# Harnessing Preattentive Processes for Multivariate Data Visualization


Christopher G. Healey
Kellogg S. Booth
James T. Enns*

Department of Computer Science
University of British Columbia
Vancouver, British Columbia, V6T 1Z2
e-mail healey@cs.ubc.ca



## ABSTRACT

A new method for designing multivariate data visualization tools is presented. These tools allow users to perform simple tasks such as estimation, target detection, and detection of data boundaries rapidly and accurately. Our design technique is based on principles arising from an area of cognitive psychology called preattentive processing. Preattentive processing involves visual features that can be detected by the human visual system without focusing attention on particular regions in an image. Examples of preattentive features include colour, orientation, intensity, size, shape, curvature, and line length. Detection is performed very rapidly by the visual system, almost certainly using a large degree of parallelism. We studied two known preattentive features, hue and orientation. The particular question investigated is whether rapid and accurate estimation is possible using these preattentive features. Experiments that simulated displays using our preattentive visualization tool were run. Analysis of the results of the experiments showed that rapid and accurate estimation is possible with both hue and orientation. A second question, whether interaction occurs between the two features, was answered negatively. This suggests that these and perhaps other preattentive features can be used to create visualization tools which allow high-speed multivariate data analysis.



## RÉSUMÉ

Une nouvelle méthode pour le design d'outils pour la visualization de données multivariées est présentée. Ces outils permettent à l'usager de réaliser rapidement et précisément des tâches simples comme l'estimation, la détection d'une cible et la détection des limites de données. Notre technique de design est fondée sur des principles de traitement préattentif en provenance du domaine de la psychologie des connaissances. Le traitement préattentif comprend des caractéristiques visuelles qui peuvent être détectées par le système visuel humain sans porter attention sur des régions particulières d'une image. La couleur, l'orientation, l'intensité, la grosseur, la forme, la courbure et la longueur de lignes sont autant d'exemples de caractéristiques préattentives. La détection est réalisée très rapidement par le système visuel, presque certainement utilisant un haut niveau de parallélisme. Nous avons choisi deux caractéristiques préattentives connues: la teinte et l'orientation. La question particulière investiguée est s'il est possible d'obtenir des estimations rapides et précises en utilisant ces caractéristiques. Nous avons conduits des expériences qui utilisaient nos outils basés sur ces deux caractéristiques préattentives. L'analyse des resultats des expériences démontre qu'une estimation rapide et précise est possible avec la teinte et l'orientation. Une seconde question ayant trait à l'intéraction entre ces deux caractéristiques fut répondue négativement. Ceci suggère que les caractéristiques préattentives peuvent être utilisées pour créer des outils de visualization qui permettent une analyse rapide de données multivariées.


## OVERVIEW

The field of scientific visualization draws on research from a wide spectrum of traditional disciplines. These include computer science, psychology, and the visual arts. The "domain of visualization", as defined by a National Science Foundation panel on scientific computing, includes the development of specific applications, the development of general purpose tools, and the study of research problems that arise in the process [McC87]. To date, most research efforts have focused


*Department of Psychology, UBC

on visualization applications for specific problems and environments. Relatively few efforts have formulated general guidelines for the design of visualization tools.

In this paper, we utilize an area of cognitive psychology known as preattentive processing in an attempt to develop such general guidelines. First, we review a set of visualization requirements that are common to applications ranging from visual interactive simulation, to volume visualization, to multivariate data analysis. Second, we summarize the area of preattentive processing in order to reveal abilities and limitations of human cognition that are relevant to these requirements. Third, we describe a specific visualization tool we have developed, based on these general considerations, to assist oceanographers in numeric estimation problems involving salmon migration simulations. Finally, we discuss the implications of our approach, both for the specific application of numeric estimation, and for the development of general guidelines in scientific visualization.

## SCIENTIFIC VISUALIZATION

Many different disciplines such as physics, chemistry, oceanography, and management science use computer simulations to model real-world phenomena. Visual interactive simulation (VIS) is a type of computer simulation system which provides immediate visual feedback and user interaction [Bel87]. A key requirement of VIS is a visualization technique which provides an informative display of results in real-time. The technique must be computationally simple, yet must allow the user to rapidly analyse the data being displayed. Researchers use VIS tools to view their results as they are being produced. This allows them to "steer" the simulation and direct its path to follow interesting trends as the data is generated. A number of researchers who built VIS tools provide various empirical and anecdotal results that show VIS to be an improvement over existing simulation models [Mel85][Set88].

The requirements for VIS are similar to another important class of problems, the visualization of output from real-time applications. Systems like air traffic control require rapid and informative visualization of multivariate data. These displays are often shared by different operators, who visually acquire different data from different parts of the display at the same time. The visualization technique must allow a variety of tasks to be performed rapidly and accurately on dynamically changing subsets of the overall display. Medical imaging systems such as CT, MRI, and ultrasound are another type of application that could benefit from real-time visualization techniques directed by the user, who analyses the data and decides how to proceed. An informative visualization technique that allows rapid and accurate visual analysis of more than one aspect of the data would decrease the amount of time needed to complete the diagnostic task. This is important, because these types of systems often cannot be time-shared and thus any improvement in visualization would increase the throughput for the system.

One explicit goal of visualization is to present data to human observers in a way that is informative and meaningful, on the one hand, and yet intuitive and effortless on the other. This goal is often pursued by attaching "features" such as colour, spatial location, and size to each data element. Features are chosen to show properties within and relationships among data elements. This technique is used to represent high-dimensional data in a low-dimensional environment. Multivariate data visualization addresses the question "How can we display the information in a low-dimensional environment, such as a computer screen or printed media?" An ad hoc assignment of features to individual data dimensions may not result in a useful visualization tool. Indeed, too often the tool itself interferes with the user's ability to extract the desired information.

Researchers have approached the multivariate data visualization problem in different ways. Enns discusses using the human visual system to efficiently process large multivariate datasets [Enn90a]; he describes geometric icons which combine the power of the computer and the human visual system [Enn90b]. Ware and Beatty designed a method that uses colour to represent multivariate data elements [War88]; subsets of the data with similar values appear as a spatial "cloud" of similarly coloured squares. Pickett and Grinstein have been using results from cognitive psychology as a basis for design of their visualization tools [Pic88][Gri89]; they display structure in the data as a set of textures and boundaries, so that groups of data elements with similar values appear as a spatial group with a unique texture in the display.

We approached multivariate visualization by defining a set of requirements which we feel are inherent to this class of problem. Specifically, we wanted to design a visualization technique which supported:

- *shared data*, the technique should be able to display independent data values simultaneously. A single user could choose to examine various relationships, or multiple users could simultaneously

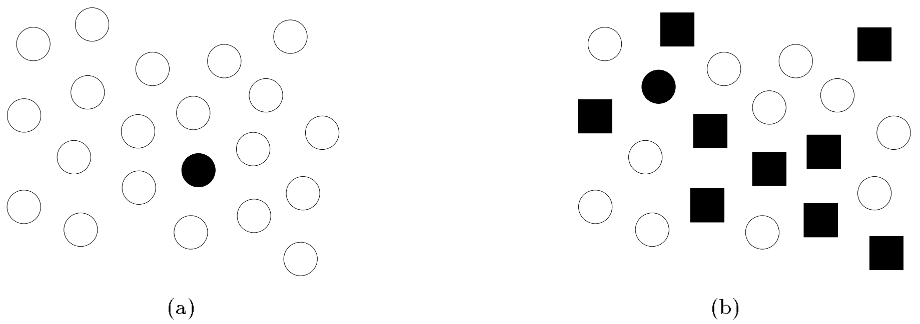


Figure 1: Examples of target detection: (a) target can be preattentively detected because it has the unique feature "filled"; (b) filled circle target cannot be preattentively detected because it has no preattentive feature unique from its distractors

examine independent data values

- *speed*, users should be able to obtain information about any of the data values quickly
- *accuracy*, information obtained by the users should accurately represent the relationship being investigated

Using an approach similar to Pickett and Grinstein, we decided to use the built-in processing of the human visual system to assist with visualization. Preattentive processing describes a set of simple visual features that are detected in parallel by the low-level human visual system. We hypothesized that the use of preattentive features in a visualization tool would allow users to perform rapid and accurate visual tasks such as grouping of similar data elements, detection of elements with a unique characteristic, and estimation of the number of elements with a given value or range of values. We tested this hypothesis using controlled psychological experiments that simulated a preattentive visualization tool. Analysis of the experiment results showed our hypothesis was true for the class of data we used. Before describing our experiments and results, we provide an introduction to preattentive processing.

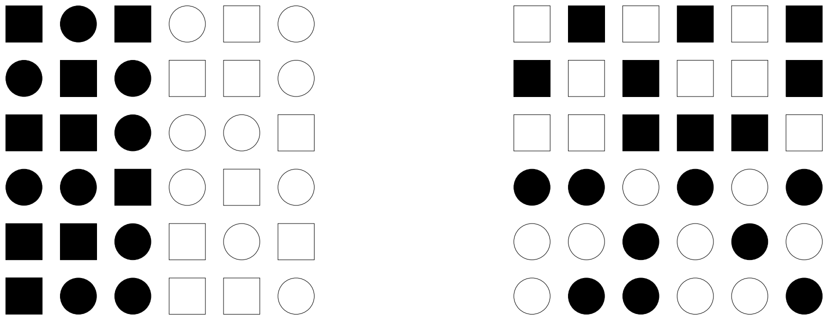
## PREATTENTIVE PROCESSING

Researchers in psychology and vision attempt to understand how the human visual system analyses images. One interesting result has been the discovery of visual properties that are "preattentively" processed. These properties are detected immediately, such that viewers do not have to focus their attention to determine whether elements with the given property are present or absent.

An example of preattentive processing is detecting a filled circle in a group of empty circles (Figure 1a). The target object has a preattentive feature "filled" that the distractor objects do not (all non-target objects are considered distractor objects). A viewer can quickly glance at the image to determine whether the target is present or absent. A conjunction occurs when the target object is made up of two or more features, each of which is contained in the distractor objects. Objects that are made up of a conjunction of unique features cannot be detected preattentively [Tri85]. Figure 1b shows an example of a conjunction target. The target is made up of two features, filled and circular. Both these features occur in the distractor objects (filled squares and empty circles). Thus, the target cannot be preattentively detected.

Properties that are preattentively processed can be used to highlight important image characteristics. Experiments in psychology by Triesman, Julész, and others have used preattentive properties to assist in performing the following visual tasks:

- *target detection*, where users attempt to rapidly and accurately detect the presence or absence of a "target" element that uses a unique preattentive feature within a field of distractor elements (Figure 1)
- *boundary detection*, where users attempt to rapidly and accurately detect a texture boundary between two groups of elements, where all the elements in each group have a common preattentive feature (Figure 2)



(a)

(b)

Figure 2: Form and hue segregation: (a) hue boundary is preattentively detected, even though form varies in both groups; (b) hue interferes with detection of form boundary

- *counting/estimation*, where users attempt to count or estimate the number of elements in a display that have a unique preattentive feature

In general, tasks which can be performed in less than 250 milliseconds are considered preattentive. Within this time frame the human visual system cannot decide to change its focus of attention. This means preattentive tasks require only "a single glance" at the image being displayed.

In addition to the tasks listed above, scientists have been examining the interaction between features within a display. Callaghan found that varying certain irrelevant features within a group can interfere with boundary detection [Cal89]. Results showed that a non-uniform hue interfered with form segregation (Figure 2b). It took subjects longer to determine where a horizontal or vertical form boundary occurred, relative to a control array where hue was held constant. However, a non-uniform form did not interfere with hue segregation (Figure 2a); a hue boundary could be detected in a fixed amount of time, regardless of whether form varied or not. Callaghan found a similar asymmetry between brightness and hue [Cal84]. Results showed that variation of brightness interfered with hue segregation. However, variation of hue did not interfere with brightness segregation.

A number of scientists have proposed competing theories to explain how preattentive processing occurs, in particular Triesman's feature integration theory [Tri85], Julész' texton theory [Jul83], and Quinlan and Humphreys' similarity theory [Qui87]. Our interest is in the use of features which have been shown to be preattentive. We examined two such features, hue and orientation, and investigated their use for a common visualization task, estimation.

## PREATTENTIVE ESTIMATION

Through experimentation, we sought to determine whether or not research in preattentive processing can help design more useful and intuitive scientific visualization tools. We addressed two specific sets of questions about preattentive features and their use in visualization tools:

- Is it possible for subjects to provide a reasonable estimation of the relative number of elements in a display with a given preattentive feature? What features allow this and under what conditions?
- How does encoding an "irrelevant" data dimension with a secondary preattentive feature interfere with a subject's estimation ability? Which features interfere with one another and which do not?

Both of these questions address the visualization requirements discussed in the previous section. Estimation is often needed for rapid and accurate analysis of visual displays. If preattentive features can be used, VIS and real-time applications could employ this technique for effective real-time visualization. Similarly, the ability to efficiently encode multiple unrelated data values in a single display would allow users to "share" the display, but only if no interference occurs. This corresponds our requirements for a potential visualization technique.

The experiments used data similar to that which occurred in a set of salmon migration simulations being run by the Department of Oceanography at the University of British Columbia [Tho92a][Tho92b]. Salmon are a well-known and economically important type of fish that live, among other areas, on the western Canadian coast. After a period of feeding and growth in the open ocean, salmon begin their migration run. This consists of an open ocean stage back to the British Columbia coast and a coastal stage back to a freshwater stream to spawn. Salmon almost always spawn in the stream where they were born. Scientists now know salmon find their stream of birth using smell when they reach the coast. The direction finding methods used to navigate from the open ocean habitat to the coast are still being researched. The simulations are studying the causal effects of ocean currents on sockeye salmon migration patterns. Results such as ocean current patterns and latitudes where each salmon arrived at the B.C. coast (latitude of landfall) were generated during the simulation. We chose to use this data to investigate the likelihood of our techniques being relevant to real-world problems and data.

We decided to examine two preattentive features, hue and orientation. This was done by running experiments which displayed data using coloured, rotated rectangles. The features hue and orientation have been shown to be preattentive in various experiments by Julész [Jul83] and Triesman [Tri85]. Two unique rotations were used: 0° rotation and 60° rotation. Two different hues, $H_1$ and $H_2$, were chosen from the Munsell colour space.

The Munsell colour space was originally proposed by Albert H. Munsell in 1898 [Bir69]. It was later revised by the Optical Society of America in 1943 to more closely approximate Munsell's desire for a functional and perceptually balanced colour system. A colour from the Munsell colour space is specified using the three "dimensions" hue, chroma, and value. Hue refers to a uniquely identifiable colour such as red, blue, or blue-green. Individual hues are further subdivided into subsections. A number before the hue specifies its subsection (e.g., 5R, 2B, or 9BG). Chroma defines a colour's strength or weakness. Greys are colours with a chroma of zero. Value refers to a colour's lightness or darkness. A Munsell colour is specified by "hue value/chroma". For example, 5R 6/6 would be a relatively strong red, while 5BG 9/2 would be a weak cyan. We chose hues which satisfied the following two properties:

- *Property 1*: the perceived brightness of both rectangles coloured using hues $H_1$ and $H_2$ was equal (isoluminence)
- *Property 2*: the perceived difference between hues $H_1$ and $H_2$ was equal to the perceived difference between a rectangle rotated 0° and one rotated 60° (where perceived difference is explained below)

A feature of the Munsell colour space is that Munsell colours with the same value are isoluminent. Property 1 was satisfied by ensuring both hues had the same value in Munsell space. We chose Munsell value 7, because that slice through Munsell space provided a large number of displayable colours for a variety of different hues.

Property 2 was satisfied by running a set of preliminary experiments. We started with a simple target detection task. Subjects were asked to detect the presence or absence of a rectangle rotated 60° in a field of distractor rectangles rotated 0°. Both the target and distractor rectangles were coloured 5R 7/8. The average reaction time for detection was computed from the trials in which the subjects responded correctly. After the first experiment, the target and distractors were changed. The target was a rectangle coloured 10RP 7/8. The distractors were rectangles coloured 5R 7/8. The target was a single "hue step" from the distractors in Munsell space. Both the target and distractor rectangles were rotated 0°. The average reaction time for detection was computed from the trials in which the subjects responded correctly.

The hues used for the target and distractors during the second experiment were very similar. Because of this, the average reaction time for the second experiment was higher than the average reaction time for the first experiment. Additional experiments were run as follows.

- the target was moved another "hue step" away from the distractors (i.e., 5RP 7/8, 10P 7/8, and so on)
- the second experiment was re-run, and average reaction time was computed
- this process continued until an average reaction time equal to or below the average reaction time of the first experiment was obtained

This process provided two isoluminent hues $H_1$ and $H_2$ with a perceived difference equal to that of a 60° rotation, where perceived difference is measured by reaction time in the target detection experiment. Analysis of the preliminary experiment results led us to choose a red hue (Munsell 5R 7/8) and a blue hue (Munsell 5PB 7/8).

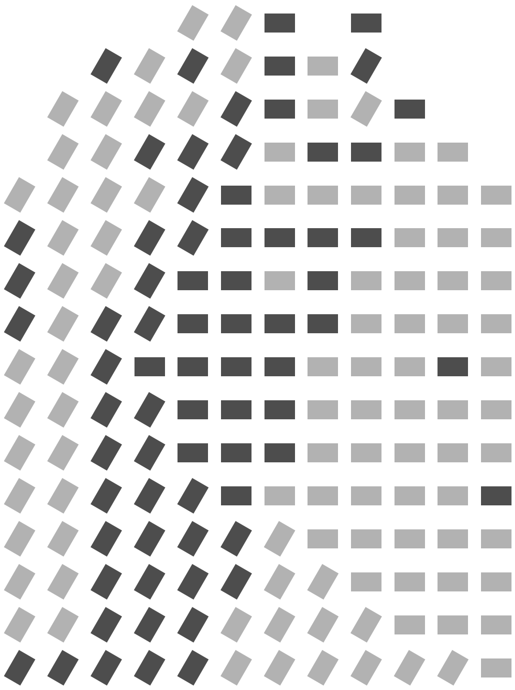


Figure 3: Example of a display from block $B_1$, data value $v_1$ (latitude of landfall) represented by hue, data value $v_2$ (ocean current) represented by orientation. Hue is represented by grey scale

Our design allowed us to use oriented, coloured rectangles to represent data elements with two associated data values $v_1$ and $v_2$. The experiment was divided into four subsections or "blocks" of experiment trials $B_1$, $B_2$, $B_3$, and $B_4$. The primary and secondary data value varied within each block, as did the primary and secondary preattentive feature. This gave us the following:

1. Primary data value was $v_1$, represented by hue; secondary data value was $v_2$, represented by orientation (Figure 3)

2. Primary data value was $v_1$, represented by orientation; secondary data value was $v_2$, represented by hue

3. Primary data value was $v_2$, represented by hue; secondary data value was $v_1$, represented by orientation

4. Primary data value was $v_2$, represented by orientation; secondary data value was $v_1$, represented by hue

During the experiment, subjects were shown a display similar to Figure 3 for 450 milliseconds. The screen was cleared, and subjects were asked to estimate the number of elements in the display with a given preattentive feature, to the nearest 10%. For example, in blocks $B_1$ and $B_3$ subjects were asked to estimate the number of rectangles coloured blue, to the nearest 10%. In blocks $B_2$ and $B_4$ they were asked to estimate the number of rectangles oriented 60°.

The two data values $v_1$ and $v_2$ represented latitude of landfall values and ocean current patterns from Oceanography's salmon migration simulations. Latitude of landfall had two possible values: "north" or "south". Ocean current had two possible values: "low" or "high". The primary data values for some trials were modified to meet statistical requirements for the data used in the experiment. For example, in blocks $B_1$ and $B_2$ the data value $v_1$ (latitude of landfall) was modified to ensure that:

1. An equal number of trials had a given percentage of data elements with a $v_1$ value of "north" (i.e., 4 trials where 5-15% of the data elements had a $v_1$ value of "north", 4 trials where 15-25% of the data elements had a $v_1$ value of "north", and so on up to 85-95%). This allowed us to compare estimation ability across a uniform range of percentages

2. Any dependence which might have existed between $v_1$ (latitude of landfall) and $v_2$ (ocean current) was removed. This ensured subjects could not infer information about the primary data value by examining the secondary data value

Trials were divided equally between control trials, where the secondary feature was fixed to a specific constant value, and experimental trials, where the secondary feature was used to represent the secondary data value which varied from element to element. This allowed us to test for feature interference. Better performance in control trials versus experimental trials would suggest that using a secondary feature to encode an "irrelevant" data value interfered with a subject's estimation ability for the primary feature. We tested

| Interval | Control 1 | | | | Control 2 | | | | Experimental | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\overline{V}$ | $\sigma(V)$ | $\overline{e}$ | $\sigma(e)$ | $\overline{V}$ | $\sigma(V)$ | $\overline{e}$ | $\sigma(e)$ | $\overline{V}$ | $\sigma(V)$ | $\overline{e}$ | $\sigma(e)$ |
| 1 | 1.25 | 0.53 | 0.25 | 0.53 | 1.33 | 0.70 | 0.33 | 0.70 | 1.29 | 0.68 | 0.29 | 0.68 |
| 2 | 1.83 | 0.82 | 0.58 | 0.58 | 2.04 | 0.86 | 0.62 | 0.58 | 2.17 | 0.83 | 0.46 | 0.71 |
| 3 | 2.71 | 0.75 | 0.46 | 0.66 | 2.75 | 0.85 | 0.67 | 0.56 | 2.79 | 0.71 | 0.54 | 0.50 |
| 4 | 4.17 | 1.13 | 0.75 | 0.85 | 3.75 | 1.11 | 0.83 | 0.76 | 3.83 | 1.49 | 1.08 | 1.03 |
| 5 | 5.50 | 1.32 | 1.00 | 0.98 | 5.08 | 1.67 | 1.42 | 0.83 | 5.54 | 1.41 | 1.25 | 0.84 |
| 6 | 5.96 | 1.27 | 0.96 | 0.81 | 6.71 | 1.23 | 1.21 | 0.72 | 6.31 | 1.17 | 0.94 | 0.76 |
| 7 | 6.83 | 1.01 | 0.75 | 0.68 | 7.42 | 0.78 | 0.67 | 0.56 | 7.19 | 0.73 | 0.52 | 0.55 |
| 8 | 8.13 | 0.80 | 0.46 | 0.66 | 8.33 | 0.56 | 0.42 | 0.50 | 8.15 | 0.62 | 0.40 | 0.49 |
| 9 | 8.71 | 0.55 | 0.29 | 0.55 | 8.96 | 0.20 | 0.04 | 0.20 | 8.65 | 0.53 | 0.35 | 0.53 |
| Total | 5.01 | 2.71 | 0.61 | 0.75 | 5.15 | 2.84 | 0.69 | 0.74 | 5.10 | 2.72 | 0.65 | 0.77 |

Table 1: Summary of block $B_1$ experiment results, showing average subject response $\overline{V}$, standard deviation of subject response $\sigma(V)$, average subject estimation error $\overline{e}$, and standard deviation of subject estimation error $\sigma(e)$ for each interval

both for orientation interfering with hue estimation and for hue interfering with orientation estimation.

Twelve subjects with normal or corrected acuity and normal colour vision were tested. The experiments were conducted in the Department of Psychology's vision laboratory, using a Macintosh II microcomputer equipped with a 13-inch RGB monitor and video hardware capable of displaying 256 colours simultaneously. The software used was designed and written by Rensink and Enns to run preattentive psychology experiments [Enn91]. Each subject completed either blocks $B_1$ and $B_3$ (blocks using hue as the primary feature) or blocks $B_2$ and $B_4$ (blocks using orientation as the primary feature).

At the beginning of the experiment, subjects were shown a sample display frame. The experiment procedure and task were explained to the subjects. Subjects were then shown how to enter their estimation. This was done by typing a digit on the keyboard between 1 and 9, which corresponded to the interval (percentage of rectangles) they estimated contained the target feature: interval 1 (5-15%), interval 2 (15-25%), and so on up to interval 9 (85-95%). Subjects were told no trial would contain less than 5% or more than 95% of the target rectangles.

Subjects began the experiment with a set of practice trials. This consisted of nine trials, one for each of the nine possible intervals. In one trial 10% of the rectangles were targets, in another 20% were targets, and so on up to 90%. The practice trials were designed to calibrate the subjects' responses and to give them an idea of the speed of the trials and the experiment. Trials were displayed one after another to the subjects. If subjects estimated correctly, they moved immediately to the next trial. If they estimated incorrectly, the trial was redisplayed, and they were told the correct answer.

Next, subjects completed a second set of practice trials. This phase consisted of 18 trials, two for each of the nine possible intervals. Trials were displayed in a random order to the subjects. This phase was designed to run more like a real experiment block. Trials in which the subjects estimated incorrectly were not redisplayed and subjects were not told the correct answer, although they did know whether their estimation was right or wrong.

Finally, subjects completed two experiment blocks, $B_1$ and $B_3$ or $B_2$ and $B_4$. Each block consisted of 72 control trials and 72 experimental trials. The 144 trials from each block were presented to the subjects in a random order. Subjects were provided with an opportunity to rest after every 48 trials. Data from all four phases were saved for later analysis.

## RESULTS

The primary dependent variable examined was estimation error, defined as the absolute difference between the subject's estimate and the percentage of target elements for the display. Statistical analyses using $t$-tests and analysis of variance (ANOVA) $F$-tests revealed the following findings:

- rapid and accurate estimation can be performed using either hue or orientation
- there is no evidence of a subject preference for either hue or orientation during the estimation task for the particular hue and orientation values used
- there is evidence of a subject preference for the spatial arrangement of data being displayed during the estimation task
- there is no evidence that orientation interferes with a subject's ability to perform hue estimation
- there is no evidence that hue interferes with a subject's ability to perform orientation estimation

The first question we asked was whether subjects were able to perform accurate estimation in a 450 millisecond exposure duration. Table 1 shows results of combined subject data for the control and experimental subsections of block $B_1$ as an example of the data calculated for each block. The results showed that accurate estimation was possible during the experiment for all four blocks. In the experimental subsections the total estimation error $\overline{e}$ ranged from a low of 0.54 in block $B_2$ to a high of 0.65 in block $B_1$. The standard deviation $\sigma(e)$ was below 1.0 in all four blocks. This indicates that subject responses were clustered close to the correct estimate. Results from the two control subsections show similar trends.

| Subsection | $n_1$ | $n_2$ | $v$ | $t$ |
|---|---|---|---|---|
| Control 1 | 432 | 432 | 862 | 0.36 |
| Control 2 | 432 | 432 | 862 | 1.43 |
| Experimental | 864 | 864 | 1726 | 0.45 |

Table 2: $t$-test results for estimation error rates from hue and orientation trials, showing the subsection, the number of hue trials $n_1$, the number of orientation trials $n_2$, the degrees of freedom $v$, and the $t$-value $t$

A point of interest was whether a subject's estimation ability differed depending on the feature being estimated. A $t$-test was computed to see if mean estimation error was equal across primary features for both the control and experimental subsections. Trials were combined into two groups: trials where orientation was the primary preattentive feature and trials where hue was the primary preattentive feature.

There appears to be no feature preference for the estimation task, since the control $t$-values (Table 2) are less than ${}_{0.975}t_{862} = 1.962$ and the experimental $t$-value is less than ${}_{0.975}t_{1726} = 1.960$. We did not expect to observe a feature preference, because we calibrated the perceived difference between our two hues and our two orientations to be equal before the experiment.

It is possible that the spatial distribution of the data affects a subject's estimation ability. It may be easy to perform estimation if the data elements cluster into two distinct groups. Similarly, if the data elements are distributed randomly throughout the display, estimation may be difficult. We used two different data sources during the experiment, $v_1$ and $v_2$, which corresponded to results from the salmon migration simulations. Both data types tended towards their own distinctive spatial distribution. A difference in mean estimation error across data types would indicate estimation ability depends, at least in part, on the spatial distribution of data being displayed. Trials were combined into two groups: trials where $v_1$ was the primary data value and trials where $v_2$ was the primary data value.

| Subsection | $n_1$ | $n_2$ | $v$ | $t$ |
|---|---|---|---|---|
| Control 1 | 432 | 432 | 862 | 2.06 |
| Control 2 | 432 | 432 | 862 | 1.73 |
| Experimental | 864 | 864 | 1726 | 1.84 |

Table 3: $t$-test results for estimation error rates from $v_1$ and $v_2$ trials, showing the subsection, the number of $v_1$ trials $n_1$, the number of $v_2$ trials $n_2$, the degrees of freedom $v$, and the $t$-values $t$.

Control subsection 1's $t$-value (Table 3) is greater than ${}_{0.975}t_{862} = 1.962$. This suggests data type did have an effect on estimation error in control subsection 1. Control subsection 2's $t$-value is less than 1.962, but it does fall between ${}_{0.95}t_{862} = 1.647 < p < {}_{0.975}t_{862}$. Similarly, the experimental subsection's $t$-value falls between ${}_{0.95}t_{1726} = 1.645 < p < {}_{0.975}t_{1726} = 1.960$. The $t$-test results indicate the possibility of data type influence on estimation error. With $\alpha = 0.10$, we would conclude data type may affect estimation error in all three subsections. Additional experiments which explicitly control the change in spatial distribution are needed before we can state specifically its effect on the estimation task.

One question of interest was whether encoding an irrelevant data value with a secondary preattentive feature affected a subject's estimation ability. We began by checking to see if orientation interfered with a subject's ability to estimate using hue. $t$-tests were computed to compare estimation error mean across control and experimental subsections for blocks $B_1$ and $B_3$, the blocks that used hue as their primary preattentive feature.

| Subsection | $n_1$ | $n_2$ | $v$ | $t$ |
|---|---|---|---|---|
| $B_1$ | 432 | 432 | 862 | 0.03 |
| $B_3$ | 432 | 432 | 862 | 0.21 |

Table 4: $t$-test results for estimation error rates from control and experimental hue trials, showing the block, the number of control trials $n_1$, the number of experimental trials $n_2$, the degrees of freedom $v$, and the $t$-value $t$

The $t$-values for both blocks (Table 4) are less than ${}_{0.975}t_{862} = 1.962$. Therefore, there appears to be no interference due to encoding of an irrelevant data value using orientation. Any difference in means is probably due to sampling error.

We continued to investigate interference by checking to see if hue interfered with a subject's ability to estimate using orientation. $t$-tests were computed to compare mean estimation error across control and experimental subsections for blocks $B_2$ and $B_4$, the blocks that used orientation as their primary preattentive feature.

| Subsection | $n_1$ | $n_2$ | $v$ | $t$ |
|---|---|---|---|---|
| $B_2$ | 432 | 432 | 862 | 0.23 |
| $B_4$ | 432 | 432 | 862 | 1.15 |

Table 5: $t$-test results for estimation error rates from control and experimental orientation trials, showing the block, the number of control trials $n_1$, the number of experimental trials $n_2$, the degrees of freedom $v$, and the $t$-value $t$

The $t$-values for both blocks (Table 5) are less than ${}_{0.975}t_{862} = 1.962$. Therefore, the appears to be no interference due to encoding of an irrelevant data value using hue. Any difference in means is probably due to sampling error.

## EXPOSURE DURATION EXPERIMENTS

Our conclusions in the first experiment apply to data displayed for an exposure duration of 450 milliseconds. This leaves two important questions unanswered. First, at what exposure duration are subjects no longer able to perform robust estimation? Second, do any interference effects begin to appear at lower exposure durations? For example, we found that orientation did not interfere with estimation of hue at a 450 millisecond exposure duration. It may be that an interference effect does exist, but 450 milliseconds gives subjects enough time to overcome this effect. Feature preference may also be dependent on exposure duration.

We conducted a second experiment in which exposure duration for each trial varied among five possible values: 15, 50, 100, 200, and 450 milliseconds. Trials were presented to subjects in the following way:

- a blank screen was displayed for 200 milliseconds
- a focus circle was displayed for 100 milliseconds
- the trial was displayed for its exposure duration (one of 15, 50, 100, 200, or 450 milliseconds)
- a "mask" of randomly oriented grey rectangles was displayed for 100 milliseconds
- the screen blanked, and subjects were allowed to enter their estimation

Because trials came from block $B_1$, our primary data value was $v_1$ (latitude of landfall), represented by hue, and our secondary data value was $v_2$ (current pattern), represented by orientation. Subjects estimated the number of blue rectangles in each trial. As before, an equal number of trials (10 control and 10 experimental) for each interval were used. Trials at each interval were split evenly among the five exposure durations, and were presented to the subjects in a random order so the various exposure durations were intermixed.

Analysis of data from the previous experiment showed estimation was accurate at every interval. Because of this, we combined trials with a given exposure duration into a single block of data. For example, trials that were displayed for 100 milliseconds formed a single group of 2 control and 2 experimental trials from each interval for a total of 18 control and 18 experimental trials. We plotted average estimation error versus exposure duration to see if estimation ability was affected by display time. Figure 4 shows the graph of average estimation error versus exposure duration for experimental trials.

Average estimation error and standard deviation of error seemed to be reasonably stable, even down to

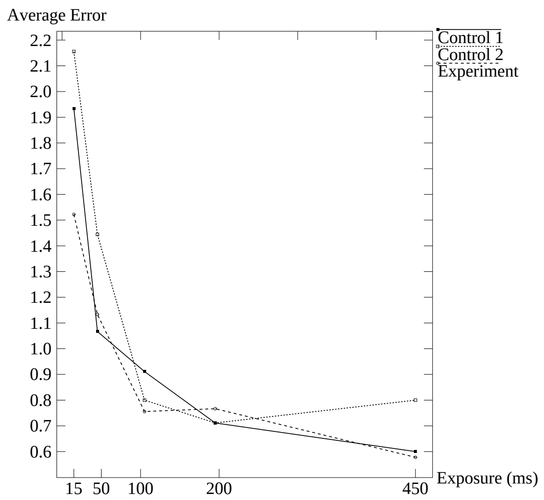



Figure 4: Graph of average error across exposure duration for combined results from exposure duration experiment

100 milliseconds. Below that duration error values increased rapidly. This indicates the minimum exposure duration for robust hue estimation lies somewhere between 50 and 100 milliseconds. We concluded our analysis by checking to see if orientation interfered with hue estimation at any of the exposure durations. $t$-tests were computed to compare mean estimation error across control and experimental subsections for all five exposure durations. The $t$-values for all durations were less than ${}_{0.975}t_{178} = 1.972$. Only the 15 millisecond exposure duration had a $t$-value which might be considered significant, ${}_{0.90}t_{178} = 1.286 < p < {}_{0.95}t_{178} = 1.653$. This suggests orientation is not interfering with hue estimation at any of the exposure durations tested.

## FUTURE WORK

Our experiments and related analysis leave open a number of interesting avenues for future work. We could examine in more detail numeric estimation and its relationship to specific visualization applications. We explicitly chose two hues whose perceived difference from one another was equal to the perceived difference between two rectangles oriented 0° and 60°. A choice of features perceptually different from one another might cause a subject feature preference during the estimation task. We could also test different features, such as intensity and size, to see how they perform during the estimation task.

Work which provides general guidelines for the use of preattentive features in the design of visualization tools should be pursued. Many visualization tasks require more than two data values to be encoded at each spatial location. Future experiments could examine how to encode higher-dimensional elements in a low-dimensional environment. This type of visualization tool could exhibit new and unexpected types of interference. There may also be a limit to the amount of information a subject can extract and process at one time.

The data values used in our experiment were derived from salmon migration studies in Oceanography. More comprehensive studies based on actual tasks performed by researchers are needed before conclusive evidence will exist for using preattentive features in real-world multivariate data analysis such as salmon migration simulations, air traffic control, and medical imaging. Other types of data should be investigated as well if general visualization tools are to be based on preattentive processing.

## ACKNOWLEDGEMENTS

C.G. Healey gratefully acknowledges the following support for his participation in this work: a Natural Sciences and Engineering Research Council graduate scholarship, and the equipment and staff of the Media and Graphics Interdisciplinary Centre, Department of Computer Science, University of British Columbia. C.G. Healey also thanks K.A. Thomson, W.J. Ingraham, and P.H. LeBlond for their direction, comments, and support.

## REFERENCES

[Bel87] Bell, P.C. and R.M. O'Keefe (1987). Visual Interactive Simulation—History, Recent Developments, and Major Issues. *Simulation,* **49**(3), 109–116.

[Bir69] Birren, F. (1969). *Munsell: A Grammar of Color.* New York, New York: Van Nostrand Reinhold Company.

[Cal84] Callaghan, T.C. (1984). Dimensional Interaction of Hue and Brightness in Preattentive Field Segregation. *Perception & Psychophysics,* **36**(1), 25–34.

[Cal89] Callaghan, T.C. (1989). Interference and Domination in Texture Segregation: Hue, Geometric Form, and Line Orientation. *Perception & Psychophysics,* **46**(4), 299–311.

[Enn90a] Enns, J.T. (1990). The Promise of Finding Effective Geometric Codes. *Proceedings Visualization '90,* 389–390. San Francisco, United States.

[Enn90b] Enns, J.T. and R.A. Rensink (1990). Sensitivity to Three-Dimensional Orientation in Visual Search. *Psychology Science,* **1**(5), 323–326.

[Enn91] Enns, J.T. and R.A. Rensink (1991). VSearch Colour: Full-Colour Visual Search Experiments on the Macintosh II. *Behaviour Research Methods, Instruments, & Computers,* **23**(2), 265–272.

[Gri89] Grinstein, G., R. Pickett, and M. Williams (1988). EXVIS: An Exploratory Visualization Environment. *Proceedings Graphics Interface '89,* 254–261. London, Canada.

[Hea92] Healey, C.G. (1992). Visualization of Multivariate Data Using Preattentive Processing. *Masters Thesis.* Department of Computer Science, University of British Columbia.

[Jul83] Julész, B. and J.R. Bergen (1983). Textons, the Fundamental Elements in Preattentive Vision and Perception of Textures. *The Bell System Technical Journal,* **62**(6), 1619–1645.

[McC87] McCormick, B.H., T.A. DeFanti, and M.D. Brown (1987). Visualization in Scientific Computing—A Synopsis. *IEEE Computer Graphics & Applications,* **7**(7), 61–70.

[Mel85] Melamed, B. and R.J.T. Morris (1985). Visual Simulation: The Performance Analysis Workstation. *IEEE Computer,* **18**, 87–94.

[Pic88] Pickett, R. and G. Grinstein (1988). Iconographic Displays for Visualizing Multidimensional Data. *Proceedings of the 1988 IEEE Conference on Systems, Man, and Cybernetics,* 514–519. Beijing and Shenyang, China.

[Qui87] Quinlan, P.T. and G.W. Humphreys (1987). Visual Search for Targets Defined by Combinations of Color, Shape, and Size: An Examination of the Task Constraints on Feature and Conjunction Searches. *Perception & Psychophysics,* **41**(5), 455–472.

[Set88] Sethian, J.A., J.B. Salem, and A.F. Ghoniem (1988). Interactive Scientific Visualization and Parallel Display Techniques. *Proceedings Supercomputing '88,* 132–139. Orlando, United States.

[Tho92a] Thomson, K.A., W.J. Ingraham, M.C. Healey, P.H. LeBlond, C. Groot, and C.G. Healey (1992). The Influence of Ocean Currents on Latitude of Landfall and Migration Speed of Sockeye Salmon Returning to the Fraser River. *Fisheries Oceanography,* **1**(2), 163–179.

[Tho92b] Thomson, K.A., W.J. Ingraham, M.C. Healey, P.H. LeBlond, C. Groot, and C.G. Healey (1992). The Influence of Ocean Currents on Return Timing of Fraser River Sockeye Salmon. *Canadian Journal of Fisheries and Aquatic Sciences,* in press.

[Tri85] Triesman, A. (1985). Preattentive Processing in Vision. *Computer Vision, Graphics, and Image Processing,* **31**, 156–177.

[War88] Ware, C. and J.C. Beatty (1988). Using Colour Dimensions to Display Data Dimensions. *Human Factors,* **30**(2), 127–142.