the largest (area 1) and smallest (area 3) display areas Figure 11 shows the number of instances judged to be satisfactory in the sense that the presentation rate was felt to be 'acceptable', 'slightly slow' or 'very slow'. To take an example, for the longest presentation time and for Area 1, five subjects (=30 – (10+9+6)) judged the presentation to be either 'slightly fast' or 'too fast'. There appeared to be very little effect of display area and, not surprisingly, an increase of acceptability with total presentation time. Nevertheless, the mixed and static modes of presentation appear to be more acceptable than the slide-show mode.

7. STATISTICAL ANALYSIS

The overriding assumption behind the statistical analysis was that it was most difficult to detect the target when images were presented for a total of 960ms and least difficult when they were presented for a total of 5760ms. From Figure 6 it is clear that this was generally the case. Therefore, for each of the combinations of presentation mode and screen size, we determined what the longest presentation time was with an incorrect detection. These errors were then scored depending on the presentation time condition in which they were obtained. For example, suppose a participant was able to correctly detect the target in the 160x120 slide-show mode presentation with total presentation times of 5760ms and 4160ms, but not in the 2560ms and 960ms conditions. In that case, a score of 2 would be noted for that combination of display area and presentation mode, because the easiest condition with an incorrect detection was the 2560ms one, the 2nd most difficult condition. After this recoding, we are left with a dependent variable with 5 levels (0-4, with 0 meaning no errors at all).

We then carried out an ANOVA using the Generalised Linear Model [8] to predict the fastest presentation rate at which subjects started making errors as revealed by the newly formed variable described above. It is first of all worthy of note that display area did not affect the point at which subjects started making errors [$F(2, 62) = 2.026$, $MSE = .877$ – not significant). This suggests that the apparent interaction between display area and presentation time in the mixed mode must be a spurious result that is not upheld by our statistical analysis. Secondly, our analysis indicates that there is a significant interaction between presentation mode and time [$F(2,62) = 22.64$, $MSE = .589$, $p < .001$]. This finding confirms the difference in the slopes of the lines in the slide-show mode compared to the mixed and static modes apparent in Figure 5. That is, the error rate increases much less with decreases in presentation time in the latter two conditions compared to the former. The three-way interaction between presentation time, display area and presentation mode was not significant [$F(4, 124) < 1.0$ – not significant].

8. EYE-GAZE BEHAVIOUR

Our attention was drawn to the strong preference expressed by subjects for the mixed-mode presentation as well as to a tendency for the mixed-mode presentation to be associated with fewer errors than the other two presentation modes. An hypothesis was formulated that these results could be consistent with a tendency for a subject to fix their gaze at the centre point of the four images and to keep it there during the entire presentation: a suggested advantage is that each image is thus presented for four times the duration associated with the slide-show mode, and that the foveal region is sufficiently extensive that a subject could decide, without saccadic movement, whether the target image is present. An experiment was carried out to test this hypothesis.

8.1 Procedure

Eye-gaze behaviour was measured using equipment manufactured by LG Technologies [7]. Subjects under measurement use a normal computer screen, keyboard and mouse, but have the movements of one eye monitored by a CCD camera placed directly under the monitor. An infra-red beam is projected axially from the centre of the camera giving reflections both from the corneal surface of the eye and from the retina. As the subject's point of gaze shifts the relative positions of these reflections move in the image of the eye obtained, from which the point of gaze can be reconstructed with an accuracy of (typically) less than 5mm relative to the screen image (technical details of this, and other gaze measuring techniques, are reviewed in [6]) The system works in real-time and returns the effective screen X, Y coordinates of the gaze point, the pupil diameter, and blink events every 60mS (once per video frame).

The system adopted is convenient as there is no need to attach any equipment to the subject or to clamp or artificially restrain their movement, readings being obtained so long as the subjects do not move their head position excessively, although if they do the experimental data obtained must be rejected. Prior to each group of measurements the subject must be positioned in the camera field of view and a calibration sequence completed in which the subject is required to look at calibration points displayed across the screen. Calibration takes about 90 seconds.

To investigate the hypothesis described, a modification of the previously described procedure was undertaken. Subjects were again shown target images drawn from the image set, followed by sequences of images (also drawn from the set) in the slide-show, mixed, and static modes, but only at full display area, and with the total presentation timings modified to 1600, 3200 and 6400ms respectively. The subjects were also shown each sequence only once (the target was always present). Eighteen (18) subjects were drawn from the University College London Interaction Centre (UCLIC) student population. As part of the requirements for informed consent, the procedure and purpose of the equipment was explained to the subjects, who were therefore made more aware of their own gaze strategy and behaviour than would (we presume) be normal. Of eighteen procedures undertaken, 12 gave rise to valid eye-gaze data, and our analysis is restricted to this sub-set. Eye tracking technology has been used previously to analyse RSVP and related technologies [5], [13].

8.2 Analysis

Figures 12, 13 and 14 show the eye-gaze patterns produced by a representative subject. The gaze path is shown as a continuous line (starting just above the centre of the image in each case). Detected fixations are shown by "F" characters. In each case, the complete eye-gaze path for the total duration of the presentation mode is shown. The background image shown is the one acting as, or containing, the target image for that presentation sequence. For the slide-show mode, the subject's gaze was, perhaps unsurprisingly, essentially fixed at the approximate centre of the display area (see Figure 12, plate 1). The faster the presentation rate, the more pronounced this tendency became.

This same behaviour was observed (see Figure 13, plate 1) for the mixed-mode presentation, but to a lesser extent. In particular, it was noted that a substantial proportion of the subjects showed gaze movements towards the target image quadrant, though due to the rapidity of presentation the gaze movements typically occurred several image presentations after the target appeared. Subjects also demonstrated a tendency to hold a gaze position for a considerable period (in comparison to the normal saccade/fixation gaze strategy normally encountered with static images).

For the static mode, where all images are presented simultaneously, the subject's eye-gaze exhibited a search pattern over the whole display area (see Figure 14, plate 1). We note that the subject's gaze was drawn to a cluster of images (laptop, inkpot, diary, and camera), which are rather similar in appearance to the designated target (the briefcase). We conjecture that purposefully grouping images by some measure of visual similarity may provide some advantage in this mode of presentation. In the longer presentation times, where the target image has been acquired, we noted a tendency to continue searching behaviour, often returning to the target image.

The results shown in Figures 12, 13 and 14 are presented to show typical examples of eye-gaze behaviour in each of the three presentation modes. We also experimented with quantitative measures of eye-gaze movement to see how they varied with presentation mode. One measure, termed ''radial dispersion'' was obtained by computing the average distance of the gaze points from the centre of the display area, to provide some idea of the extent of the region within which the eye-gaze trajectory lies. For the three presentation modes the average radial dispersion for twelve subjects, and its standard deviation, is shown in Figure 15. We note the much higher values of radial dispersion for the static mode than for the mixed and slide-show modes, and the lack of effect of presentation time. Statistical analysis confirmed the lack of effect of presentation time, and confirmed that the average

distance of eye-gaze from the centre of the display was significantly ($p<0.01$) greater for the static mode compared with the other two modes. Analysis also detected a significant difference between the mixed and slide-show modes for a total presentation time of 1600ms.

The second quantitative measure investigated was the total distance travelled between gaze-points and the corresponding average speed. Table 1 shows the results averaged over twelve subjects for the 3200ms presentation time. Again there is a noticeable difference between the static mode and the other two modes with regard to both distance and speed. A Repeated Measures ANOVA revealed that both presentation mode and time had significant effects on the average distance travelled by the eyes during presentation [$F(2, 22)= 58.57$, and $F(2, 20)=123.35$ respectively, $ps<.001$]. However, there was also a significant interaction between these two factors [$F(4, 40)= 15.8$, $p<.001$], suggesting that the increase in eye movements with increasing presentation time was indeed much more dramatic in the static mode than in the semi-static and mixed modes. A Repeated Measures ANOVA revealed that the average speed of eye movements was significantly affected by the mode of presentation [$F(2, 20)= 150.74$, $p<.001$], confirming the suggestion that the speed at which the subjects' eyes travelled was higher in the static mode than in the mixed and slide-show modes of presentation.

From Table 1 it is clear that the effect of presentations mode on distance travelled is similar to that on average speed, as would be expected given that the duration of saccades is more or less independent of distance travelled. The speed at which the eye gaze travels increases, therefore, with increases in the distance travelled.

Table 1. Total distance travelled between gaze-points, and speed of gaze movement, averaged over 12 subjects. 60Hz samples.

	Static	Mixed	Slide-show
Average total distance (pixels per sample)	3582	1885	1970
Average speed (pixels per sample)	18.83	7.85	5.03

9. RAMIFICATIONS FOR INTERACTION DESIGNERS

Our original motivation for the study reported above was to inform the interaction designer. While many of the questions that can be posed by an interaction designer still remain unanswered we can nevertheless make some pertinent and potentially useful observations:

(1) For all three modes error rates become substantial as the time per image falls below 100ms (as expected);
(2) As time per image falls below 100ms the error rate associated with mixed and static modes tends to be less than in the slide-show mode;

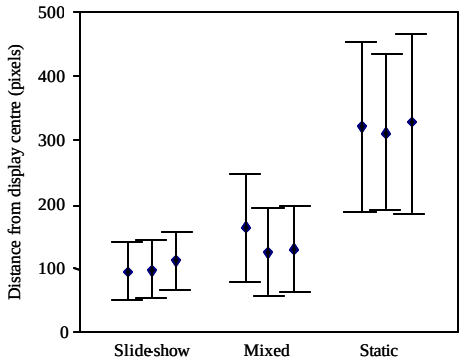

Figure 15: Average radial dispersion for each mode. 1600, 3200, and 6400 millisecond presentations, respectively for each mode, with standard deviations.

(3) Mixed mode was the least error prone of the three modes over a substantial region of resource space;
(4) Static mode was the least error-prone for only a small number of locations;
(5) User preference was strongly in favour of the mixed mode;
(6) Statistical analysis showed a significant interaction between presentation mode and the shortest presentation time at which subjects started to make errors;
(7) Statistical analysis showed that display area did not significantly affect the point at which subjects started making errors.

10. DIRECTIONS FOR RESEARCH

Though the reported results have shed some light on the ways in which a collection of images might be presented, they answer very few of the many questions that arise when slide-show or other modes are being considered for use in a given application. However, as a result of our study we have been able to identify potentially fruitful directions of research.

Task: To limit the scope of our study we selected a single task, that of identifying a known image among a collection of images. Of equal interest is a task in which a user has no previous knowledge of the collection, but wishes to gain insight into it. Research into such a task has already been reported by Tse et. al. [12], in the context of discovering the 'gist' of a collection of key-frames taken from a video. It is also frequently the case, as with online sales, that a user does not initially know what is being sought except in the most general terms ("a present for Mom on Mothers' Day"). Both these tasks deserve study in the context of available presentation modes.

Control: It may on occasion be more effective if the rate of presentation of images is under the control of the user. Again, some relevant work has appeared in the literature: Wittenburg et. al. [16] have not only demonstrated such control, but have commented upon the design issues that arise. There are many opportunities for further research.

Trajectories: Images within a collection can move along many different paths, their size can vary with time, they can overlap with their neighbours to a greater or lesser degree and it may be useful for them to halt momentarily at some point on the trajectory [15]. Research is needed to establish the comparative benefit of a variety of space and time trajectories and, most importantly, to identify a basis – perhaps eye-gaze – from which these benefits can be predicted.

11. SUMMARY AND CONCLUSIONS

Rather than focus on traditional slide-show RSVP we have instead addressed the more relevant and general question as to how a collection of images should be presented to support a given task. To this end, we suggested a resource basis for the comparison of presentation modes, the resources being *display area* and *total presentation time*. For comparisons to be independent of the number of images in a collection, we introduced a parameter T_i, being a time period assigned to each image in a collection irrespective of the presentation mode, the individual values of this parameter being summed to yield the total presentation time.

The task of identifying the presence or absence of a familiar image within a collection was investigated in experiments on a number of subjects. The results identified user preferences regarding presentation mode and presentation rate, the most striking outcome being a strong preference for the mixed mode. With regard to the error rate of target recognition, different modes appeared to be associated with distinct regions of the resource plane; slide-show mode was often the most error-prone, while the mixed mode was often the least error-prone. The static mode was rarely the least error-prone. Statistical analysis showed that as total presentation time decreases, error rate increases much less rapidly with the mixed and static modes than with the slide-show mode, and that display area had no significant effect.

The apparent advantage of the mixed mode of presentation stimulated a study of eye-gaze behaviour. From experiments on twelve subjects, three measures of eye-gaze behaviour – radial dispersion, average speed, and total distance travelled – were computed over the subjects. Distinctive combinations of the three measures were exhibited by the three presentation modes, and statistical analysis showed less radial dispersion with slide-show and mixed modes of presentation, than with static mode. Additional analysis showed that eye movements increased for static mode presentation as presentation time increased.

12. ACKNOWLEDGEMENTS

Support from the EPSRC (Grant GR/N20881) is gratefully acknowledged. We are grateful to Basil Fakhry and Karen Lam for their valuable contributions.

13. REFERENCES

[1] Boff, K.R. and Lincoln, J.E. (1988) "Engineering Data Compendium: Human Perception and Performance", Wright-Patterson AFB, OH: USAF Harry G. Armstrong Aerospace Medical Research Laboratory (AAMRL).

[2] Coltheart, V. (Ed.) (1999) "Fleeting memories", Cambridge, MA: MIT Press.

[3] de Bruijn, O. and Spence, R. (2000) "Rapid Serial Visual Presentation: a Space-time Trade-off in Information presentation", Proc. Int. Conf. on Advanced Visual Interfaces (AVI-2000), pp. 189-192.

[4] de Bruijn, O., Spence, R. and Chong, M. (2001) "RSVP Browser: web browsing on small screen devices", Proc. Mobile HCI 2001 – 3rd Int. Workshop on Human Computer Interaction with Mobile Devices, Lille, France.

[5] de Bruijn, O. and Spence, R. (2002) "Patterns of Eye Gaze During Rapid Serial Visual Presentation", Proc. Int. Conf. on Advanced Visual Interfaces (AVI-2002), pp.209-217.

[6] Duchowski, A.T. (2003) "Eye Tracking Methodology : Theory and Practice", Springer.

[7] LC Technologies (2003) "Eye-tracking from LC Technologies, Inc.", http://eyegaze.com

[8] Meyers, R.H., Montgomery, D.C., & Vining, G.G. (2002). Generalized Linear Models With Applications in Engineering and the Sciences. Wiley series in probability and statistics, Wiley and Sons.

[9] Potter, M. C. (1999). "Understanding sentences and scenes: The role of Conceptual Short Term Memory", in V. Coltheart (Ed.), Fleeting memories, Cambridge, MA: MIT Press, pp. 13-46.

[10] Rodden, K., and Wood, K.R., "How Do People Manage Their Digital Photographs?", Proceedings of CHI 2003 (Ft. Lauderdale, FL, April 2003), pp 409-416.

[11] Spence, R. (2002) "Rapid, Serial, and Visual: A Visual Presentation Technique with Potential", Information Visualization, Vol. 1, No. 1, pp. 13-19.

[12] Tse, T., Marchionini, G., Ding, W., Slaughter, L. and Komlodi, A. (1998) "Dynamic Keyframe Presentation Techniques for Augmenting Video Browsing", in Proc. of Advanced Visual Interfaces (AVI-98)

[13] Witkowski, M., Arafa, Y. and de Bruijn, O. (2001) "Evaluating User Reaction to Character Agent Mediated Displays Using Eye-tracking Technology", in Proc. AISB'01 Sym. On Information Agents for Electronic Commerce, pp. 79-87.

[14] Witkowski, M., Neville, B. and Pitt, J. (2003) "Agent Mediated Retailing in the Connected Local Community", Interacting with Computers 15, pp. 5-32.

[15] Wittenburg, K., Forlines, C., Lanning, T., Esenther, A., Harada, S., Miyachi, T., (2003) "Rapid Serial Visual Presentation Techniques for Consumer Digital Video Devices", ACM Symposium on User Interface Software and Technology (UIST), pp. 115-124.

[16] Wittenburg, K., Chiyoda, C., Heinrichs, M. and Lanning., T. (2000) "Browsing through rapid-fire imaging: requirements and industry initiatives", in Proceedings of Electronic Imaging 2000, San Jose, CA.

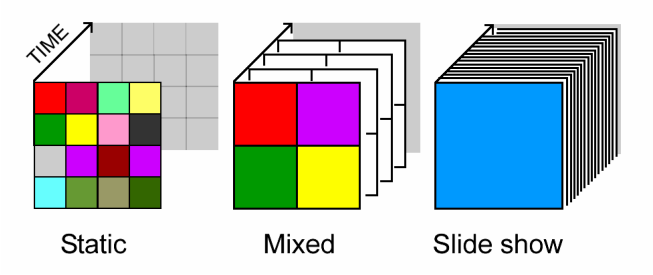

Figure 1. A schematic representation of the three presentation modes investigated.

Figure 2. Examples of images used in the experiments.

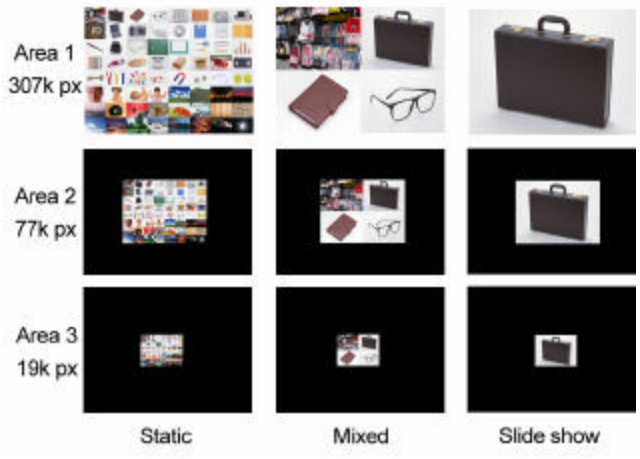

Figure 3. Presentation mode/display area combinations used in the experiments.

Figure 12. Eye-gaze pattern, subject 5, slide-show mode, 3200ms.

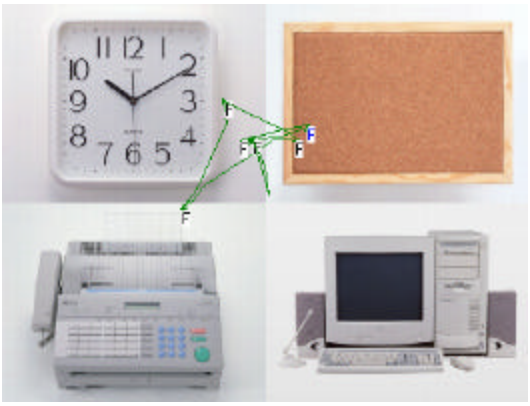

Figure 13. Eye-gaze pattern, subject 5, mixed mode, 3200ms.

Figure 14. Eye-gaze pattern, subject 5, static mode, 3200ms