# An Evaluation of Interaction Methods for Controlling RSVP Displays in Visual Search Tasks

Jamie Waese
Cell and Systems Biology
University of Toronto
Toronto, Canada

Wolfgang Stuerzlinger
Interactive Arts + Technology
Simon Fraser University
Surrey, Canada

Nicholas J. Provart
Cell and Systems Biology
University of Toronto
Toronto, Canada

***Abstract*— Accurately identifying images with subtly varying features from a large set of similar images can be a challenging task. To succeed, viewers must perceive subtle differences between multiple nearly identical images and react appropriately. The Rapid Serial Visual Presentation (RSVP) display technique has the potential to improve performance as it exploits our ability to preattentively recognize differences between images when they are flashed on a screen in a rapid and serial manner. We compared the speed and accuracy of three RSVP interface methods ("Hover", "Slide Show" and "Velocity") against a traditional "Point & Click" non-RSVP interface to test whether an RSVP display improves performance in visual search tasks. In a follow-up study we compared "Hover" and "Velocity" RSVP interface methods against a "Small Multiples" non-RSVP interface to explore the interaction of interface type and target size on visual search tasks. We found the "Hover" RSVP interface to significantly reduce the time it takes to perform visual search tasks with no reduction in accuracy, regardless of the size of the search targets. Beyond the gene identification task tested here, these experiments inform the design of user interfaces for many other visual search tasks.**

***Keywords*— *Rapid Serial Visual Presentation (RSVP), visual search, user interface design, preattentive processing***

## I. Introduction

Flipping through a set of images to find one with a desired pattern is a relatively common visual search task. Biologists often do this with electronic fluorescent pictographs (eFPs), a widely used data visualization technique for identifying gene expression patterns in plants (Fig. 1). eFP images display gene transcript levels on a diagram of the plant, with low expression levels coloured yellow and high expression levels coloured red [1]. They enable visual exploration of big data sets with each image summarizing up to 300 samples from databases of 30+ million records. In order to identify genes that contribute to a particular biological function, researchers create eFP images for several genes of interest, visually scan them, then select the one that best satisfies their search criteria [2]. Some researchers may look for genes with the maximum expression level overall, while others are more interested in specificity of an expression pattern, which involves looking at the overall visual pattern in each eFP image and deciding if it matches what they are looking for.

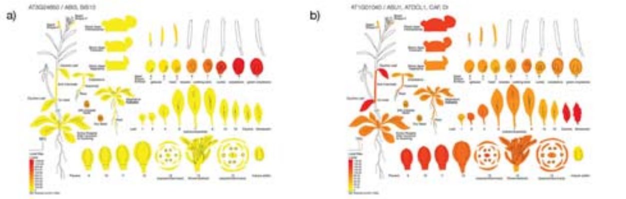


Fig 1. eFP images showing the expression patterns of two *Arabidopsis thaliana* genes. a) ABI3's expression is localized to mature seeds. b) ASU1 has a broad expression pattern, with a visible absence in the mature pollen.

Consider the cognitive processes involved in a similar spot-the-differences task (Fig. 2). Information retrieved by rapid movements of the eye must be stored in working memory when switching gaze between the images and within features of the images [3]. The viewer must use "top down" processes to direct his/her gaze and attention in order to judge unique vs. non-unique features between the two images [4], [5]. This information processing comes at a cost in time and resources.

The proximal arrangement of images in Fig. 1 is sometimes referred to as "small multiples", a data presentation technique popularized by Edward Tufte [6] for comparing simple images, or identifying images with unique categorical features. However, due to the challenge of recognizing subtle differences between complex images, a small multiples display is not conducive for visual search tasks that include a variety of small features. Most critically, with an increasing number of stimuli the (already small) images get smaller and smaller.

A faster way to perceive differences between similar images is to align one on top of the other and then rapidly flip between them. This takes advantage of our ability for preattentive visual processing as the differences between the two images will appear to flicker. Typically, all processing of visual information that is performed in less than 250 ms is termed preattentive [7]. A display method that takes advantage of preattentive processing could reduce the amount of time it takes researchers

Fig 2. Spot the differences between two similar images that are displayed side by side.

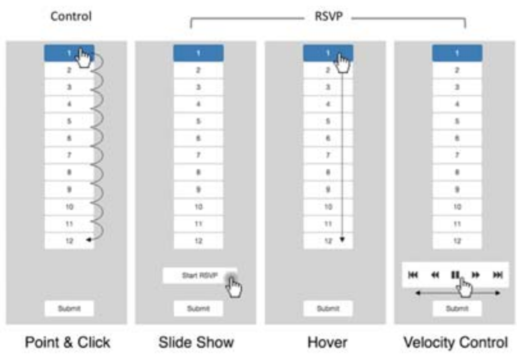


Fig 3. Four control panels for viewing and selecting images.

to identify genes of interest, but it could also be relevant for other domains. Carl Pulfrich developed a similar method, the blink comparator, in 1904 to detect the movement of asteroids and comets in sequential photographs of the night sky [8].

Rapid Serial Visual Presentation (RSVP) is a technique that displays a large amount of visual information in a limited space by rapidly presenting a set of images as a sequence over time [9]. Each image is displayed for a fraction of a second, ranging from 110 to 750 ms per image [10], [11]. Several variations of RSVP have been explored for computer displays [12], different presentation modes [11] and methods of control [13]. RSVP is known for its use in reading applications but the principle can be applied to visual search tasks as well. A common example would be flipping through a photo album to find a picture of one's dog. Visual search using RSVP uses a combination of temporal juxtaposition and spatial superposition as it "*places objects to be compared in the same space so that differences can be detected as low-level visual features, i.e., blinking*."

Our literature review identified three types of user interfaces that have been used to control RSVP displays: Slide Show, Velocity, and Hover. A "Slide Show" interface rapidly flashes images in sequence at the same location. The only control available to the user is a button to start and stop the slide show and perhaps a controller to adjust the overall speed. "Velocity" interfaces are similar to the "Slide Show" method but allow the user to select the speed and direction of image playback using buttons or a slider to adjust the total time for which an image is visible. A "Hover" interface allows the user to hover the mouse over a list of options to select the image that is displayed. This is similar to video scrubbing in which the user drags a playback head over a video timeline to select the frame they wish to stop on. However, there is no need to hold the mouse button down while sliding the mouse back and forth. Currently, there is no empirical research as to which RSVP interface method is best suited for visual search tasks.

In this paper we present two studies exploring whether an RSVP display can help improve performance on visual search tasks using eFP images, and if so, which method for controlling the RSVP display is most effective. We define performance as a combination of task completion time and accuracy.

The real task of gene selection using eFP images by biologists is different from the spot-the-differences task in Fig. 2. Yet, both tasks require looking at a set of images with small local variations. Though searching for a particular eFP colour pattern is a biology-specific task, rapid image search has broad implications on the viability of RSVP as a visual search tool.

## II. Experiment One

### A. Hypothesis

H1. All three RSVP interfaces will result in faster task completion times than the non-RSVP "Point & Click" interface.

H2. Faster completion times will not compromise accuracy.

### B. Apparatus

In accordance with research that shows user testing that measures online performance in terms of time and accuracy to be valid and reliable [14]–[16], participants were asked to perform the experiment using their own personal computers from home via the Internet. An experiment website was created with Ruby-on-Rails and HTML/JavaScript and hosted on Heroku. A pilot test was conducted to ensure the website worked as anticipated. This website randomly assigned participants to one of four groups, delivered a series of trials according to the condition a participant was assigned to, and stored all responses and time logs in an online database. Participants were instructed to log in to the experiment website using a laptop or desktop computer. Touch screen devices were not permitted. The website was designed to fit in a window 1024 pixels in width and 640 pixels in height to fit laptop screens without scaling. The experiment website is at: http://bar.utoronto.ca/userTesting

### C. Participants

Eighty-one volunteers, 19-29 years old, participated in the experiment. Forty-one of them were female. All were enrolled in a third year undergraduate university course in Bioinformatics Methods and had prior experience identifying gene expression levels using eFP images. Participants received a course bonus point for taking part in the study.

### D. Experimental Design

Participants were assigned to one of four groups using block randomization to ensure similar group sizes. Each group was given a different user interface for controlling the speed and direction of RSVP image advancement in a display window.

The left side contained one of four possible control panels (as depicted in Fig. 3) and the right side displayed the sequence of images. Participants were instructed to select an eFP image that was reddest in a particular area from a set of twelve possible answers. Each participant responded to ten such trials, with each group viewing the same set of images, numbered in the same order, with the same order of questions. While this task is simpler than the task that fully trained biologists perform, it matches the preattentive criterion and was simple enough that all participants easily understood and were able to perform the task.

This experiment had one independent variable with four levels. It had a between-subject design with each participant only being exposed to a single user interface. One condition served as our control group, a non-RSVP display method that required users to "Point & Click" on the button associated with the image they wished to display. The remaining three groups were "Slide Show", "Hover" and "Velocity" depending on the RSVP interface participants were assigned to.

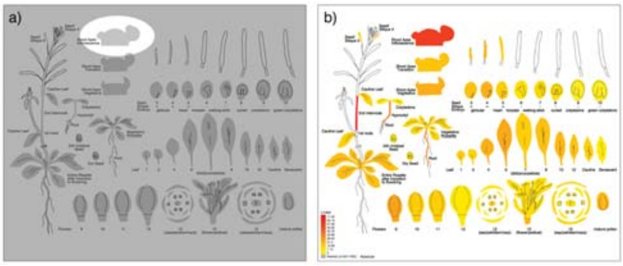


Fig 4. (a) Priming image for a guided visual search. (b) The eFP image that is reddest in the area that was highlighted in the priming image.

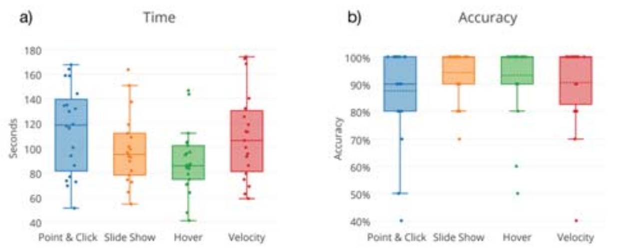


Fig 5. Time and accuracy results for each condition. Box plot whiskers indicate highest and lowest data points within 1.5 x IQR range.

Time to complete each trial and accuracy of the answers were measured for each participant. Time was measured in milliseconds. The experiment was designed to be completed in one sitting, and lasted roughly three minutes per participant.

### *E. Tasks and Procedures*

Participants were asked to answer ten visual search questions that involved selecting one target image from a set of twelve eFP images. The eFP images were randomly assigned a position in the sequence using a simple randomization script. All participants viewed the sequence and answered the questions in the same order. Each question consisted of a priming image (Fig. 4a) that highlighted which area of the screen participants should focus on, and the instruction: "Select the image on the following page that has the most red in the area that is highlighted below." Fig. 4b shows one of the eFP images participants had to select from. Thus, it was a guided visual search task [17] in which the participant selects one target image from a set of eleven distractor images based on differences in colour using a combination of top-down and bottom-up processing.

For each interface, participants had to click the 'Submit' button to indicate their selected eFP image. The rest of the user interface was modified for each group as follows:

- **"Point & Click":** The user selects which eFP image is displayed by clicking on the numbered buttons.
- **"Slide Show":** Pressing the 'Start RSVP' button at the bottom of the panel initiates the "Slide Show" RSVP display method, automatically changing images serially every 200 milliseconds, as described in [12]. After displaying the last image in the set, the cycle is restarted. The 'Stop RSVP' button stops the automatic progression at the current image, and the user can restart it again.
- **"Hover":** This group was similar to the "Point & Click" group but instead of clicking the numbered buttons to move on to the next eFP image, the participant only needs to hover their mouse pointer over the buttons. Moving the mouse up and down over the list quickly produces a *user controlled* RSVP effect.
- **"Velocity":** This group was similar to the "Slide Show" group except that participants could control the speed and direction of the slide show by hovering over the VCR-like velocity control buttons at the bottom of the panel. The center button paused the display. Regular RSVP speed was 250 ms and fast 125 ms.

### *F. Data Collection and Measures*

As mentioned above, all data was collected through a web site that participants accessed from their own homes. The pre-test questionnaire data was examined to rule out potential biases such as gender, colour insensitivity, and participants' prior exposure to reading eFP images.

Due to a participant's early withdrawal, one incomplete data set was excluded from the study. There were 20 people in the "Point & Click" control group, 19 in the "Slide Show" group, 22 in the "Hover" group and 19 in the "Velocity" group.

We did not include the first two trials in the time analyses because a scatter plot of the results clearly showed that participants were slower on the first two trials than the remaining eight, suggesting that they were still learning the task. Since participants were not supervised during the experiment, we were concerned about the potential for external interruptions possibly influencing the data. To prevent outliers from biasing our results, we used John Tukey's method of identifying outliers. For each group, we computed its first and third quartile (Q1 resp. Q3) and interquartile range (IQR). We then replaced data points that were less than Q1 – 1.5IQR and more than Q3 + 1.5IQR with the mean of that participant's other responses. Through this method, ten of our 640 trials (1.56%) were filtered. One was from the "Point & Click" group, two were from the "Slide Show" group, three from the "Hover" group, and four from the "Velocity" group. The accuracy data was not filtered for outliers.

### *G. Results: Time*

Fig. 5a shows the completion times for each group. The mean time per participant for the "Point & Click" group was 114.9 seconds (SD 35.3), the "Slide Show" group was 97.2 seconds (SD 29.8), the "Hover" group was 89.0 seconds (SD 25.2), and the "Velocity" group was 109.6 seconds (SD 35.7). The data met preconditions for ANOVA. The effect of user interface method on task completion time was statistically significant ($F_{3, 76} = 2.86$, $p < 0.05$).

A Bonferroni and Holm post-hoc test identified a significant difference between the "Point & Click" group and "Hover" ($T = 2.65$, $p < 0.05$). The "Slide Show" and "Velocity" interfaces were not significantly different from the "Point & Click" group, nor from each other.

### *H. Results: Accuracy*

Fig. 5b shows the accuracy scores for each group. The histogram of accuracy scores per trial indicated that accuracy was not impacted by learning effects so all ten trials were included in the accuracy analyses. For each participant, a value of 1 was added for each correct answer and 0 for each incorrect one. Thus a score of 10 is equivalent to 100% accuracy. The mean accuracy for all four conditions was 91.3% (SD 14.2). The "Point & Click" group had a mean accuracy of 87.5% (SD 17.1), the "Slide Show" 94.2% (SD 9.0), the "Hover" 93.2% (SD 13.6), and the "Velocity" group had a mean accuracy of 90.5% (SD 15.5). The data is negatively skewed with the percentage of participants who had 100% correct answers by group as follows: "Point & Click" 45%, "Slide Show" 63%, "Hover" 68%, and "Velocity" 57%. The data met other preconditions for ANOVA. Analysis of variance for the four conditions showed no significant effects ($F_{3,76} = 0.8970$, ns).

*Results: Input Device*

44.2% of participants used a mouse while the rest used a trackpad. Analysis of variance showed no significant difference of this ratio across the four groups ($F_{3,76} = 0.3295$, ns) so we did not explore the potential impact of input device further.

### I. *Results: User Satisfaction*

In a post-test questionnaire, many participants reported that the "Slide Show" interface was difficult to use because the auto advance was too fast. The "Hover" group received the most positive feedback. Participants did not report any difficulties specific to using RSVP for visual search tasks.

### J. *Discussion*

H1 is partially supported by the results in that the "Hover" method proved to be superior to the "Point & Click" control group. Thus, there is some evidence that an RSVP display can improve performance on visual search tasks. However, the other RSVP conditions ("Slide Show" and "Velocity") performed similarly to the non-RSVP "Point & Click" control group.

We speculate that the "Hover" interface had the best performance because it combines the benefits of an RSVP display with a method of control through gross motor actions. Instead of manipulating the RSVP display by triggering the automatic playback of images, the "Hover" method makes a direct connection between the numbered buttons and their associated images. Additionally, hover controls can be used without participants having to make eye saccades to find the exact position of the buttons. This could account for some of the performance increase over the other methods tested, since such saccades take time and might change the minimal memory strategies employed for the task. Eye-tracking could test this hypothesis, but clicking on buttons is subject to Fitts' law, while hovering does not require looking at the panel. Also, with the "Slide Show" speed set at 200 ms/image, there is an inherent potential for overshoot, due to the 200-350 ms delay between making a decision and its execution [18].

Our hypothesis that *all* the RSVP interfaces would perform better than the "Point & Click" interface was not confirmed. Perhaps this is because the "Slide Show" and "Velocity" interfaces rely on the user understanding how the stop/start or velocity control buttons had an ongoing effect without further action after the initial click. It is possible that variations of these methods could lead to better performance. We speculate that these interfaces are a less intuitive mapping for our context than direct interaction with the numbered buttons.

Regarding H2, none of the interfaces we tested showed any significant effects for accuracy. Simply put, there was no trade-off between faster completion times and accuracy scores. While the data were negatively skewed, ANOVA can tolerate data that is non-normal with only a small effect on the Type I error rate [19]. Group differences might appear with a more difficult task, such as identifying a particular mid-range shade instead of a maximum, but gene identification tasks rarely call for this.

Collecting data remotely raises methodological concerns since random contextual factors cannot be fully accounted for outside of a lab. However as noted in [12], *"any significant effects observed in an online experiment would be far more generalizable than those obtained in a controlled laboratory setting, considering that these would have been detected amidst the noise of a diversity of subjects and implementations."* Also, the relatively low standard deviation within each of the groups increases our confidence that the differences between the groups must be due to the conditions themselves.

## III. Experiment Two

We ran a second experiment in a controlled setting to confirm and expand results from study one. We also aimed to establish if a "Hover" interface remains effective for a larger image set and to compare it against a modified "Velocity" interface that includes "step forward/backward" buttons for fine-tuning, which might improve performance. There is an inherent delay between deciding to stop a RSVP display and actually stopping it, due to human reaction speeds. Finally, we investigate if the advantages of RSVP display can be attributed to preattentive processing, which is one of our underlying assumptions. If so, RSVP displays should not be impacted by the assigned search target size, as they are predicated on a colour shift in the region participants have been primed to focus on. In comparison, a "Small Multiples" (non-RSVP) display may work well (or even better in some instances) with large targets, but it should be less effective when the search targets are small because preattentive processing cannot be used. We dropped the "Point & Click" and "Slide Show" conditions as they do not scale to larger image sets and participants did not prefer them.

### A. *Hypothesis*

H1. RSVP interface performance time and accuracy scores will not be affected by search target size.

H2. A "Small Multiples" interface will perform worse than the RSVP interfaces on small search targets.

H3. Faster RSVP completion times will not compromise accuracy.

### B. *Apparatus*

Data was collected in a computer lab with 40 desktop computers that all used a mouse pointer device, i.e., a reasonably controlled setting. An experiment website was created with PHP and HTML/JavaScript and hosted on a university server. A pilot test was conducted to ensure the website worked as anticipated. The website randomly assigned participants to one of three conditions, delivered a series of trials, and stored all responses and time logs in an online database. Participants were instructed to log in to the website during a class session. The experiment website can be seen at: http://bar.utoronto.ca/RSVP

### C. Participants

A separate cohort of 90 volunteers, 19-25 years old, participated in the experiment. 52 of them were female. All were enrolled in a third year undergraduate university course in Bioinformatics Methods. Participants received a course bonus point for taking part in the study.

### D. Experimental Design

Participants were assigned to one of three groups using block randomization to ensure similar group sizes. The first two groups were given "Hover" (Fig. 6a) and "Velocity" (Fig. 6b) interfaces for controlling the speed and direction of RSVP image advancement. The third group was given a "Small Multiples" interface for selecting images (Fig. 7b). Before each of the ten trials, participants were given a priming image instructing them to select the image that was reddest in a particular area (Fig. 7a). Questions were randomized so answers could not be shared with other participants. Questions alternated between spatially large and spatially small search targets so practice effects should apply to both groups equally. For each participant, we measured time and accuracy for all large target trials, and time and accuracy for all small target trials. The experiment lasted approximately three minutes including time to review the next priming image.

### E. Tasks and Procedures

Participants were asked to perform ten visual search tasks on a set of thirty-six images. To broaden the applicability of our findings, we did not use eFP images in this experiment. Rather, synthetic images with spatially large, medium and small target shapes were created with Processing. Each image had a single red target and used a random assignment of yellow to orange shades for all the other shapes. Images were numbered randomly from one to thirty-six and participants viewed the sequence in the same order. Tasks were selected randomly from a subset of images where a single large or small shape was highlighted in red. Odd numbered tasks required searching for images with a large red target, while even numbered tasks required searching for images with a small red target. Each question consisted of a priming image that indicated which area of the screen participants should focus on, and the instruction: "On the next page, select the image that has the most red in the area that is highlighted below." The user interfaces for the three groups were as follows:

- **"Hover":** Moving the mouse up and down over a column of buttons quickly produces a user controlled RSVP effect. To select an image, participants must click the button with the image number they are hovering over.
- **"Velocity":** Dragging a horizontal slider back and forth adjusts the forward or backward speed of the RSVP display. Letting go of the slider handle makes it snap back to center, causing the RSVP display to immediately stop. The fastest RSVP speed in either direction was 50 ms and the slowest was 1000 ms. Step-forward and step-backward buttons underneath the slider could be used to advance/back up the display one image at a time. To select an image, participants must stop on the image they want and then press the "Submit" button.
- **"Small Multiples":** Thumbnails of all thirty-six images in the set were displayed on a 6x6 table, each occupying just over one square inch of screen space. To select an image, participants must click the thumbnail that meets their search criteria.

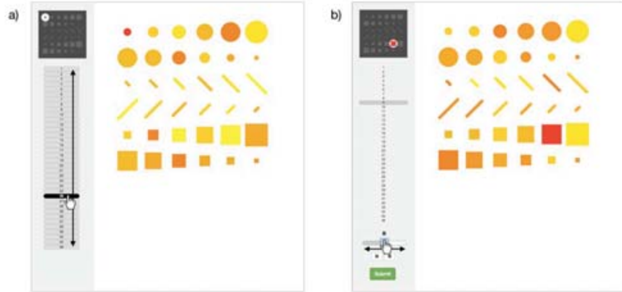


Fig 6. (a) The "Hover" interface with an image containing a small red target selected (i.e., the red circle in the top left corner). b) The "Velocity" interface with an image containing a large red target selected (i.e, the red square in the fifth row).

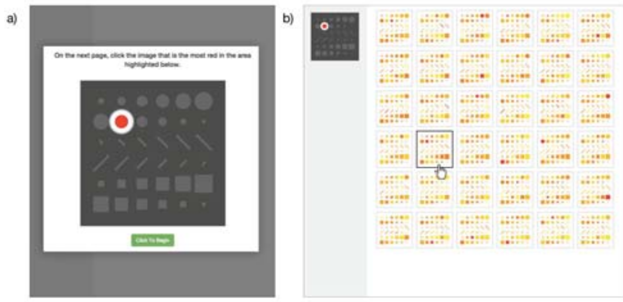


Fig 7. a) A priming image describing which of the thirty six images participants must search for. b) The "Small Multiples" interface with the correct image selected.

### F. Data Collection and Measures

Data were collected through a web site, accessed from a university computer lab during two separate class sessions, supervised by two teaching assistants and an instructor. Pre-test data was examined to rule out potential bias from age, gender and colour insensitivity.

One set of incomplete data, due to one participant's early withdrawal, was excluded from the study. There were 30 people in the "Hover" group, 29 in the "Velocity" group, and 30 in the "Small Multiples" group.

A scatter plot of the results suggested participants were still learning the task during the first two trials. To verify this, a series of paired samples t-tests were conducted to confirm whether there was a trial/learning effect for both response times and accuracy. To not confound the potential impact of trial/learning and target size, large-to-large and small-to-small trials with increasing numbers (i.e., trials 1-3, 2-4, 3–5, 4-6, 5–7, 6-8, 7-9 and 8-10) were compared to test whether performance changed over time. For response time, there was a significant drop from trial one to trial three, $t(88) = 6.23$, $p<.001$, as well as from trial two to four, $t(88) = 2.60$, $p < .05$. There were no significant differences between the response times of matching trial pairs

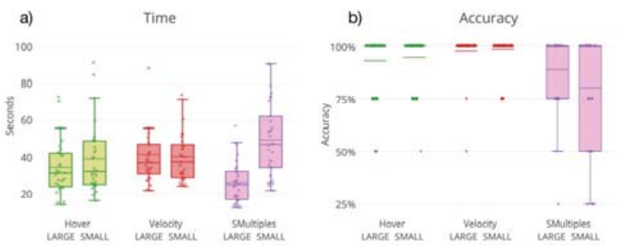


Fig 8. Time and accuracy results for each condition. Box plot whiskers indicate highest and lowest data points within 1.5 x IQR range.

beyond the second trial. For accuracy, scores improved significantly from trial one to three t(88) = -5.78, $p < .001$ but remained stable across all subsequent trials. Given this pattern, trials one and two (i.e., the first trial of each size) were dropped from subsequent analyses.

To rule out the potential of outliers biasing the analyses, the Tukey method was again used to identify participants with outlying response times. Two participants from the "Hover" group, four from the "Velocity" group, and two from the "Small Multiples" group were identified. Closer examination of their data, excluding the initial trials considered as training, identified 8 response times (1.1% of 712 trials) that were two standard deviations above the mean of all other trials by that participant for that target size. These were replaced with that participant's mean for their other trials for that target size. Analysis of this filtered data was conducted and showed an identical patterns of results. Since the outlier removal had no effect on the results, raw (unfiltered) data is reported for the analyses below.

### G. Results: Time

Fig. 8a shows the completion times for each group. For large targets, the mean time for the "Hover" group was 34.4 seconds (SD 14.4), for "Velocity" 41.5 seconds (SD 18.5), and for "Small Multiples" 26.5 seconds (SD 11.4). For small targets, the mean time for the "Hover" group was 39.0 seconds (SD 18.6), for "Velocity" 40.2 seconds (SD 13.8), and for "Small Multiples" 49.7 seconds (SD 20.1).

A two (target size: "Large", "Small") by three (interface type: "Hover", "Velocity", "Small Multiple") factor mixed analysis of variance (ANOVA) was conducted. The target size factor was within subjects while the interface type factor was between subjects. The data meets preconditions for mixed ANOVA. Levene's test of equality of error variances was not significant for target size. Box's test of equality of covariance matrices was significant ($F_{6,\,183031} = .2.58$, $p < .05$, M = 16.03), but this is a very sensitive test and only p values lower than .001 are of concern [20]. The mixed ANOVA results were significant for target size ($F_{1,\,86} = 21.76$, $p < .001$), and the interaction of interface type and target size ($F_{2,\,86} = 10.76$, $p < .001$). The interface type factor was not significant on its own.

Based on this pattern of results, our hypotheses were tested using a series of independent sample t-tests. This identified significant differences between "Small Multiples: Large" and "Hover: Large" t(58) = -2.36, $p < .05$; "Small Multiples: Small" and "Hover: Small" t(58) = 2.15, $p < .05$; "Small Multiples: Large" and "Velocity: Large" t(57) = -3.742, $p < .001$; and "Small Multiples: Small" and "Velocity: Small" t(57) = 2.133, $p < 0.05$. No significant differences were found between "Hover: Large" and "Velocity: Large", or "Hover: Small" and "Velocity: Small".

Paired samples t-tests found significant differences between "Small Multiples: Large" and "Small Multiples: Small" t(29) = -6.175, $p < .001$. No significant differences were found between "Hover: Large" and "Hover: Small", or "Velocity: Large" and "Velocity: Small".

These results indicate that size only matters for the "Small Multiples" group. "Hover" and "Velocity" are comparable in terms of response times. When search targets are large, both have longer response times than "Small Multiples". When search targets are small, both have shorter response times than "Small Multiples".

### H. Results: Accuracy

Fig. 8b shows the accuracy for each group. After removing the first two training trials, there were four trials in which participants had to search for images with large targets, and four with small targets. Each correct answer was assigned a value of 1, and each incorrect answer was assigned a value of 0. Thus, for each of the two target size conditions, a score of 4 is equivalent to 100% accuracy. For large targets, the mean accuracy for the "Hover" group was 92.5% (SD 19.9), for "Velocity" 97.4% (SD 10.2), and for "Small Multiples" 88.3% (SD 18.3). For small targets, the mean accuracy for the "Hover" group was 95.0% (SD 12.6), for "Velocity" 98.3% (SD 6.5), and for "Small Multiples" 80.0% (SD 26.6).

A two (target size: "Large", "Small") by three (interface type: "Hover", "Velocity", "Small Multiple") factor mixed analysis of variance (ANOVA) was conducted. The target size factor was within subjects while the interface type factor was between subjects. ANOVA results show a significant effect for interface type, ($F_{2,\,86} = 7.11$, $p < .001$) and the interaction of interface type and target size ($F_{2,\,86} = 3.79$, $p < .05$). The target size factor was not significant on its own.

With regards to meeting preconditions for mixed ANOVA, Levene's test of equality of error variances was significant for target size ($F_{2,86} = 8.31$, $p < .001$), and Box's test of equality of covariance matrices was also significant ($F_{6,183031} = 9.53$, $p < .001$). This indication of non-normally distributed data could be attributed to a substantial ceiling effect as more than half of all participants received an accuracy score of 100%. The percent of participants who had 100% accuracy by group is as follows: "Hover: Large" 76.66%, "Hover: Small" 80%, "Velocity: Large" 93.10%, "Velocity: Small" 93.10%, "Small Multiples: Large" 63.33%, and "Small Multiples: Small" 56.66%. Because the data were skewed in this way, we also ran the analyses using Generalized Estimating Equations (GEE), a semi-parametric statistical technique that is appropriate for a wide variety of variable distributions (normal and skewed, continuous, dichotomous, ordinal, etc.). Results of the GEE and pairwise comparisons replicated the pattern of results we found using the mixed ANOVA and post hoc t-tests. To keep reporting of results consistent across dependent variables, the mixed ANOVA and t-test results are presented here.

Independent samples t-tests found significant differences between "Small Multiples: Large" and "Velocity: Large" t(57)

= -2.35, $p < .05$; "Small Multiples: Small" and "Velocity: Small" $t(57) = -3.60$, $p < .001$; and "Small Multiples: Small" and "Hover: Small" $t(58) = -2.64$, $p < .005$. No significant differences were found between "Small Multiples: Large" and "Hover: Large", "Hover: Large" and "Velocity: Large", or "Hover: Small" and "Velocity: Small".

Paired samples t-tests found significant differences between "Small Multiples: Large" and "Small Multiples: Small" $t(29) = 2.76$, $p < .01$. No significant differences were found between "Hover: Large" and "Hover: Small", or "Velocity: Large" and "Velocity: Small".

These results indicate that, once again, size only matters for the "Small Multiples" group. Overall, "Small Multiples" was less accurate than "Hover" and "Velocity" (the exception to this was that "Hover: Large" and "Small Multiples: Large" were not significantly different in terms of accuracy). The "Hover" and "Velocity" interfaces have similar accuracy rates, regardless of search target size.

### I. *Discussion*

H1 was supported by the results. As we expected, using an RSVP display for this sort of visual search task takes advantage of preattentive colour processing within the target area participants are primed to focus on. Since even small search targets can exhibit strong colour shifts, performance times and accuracy rates are not negatively affected by size when participants use an RSVP display.

H2 was also supported by the results. The "Small Multiples" display was clearly the fastest interface for finding images with large targets but it was significantly slower than "Hover" and "Velocity" for finding images with small targets. With regards to accuracy, the "Small Multiples" display performed significantly worse than the RSVP interfaces for both large and small targets. Despite its speed advantages for large targets, its accuracy scores were simply too low to be relied on for important tasks such as gene identification.

Please note that this experiment only used thirty-six images. Whereas an RSVP display could potentially scale up to several hundred images without fundamentally altering the user interface, more images on a "Small Multiples" display would require shrinking the thumbnails down even further. At some point, the small targets would be too small to be usable.

Finally, with regards to H3, there was no trade-off between faster completion times and accuracy scores. Although the RSVP conditions did not perform faster than the "Small Multiples" condition with large targets, the faster completion times of the "Hover" and "Velocity" conditions did not compromise accuracy when searching for small targets.

## IV. General Discussion and Limitations

Our experiments explore the trade-off between target size and colour discrimination to inform the design of user interfaces for visual search tasks. Although the gene identification and primitive shape tasks studied here are very specific, the results may well generalize to other domains that require visual search across many images. Examples include MRI and sonographic images, biological samples, satellite imagery, climate maps, and virtually any data chart presented in the form of a heat map.

Broader applications may include identifying images in a museum archive, or items for sale on a shopping website, however these may require more semantic interpretation which does not necessarily take advantage of preattentive processing. This would be akin to a familiar visual search task such as "rifling" through a book to see if it contains an image of the Mona Lisa [21]. Priming may play an additional role by indicating which part of the image viewers should focus on, but it is the preattentive recognition of a specific colour that viewers are looking for. Whether the underlying processes can rightfully be described as "preattentive" or not, any tool that helps users quickly select a target image from a large set of distractor images could be useful for big data visual analytics.

We did not explore alternate interface designs that involve explicit encoding, search widgets, image cropping, image folding or video scrubbing, e.g., [22]–[24], because visual search tasks like gene identification require Gestalt-like judgements, such as finding the maximum or minimum in one or more tissue samples with no irregular patterns across all other samples. Such queries can be easily and quickly perceived and judged visually, but are hard to express in a widget-based user interface. Since there is no inherent order to a typical set of eFP images, interfaces that presuppose a (partially) ordered sequence would not work. A custom sorting tool for the underlying eFP data ignores the broader scope of more "Gestalt-like" tasks around biological data and ignores other benefits of the visual nature of the eFP images. Also, we did not explore the minimum or most natural RSVP exposure time, or how long RSVP image sequences can become, as these are separate research questions. Finally, our studies focus on search tasks where the target is indicated through differences in colour – other changes in visual channels such as shape or size are beyond our scope.

Differences between the interfaces may have impacted participant reaction time differently. In the "Slide Show" and both "Velocity" conditions, stopping on an image that has been preattentively identified requires an intentional action – either pressing the "stop" button or letting go of the velocity control handle. Instead of stopping on the image that they wanted, many participants stopped on the first or second image following it, requiring additional search to reach the desired image. Here, the "Velocity" interface in the second experiment was a big improvement over the "Velocity" interface in the first because it included "step forward" and "step backward" buttons, which participants used an average of 19 times per question (SD 11 for large targets, SD 9 for small targets). Note that this is fewer than the 36 images in the set, so participants were clearly using the velocity slider and buttons in combination.

The accuracy analyses may have been compromised by the ceiling effect of the tasks. While this resulted in our data being skewed, Glass et al. point out, *"The relevant question is not whether ANOVA assumptions are met exactly, but rather whether the plausible violations of the assumptions have serious consequences on the validity of probability statements based on the standard assumptions."* [25]. The face validity of our results, as well as the fact that they were replicated using GEE in the second experiment, strengthens our confidence in our findings. Future studies can build on this with more challenging tasks.

978-1-5090-5272-1/16/$31.00 ©2016 IEEE

This study is by no means an exhaustive analysis of which visual search method or RSVP interface is "optimal". In future studies we plan to examine whether a "Scroll Wheel", "Touch Screen", or "Keyboard Buttons" offer any improvement over the methods we tested, and whether RSVP search is in fact more efficient than text based search or task-specific filtering or aggregating methods. Further, we did not attempt to explore image comparison problems beyond the scope of visualizations in the same spatial frame. RSVP can improve gene identification using eFP images, but whether they are appropriate for general image comparison is a separate question.

The tasks evaluated here serve as a proxy for more realistic tasks in biology that require visual analysis of multiple features and evaluation of associated trade-offs. Yet, performing such tasks in any realistic way requires (much) more training. In our second experiment we used generic images with primitive shapes in order to broaden the applicability of our findings. Based on the fact that the findings are similar, we believe that our results should generalize to other domains.

## V. Conclusion

We presented experimental work that evaluates the effects of three different RSVP interaction methods and two non-RSVP interaction methods on time and accuracy for completing a visual search task. Our results indicate that such tasks can indeed benefit from an RSVP display. Among the different RSVP interfaces we tested, "Hover" proved to be the most promising method for identifying target images quickly and accurately, with the "Velocity" method also performing well as long as it includes "step forward" and "step backward" buttons.

What are the implications for visualization practitioners? "Point & Click" and "Small Multiples" methods are de facto standards among window-based operating systems. They are frequently used for visual search tasks (e.g., searching for a photograph in a file browser either with a preview panel or using thumbnail images). Our findings suggest that RSVP may offer improvements over these methods for certain tasks, and thus it should be considered as an alternative approach – especially in combination with a "Hover"-style interface.

While gene identification using eFP images is an example of big data visual search, further work must be done to determine how a "Hover" interface might be scaled to more than a few hundreds of images. Beyond the tasks investigated here, we believe that RSVP displays can aid visual search tasks across multiple domains, such as climate research, medical imaging, online shopping, photo cataloguing, and file management.

## VI. Acknowledgments

We thank Olivier St-Cyr, Simon Cook, Hubert Hu, Xiwen Zhou and Pan Zhang for their help with the first study. We also thank hellokids.com for the illustration in Fig. 1. This work received University of Toronto Research Ethics Board approval under protocol reference number 32629.

## References

[1] D. Winter, B. Vinegar, H. Nahal, R. Ammar, G. V. Wilson, and N. J. Provart, "An 'Electronic Fluorescent Pictograph' Browser for Exploring and Analyzing Large-Scale Biological Data Sets," *PLoS ONE*, vol. 2, no. 8, p. e718, Aug. 2007.

[2] R. S. Austin, D. Vidaurre, G. Stamatiou, R. Breit, N. J. Provart, D. Bonetta, J. Zhang, P. Fung, Y. Gong, P. W. Wang, P. McCourt, and D. S. Guttman, "Next-generation mapping of Arabidopsis genes," *Plant J.*, vol. 67, no. 4, pp. 715–725, Aug. 2011.

[3] R. M. McPeek, V. Maljkovic, and K. Nakayama, "Saccades require focal attention and are facilitated by a short-term memory system," *Vision Res.*, vol. 39, no. 8, pp. 1555–1566, Apr. 1999.

[4] M. J. Bravo and K. Nakayama, "The role of attention in different visual-search tasks," *Percept. Psychophys.*, vol. 51, no. 5, pp. 465–472, May 1992.

[5] C. Ware, *Information Visualization: Perception for Design*. Elsevier, 2012.

[6] E. R. Tufte and P. R. Graves-Morris, *The visual display of quantitative information*, vol. 2. Graphics press Cheshire, CT, 1983.

[7] C. G. Healey and J. T. Enns, "Attention and visual memory in visualization and computer graphics," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 7, pp. 1170–1188, Jul. 2012.

[8] D. Hoffleit, "A History of Variable Star Astronomy to 1900 and Slightly Beyond," *J. Am. Assoc. Var. Star Obs. JAAVSO*, vol. 15, pp. 77–106, 1986.

[9] M. C. Potter and E. I. Levy, "Recognition memory for a rapid sequence of pictures," *J. Exp. Psychol.*, vol. 81, no. 1, pp. 10–15, 1969.

[10] H. Intraub, "Rapid conceptual identification of sequentially presented pictures.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 7, no. 3, pp. 604–610, 1981.

[11] R. Spence, "Rapid, Serial and Visual: A Presentation Technique with Potential," *Inf. Vis.*, vol. 1, no. 1, pp. 13–19, Mar. 2002.

[12] R. Spence and M. Witkowski, *Rapid serial visual presentation: design for cognition*. New York: Springer, 2013.

[13] T. Brinded, J. Mardell, M. Witkowski, and R. Spence, "The Effects of Image Speed and Overlap on Image Recognition," in *2011 15th International Conference on Information Visualisation (IV)*, 2011, pp. 3–11.

[14] J. R. de Leeuw and B. A. Motz, "Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task," *Behav. Res. Methods*, Mar. 2015.

[15] J. Cheng, J. Teevan, and M. S. Bernstein, "Measuring Crowdsourcing Effort with Error-Time Curves," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, New York, NY, USA, 2015, pp. 1365–1374.

[16] K. Enochson and J. Culbertson, "Collecting Psycholinguistic Response Time Data Using Amazon Mechanical Turk," *PLoS ONE*, vol. 10, no. 3, p. e0116946, Mar. 2015.

[17] J. M. Wolfe, "Guided Search 2.0 A revised model of visual search," *Psychon. Bull. Rev.*, vol. 1, no. 2, pp. 202–238, Jun. 1994.

[18] J. Fuster, *The Prefrontal Cortex*. Elsevier, 2008.

[19] "One-way ANOVA - Violations to the assumptions of this test | Laerd Statistics." [Online]. Available: https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide-3.php. [Accessed: 10-Jun-2016].

[20] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, 4th ed. New York: Allyn & Bacon, 2001.

[21] M. Witkowski and R. Spence, "Rapid serial visual presentation: An approach to design," *Inf. Vis.*, vol. 11, no. 4, pp. 301–318, Oct. 2012.

[22] J. Schmidt, M. E. Gröller, and S. Bruckner, "VAICo: visual analysis for image comparison," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2090–2099, Dec. 2013.

[23] C. Tominski, C. Forsell, and J. Johansson, "Interaction Support for Visual Comparison Inspired by Natural Behavior," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2719–2728, 2012.

[24] J. Matejka, T. Grossman, and G. Fitzmaurice, "Swifter: Improved Online Video Scrubbing," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2013, pp. 1159–1168.

[25] G. V. Glass, P. D. Peckham, and J. R. Sanders, "Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance," *Rev. Educ. Res.*, vol. 42, no. 3, pp. 237–288, 1972.