

Exploring the Effects of Group Size and Display Configuration on Visual Search

Clifton Forlines¹, Chia Shen¹, Daniel Wigdor^{1,2}, Ravin Balakrishnan²

¹Mitsubishi Electric Research Labs
201 Broadway, 8th Floor
Cambridge, MA, 02139, USA

{forlines, shen}@merl.com

²Department of Computer Science
University of Toronto
Toronto, ON, Canada

{dwigdor, ravin@dgp.toronto.edu}

ABSTRACT

Visual search is the subject of countless psychology studies in which people search for target items within a scene. The bulk of this literature focuses on the individual with the goal of understanding the human perceptual system. In life, visual search is performed not only by individuals, but also by groups – a team of doctors may study an x-ray and a team of analysts may study a satellite photograph. In this paper, we examine the issues one should consider when searching as a group. We present the details of an experiment designed to investigate the impact of group size on visual search performance, and how different display configurations affected that performance. We asked individuals, pairs, and groups of four people to participate in a baggage screening task in which these teams searched simulated x-rays for prohibited items. Teams conducted these searches on single monitors, a row of four monitors, and on a single horizontal display. Our findings suggest that groups commit far fewer errors in visual search tasks, although they may perform slower than individuals under certain conditions. The interaction between group size and display configuration turned out to be an important factor as well.

Categories and Subject Descriptors

H.5.3. [Information Interfaces and Presentation (e.g., HCI)]: Group and Organization Interfaces.

General Terms

Design, Experimentation, Human Factors.

Keywords

Visual search, group size, horizontal display, shared-display groupware, multi-monitor interfaces.

1. INTRODUCTION

Visually searching images for target objects is a critical task in a variety of domains. These include screening for prohibited items in x-ray imagery of luggage in airports, identifying anomalies in medical imagery, and looking for unusual activity and changes over time in satellite imagery. Unlike many other human-machine interface tasks where a small error rate is quite acceptable, a crucial element of visual search tasks in these application areas is that a single mistake can have catastrophic consequences. Of particular concern is recent research [19] which demonstrates that

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

CSCW'06, November 4–8, 2006, Banff, Alberta, Canada.

Copyright 2006 ACM 1-59593-249-6/06/0004...\$5.00.

when the occurrence of a target is rare – as is typical in the airport screening and medical imaging examples – observers' ability to detect those targets are disturbingly poor. In other words, unless observers see what they are looking for fairly frequently, they will often miss it when it does appear. Thus, any improvements – in technology or human processes – that can be made to aid visual search would be invaluable in these safety critical areas.

Although visual search has been studied extensively [18], most of this literature focuses on understanding an individual's performance in a search task and the underlying human perceptual processes. In real-life applications, however, visual search may be conducted by groups of people rather than an individual – several doctors often study an x-ray, teams of security personnel work at screening checkpoints, and many analysts may pore over satellite imagery. When groups of people work together in this manner, it is plausible that their performance is a function of both their individual perceptual capabilities and the dynamics of the group, with factors such as competition and discussion amongst group members possibly playing a significant role. Whether or not these group dynamics help or hinder the search task is an important question that has not been explored in the literature.

From a technology standpoint, most visual search studies have been conducted with an individual observer viewing a single scene on a single vertical display. In practice, however, visual search need not be restricted to vertical displays and could benefit from new form factors such as horizontal displays and multiple monitors. While much recent research has explored using these new displays, their benefits, if any, are not fully understood either in general or in the context of visual search per se. For example, it is unclear whether using the increased pixel count of multiple displays to provide multiple, simultaneous orientations of an image, without degradation in quality, adds any value to visual search tasks. Similarly, it is unclear if and how the horizontal orientation of a tabletop display might impact visual search. Perhaps most interesting is the question of how different display factors might affect the behavior and performance of groups of people performing search tasks.

In this paper, we present an experiment designed to probe these questions. In particular, we consider the impact of group sizes of 1, 2, and 4 people and three display configurations – a single vertical display, a horizontal tabletop display, and four vertical displays each showing the same image at one of four orientations – on the performance of a visual search task where target occurrence is rare. The display configurations chosen are ones that could be deployed in practical settings, while nonetheless providing sufficient variability in several dimensions to be interesting from a theoretical/experimental perspective. Our results show the value of having groups perform visual search, and can guide the design of future displays for such tasks.

2. RELATED WORK

2.1 Visual Search

Visual search has been studied extensively for over a century and a full discussion of the field is well beyond the scope of this paper. Many underlying theories have been presented as to the nature of the human visual system; however, no single model explains the variety of experimental results in the literature [18].

In a typical visual search experiment, participants are asked to look for a specific target object within a stimulus image that may or may not contain the target and does contain a varying number of distracter objects. Normally, the target object is present in 50% of the stimulus images and the participant responds to each stimulus image by indicating whether or not they believe the target object is present. Participants are asked to respond as quickly and as accurately as possible, and the dependent variables measured are reaction time and error rate. Typically, reaction time is plotted as a function of number of distracter objects, and the slope of this line tells the experimenters something about the increase in search complexity for each additional item in the stimulus image.

Slopes near or at zero indicate no additional cost to searching through a larger set of items. Such searches are called pre-attentive, or efficient, and typically involve searching for an item that differs from all of the distracter items in terms of a basic, easily recognizable feature (e.g. searching for a blue circle in a collection of red circles). A steep slope indicates a less efficient search and is often called attentional. Participants performing an attentional search task must focus their attention on each item, normally skipping randomly from item to item around the image.

Reaction times are typically higher for target-absent trials than for target-present trials. Participants are typically described as having an internal confidence threshold that they must pass in order to form a conclusion as to the presence or absence of the target object in the stimulus image. In target-absent trials, the greater the number of items to search through, the longer participants take to pass this threshold. In target-present trials, the threshold is immediately passed upon the recognition of the target object. Error rates are typically higher for images that contain a greater number of distracter objects. Missing a target object is much more common than false positives.

Error rates are typically around 10% as participants naturally slow down if they are making too many errors and gradually speed up if they are committing few errors; however, the majority of visual searching studies present the target object in 50% of the stimulus images. In a recent study, Wolfe et al. [19] tested target object prevalence of 50%, 10%, and 1%, pointing out that in many important real-world tasks, such as baggage screening and x-ray analysis, targets of interest, such as knives and tumors, were very rare. Their somewhat disturbing results showed that error rate grew significantly as target prevalence dropped – from 7% error in the 50% prevalence trials, to 16% error for 10% prevalence, to 30% error for 1% prevalence.

2.2 Effects of Object Rotation on Search

As with many theories in the field of visual search, there are competing models to predict the effects of object orientation on search time and accuracy. *Geon Structural Descriptions* theory states that the shape of an object is a composite of simple three-dimensional parts (called geons) and that object recognition is

viewpoint invariant (i.e. rotation invariant) [2]. This theory would predict no benefit in terms of object recognition from the multiple viewpoints that a group is provided as they sit around a shared horizontal display.

Geon Structural Descriptions theory stands in contrast to the multiple-view theory [9], which states that individuals have a collection of prototypical views of an object, and that objects are more easily recognized when they are presented from a viewpoint that is close to one of these prototypical views. Indeed, object rotation has been shown to affect reaction time and error rate [9]. Objects with a strong axis of orientation are more easily identified when they are presented right-side-up than when they are shown at an arbitrary rotation (Figure 1). Given the multiple-view theory, we should expect a team of individuals to benefit from the multiple views on a dataset that standing around a horizontal display provides. The multiple-views minimize the maximum amount of rotation from the canonical view of an object when the group is looked at as a whole. In other words, one group member is guaranteed to have a good view of the object that the team is searching for. Unsurprisingly, there are a number of experimental results that show both viewpoint-invariant and viewpoint-dependent performance for object recognition; however, for tasks that are shown to be viewpoint-dependent, a horizontal display should aid group search.

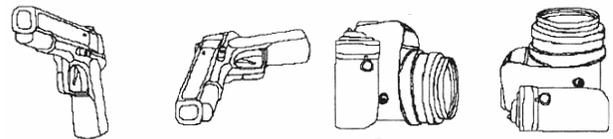


Figure 1: Objects are more easily recognized when viewed at their canonical orientation (left) than when rotated (right). Images from [10]

2.3 Benefits of New Display Configurations

The use of multiple displays for an individual's workstation has become more common in recent years, and the performance and preferential benefits of multiple displays has been a topic of much research [1, 3, 5, 7, 17]. This body of work generally points out not only performance, but also preferential, benefits of multiple displays for single-user applications. The large pixel resolution of these workspaces allow for multiple views of a dataset to be presented simultaneously [12], which may aid not only an individual's understanding of a dataset [13], but also the coordination of an interdisciplinary team.

Ryall et al. [14] presented the results of a study in which groups of two, three, and four completed a collaborative task on an interactive tabletop during which teams built target poems out of individual word tiles spread randomly among distracter word tiles. Their findings show a linear decrease in task time with the addition of each extra group member, and the bulk of their discussion center on the social interaction among group members and the roles that group members adopted. While their study did involve large amounts of visual searching of text, the complex task also included document manipulation as well as the management of shared resources. The effects of each portion of this task are difficult to separate from one another, and they did not compare task performance on different display configurations.

3. EXPERIMENT

3.1 Goals and Considerations

The goals of our experiment were twofold: to investigate both the impact of group size on visual search performance, and how different display configurations affected that performance. We chose a simulated baggage screening task modeled after that presented in [19]. Participants visually searched an image looking for knives or guns among a random collection of objects. In designing this experiment, there are several issues regarding group size and display configuration that are worth considering.

3.1.1 Behavior of Individuals Within Groups.

It is known that the larger the group the less responsibility each individual feels toward the shared responsibilities of the group [4]. For a task like visual searching, are any potential benefits accrued by having additional “pairs of eyes” outweighed by the reduction in responsibility that each individual feels? Similarly, would individuals feel social pressure to conform to some consensus answer that the group is converging upon, as suggested by Janis [8]. Thus, while it may appear obvious that having more people do the search task should improve performance, these issues question that premise, making empirical investigation necessary.

3.1.2 Search Strategy

When performing visual search in a group setting, it is possible that the strategies used could be significantly impacted by the group’s size and display configuration. Larger groups might choose to distribute the work in a variety of ways amongst group members – perhaps differently depending on the display configuration – rather than duplicating efforts. Groups could also benefit from discussion and other problem-solving strategies that are not available to individuals. Studies have explored these issues in non-computer augmented settings [16] but not in shared-display groupware settings in the context of visual search.

3.1.3 Object Visibility

In a visual search task, objects that are on the far side of a large display may be hard to see. This drawback might be mitigated in group visual search since individuals could have different views of the displayed imagery, depending on the display configuration. While limited visibility of objects for some members of the group reduces the shared context of the group, a large workspace, multiple monitors, or horizontal workspaces around which group members sit may implicitly partition the workspace and clearly assign certain portions of the image to each individual, making the overall task more efficient.

3.1.4 Object Rotation

For a group sitting around a table, no single orientation works best for everyone: a text document that appears right-side-up for one individual appears up-side-down to another. These multiple points-of-view are normally viewed as a problem to overcome [15]; however, they may be beneficial for groups performing visual search tasks. In an image of arbitrarily rotated objects, such as an x-ray of a suitcase, there is a greater chance of a group member having a visually advantageous view of the object and thus identifying it if the group is seated around a table rather than in front of a vertical display where everyone views the image in the same orientation.

3.2 Participants and Group Configurations

We recruited a total of 43 participants for our study who participated as six groups of four, seven groups of two, and five individuals. Participants were recruited from local universities and from local on-line bulletin boards. Individuals were paid \$20 compensation for participating, pairs \$40, and groups of four \$160 because of our initial difficulty in recruiting large groups. Experimental sessions lasted a little more than an hour depending on the speed of the groups. Participants’ ages ranged from 18 to 55 and were roughly evenly split by gender. All members of each group of two and four knew each other before participating in the study.

3.3 Display Configurations

We had three display configurations (Figure 2), all using NEC MultiSync LCD 2080UX+ monitors, running at half of their 1600x1200 native resolution.

Single Vertical Display. Participants sat shoulder-to-shoulder in front of a single display which rested on a desktop and was positioned at a comfortable working height. Chairs were positioned 60 cm from the display’s image plane.

Multiple Vertical Displays. Four displays were positioned in a row on the same desktop used in the single vertical display configuration. Each display showed the same image rotated in 2D on the screen in 90° increments. Single participants were seated in the center of the four displays, pairs were positioned in front of the center two displays, and groups of four had one participant in front of each display.

Single Horizontal Display. A single display was positioned horizontally at a height of 70 cm. Single participants sat along the “bottom” edge of the display, pairs sat across from one another in the middle of the long sides of the display, and groups of four had each participant sitting one on each side.

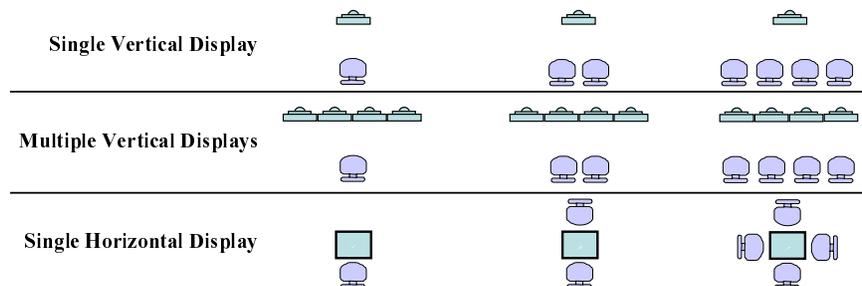


Figure 2. Top row: a bird’s eye view of the single vertical display configuration for individuals, pairs, and groups of four. Middle row: the multiple-vertical display configuration. Bottom row: the single horizontal display configuration.

These three display configurations were chosen firstly because they are practical and can be immediately implemented in security checkpoints, radiologist offices, and analysts' workspaces. Secondly, they provide a continuum of information content and a continuum of shared context among members in a group.

The single vertical display configuration presents the image in a single orientation to all group members, with no person getting any additional information from the others. Everyone in a group has the same shared context. The only potential advantages of an increased group size in this configuration are the "extra pairs of eyes" and discussion within the group.

The multiple vertical displays configuration presents the same image four times, each at a different 90° rotation. This 2D rotation is similar to what a typical baggage-screening machine is able to display but much less sophisticated than the 3D rotations offered by medical imaging devices. Even so, the amount of information available to all viewers is ostensibly increased due to being able to view the same image from different orientations, although the display directly in front of each user would likely act as the primary information source for that user. Further, larger groups might exhibit interesting group dynamics such as dividing up the task so that each user only focused on one or two of the four displays, or they might collaboratively discuss all four images.

The single horizontal display configuration presents a single image. As such, a single user should find this configuration roughly equivalent to the single vertical display in terms of information content, unless the horizontal vs. vertical orientation of the display itself affects visual search performance. With groups, however, since users are seated at different locations around the display, each user gets a different view of the image than others in the group. The amount of information available to users in this configuration therefore arguably increases as the group size increases. An additional interesting aspect of this horizontal configuration is that groups of users are better located for face-to-face communication. In contrast, the two vertical configurations are such that groups of users have to sit roughly shoulder-to-shoulder, which might impede communication.

3.4 Apparatus and Stimulus

We ran the experiment on a 4GHz AMD 64-bit PC running WindowsXP. The experiment application was programmed in Java. Stimulus images were 800x600 pixel collages made from a sampling of 150 images from the Hemera Photo-Objects Collection [6]. Images were pre-generated by starting with one of ten random background noise images, and adding either 12 or 16 distracter objects which were drawn with 40% opacity. Half the images had 12 distracters, the other half had 16. Object images were de-saturated and were positioned randomly in the image with the only constraint being that the object did not cross out of the image. Objects were rotated randomly around their center in the image plane. A target knife or a gun image was present in 10% of the stimulus images, and absent in the other 90% (i.e. 10% target prevalence). Figure 3 shows two example images from our study. Participants indicated whether or not either of the target images was present in the stimulus image by using a small keypad with two keys, one for target-present and one for target-absent. Each participant in a group had their own keypad.

Participants were given feedback as to their performance and encouraged to work quickly and accurately in three ways – visually, temporally, and with a scoring system. When the group was correct in identifying a target object, the system responded by displaying the target object in isolation on a background of green (Figure 3, *right*). This image remained on-screen for two seconds, long enough for the group to see that they had correctly found a target object. When the group missed a target object, the system displayed the missed target on a red background. This image remained on-screen for ten seconds, which not only notified the group of their error, but also penalized the team in that it took them longer to complete the experiment. This pause was meant to reduce the chance that a participant would race through the trials with careless responses in order to quickly finish the experiment.

The final means of feedback was provided with a simple scoring system. Participants were given 10 points for correct answers in target-absent trials, and 1000 points for correct answers in target-present trials. Missing a target resulted in a 1000 point deduction, and false-positives (saying the target was present when it was not) resulted in a 100 point penalty.

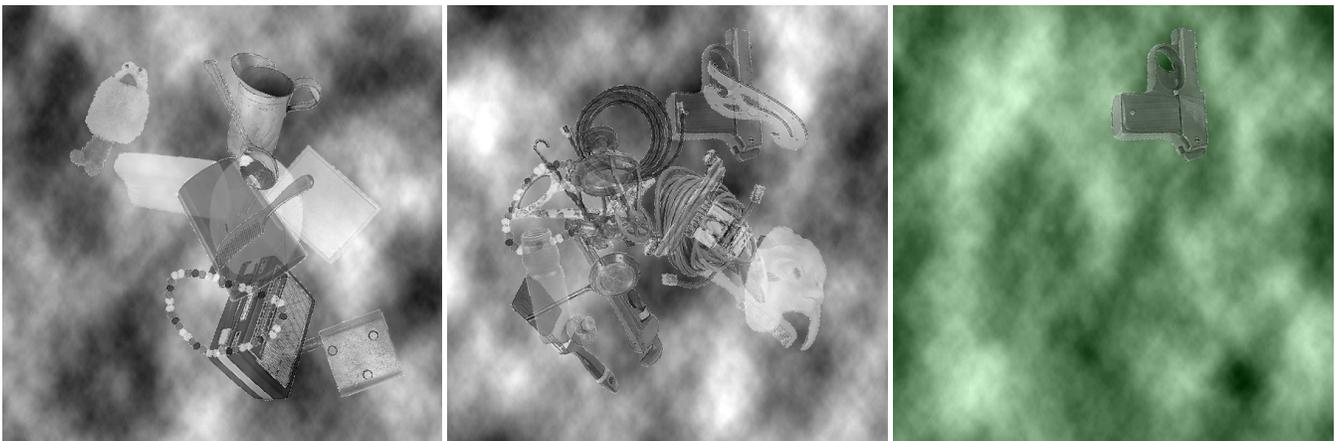


Figure 3. Example images in the experiment (left) Image with target object absent. (middle) Image with target object present. (right) Positive feedback image displayed when participants correctly identified a target in an image.

In the case where target objects were absent in the image, every member of the group had to answer before the testing application moved onto the next image, and the reaction time recorded for the team was the reaction time of the slowest individual. In the case where a target object was present in the image, the testing application would quickly display the target on its own, and then moved on to the next image immediately after the first participant indicated the target was present in the image and the reaction time recorded was this fastest reaction time. This approximates the real-world scenario where if anyone spots a suspicious object in an x-ray of a bag, the bag would be subjected to detailed inspection rather than requiring others to scrutinize the image further.

3.5 Procedure

Groups arrived at the testing location and were instructed in the operation of the input device.

A within-participant fully-counterbalanced design was used. All participants performed trials in all display configurations. The order of presentation of display configurations was counterbalanced across groups within each group size. The order of presentation of images within each display configuration was randomized, with each group seeing the same 150 images in total for each display configuration for a total of 8100 trials in the study. Participants in groups were allowed to talk freely amongst themselves during the experiment. This approximated a real world co-located collaboration scenario where users would naturally communicate in such tasks. Participants could take breaks between display configurations, but not within the 150 trials per display configuration.

3.6 Hypotheses

Based on the literature and our preliminary exploration of the issues, we formulated the following hypotheses:

- H1. Group size would significantly affect reaction time for trials where a target object was present in the image.
- H2. Display configuration would significantly affect reaction time.
- H3. Group size would have a significant interaction with display configuration in terms of reaction time.
- H4. Group size would significantly affect error rate. Because any individual in a group could identify the presence of a target object in the stimulus image, we expect that more people searching the image should lead to fewer errors.
- H5. Display configuration would significantly affect error rate.
- H6. Group size would have a significant interaction with display configuration in terms of error rate.

3.7 Results

We computed a repeated measures ANOVA for both reaction time and error rate using *display configuration*, and *number of distracters* as within-participant factors, and *group size* as a between-participants factor. Because of the different number of trials in the target-present and target-absent conditions (given the 10% target prevalence), we analyzed the two *target presence* conditions independently. Our findings support hypotheses H2, H3, and H4, while they allow the rejection of hypotheses H1, H5, and H6.

3.7.1 Reaction Time Analysis

1.07% of the trials had a reaction time of more than 3 standard deviations faster or slower from the mean reaction time for that trial's *group size / display configuration / number of distracters* combination. These outliers were removed and not included in our analysis of reaction time or error rate.

3.7.1.1 Target-present Reaction Time

The order of *display configuration* presentation did not significantly affect reaction time ($F_{2,9} = 1.76, p = 0.23$) and did not significantly interact with any of the independent variables, indicating that no asymmetrical learning effects occurred relative to reaction time and that our analysis could include a within-participant comparison of *display configurations*.

Contrary to our hypothesis H1, *group size* had no significant effect on reaction time when targets were present in the image ($F_{2,9} = 2.68, p = 0.12$). Although not statistically significant, on average, larger groups had lower reaction times, with mean reaction times of 4.23s, 3.87s, and 3.03s for individuals, groups of two, and groups of four respectively.

In contrast to hypothesis H2, *display configuration* had no significant effect on reaction time during target-present trials ($F_{2,18} = 0.66, p = 0.53$), with mean reaction times of 3.51s, 3.75s, and 3.88s for horizontal, single vertical, and multiple vertical displays respectively. Furthermore, although predicted by hypothesis H3, *display configuration* had no significant interaction with *group size* ($F_{4,18} = 1.38, p = 0.28$) in terms of target-present reaction time. Figure 4 shows the mean reaction times for each *display configuration* for each *group size*.

As is the case with attentional visual searching tasks, *number of distracters* had a strong effect on target-present reaction time ($F_{1,9} = 83.51, p < 0.001$), with stimulus images containing 12 distracters having a mean reaction time of 2.83s and those with 16 distracters a mean reaction time of 4.59s.

3.7.1.2 Target-absent Reaction Time

Again, *group size* had no significant effect on reaction time in target-absent trials ($F_{2,9} = 0.51, p = 0.61$). Overall, groups of four had a slower mean reaction time than the other group sizes, with means of 6.28s, 6.20s, and 7.00s for individuals, pairs, and groups of four respectively.

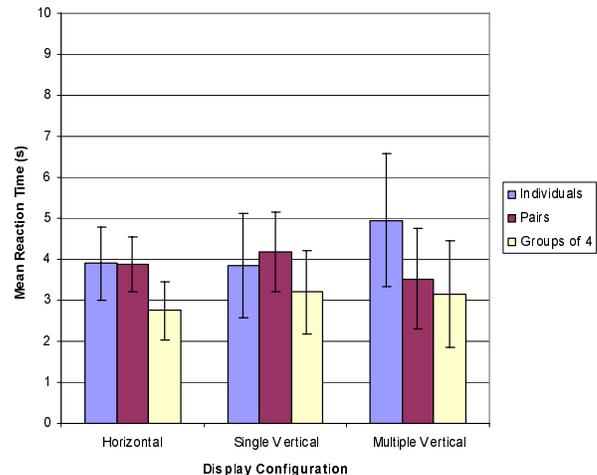


Figure 4: Mean target-present reaction times for each display configuration for each group size.

In support of hypothesis H2, *display configuration* had a significant effect on reaction time during target-absent trials ($F_{2,18} = 8.92$, $p = 0.002$), with mean reaction times of 6.22s, 6.05s, and 7.21s for horizontal, single vertical, and multiple vertical displays respectively. A post-hoc pair-wise means comparison showed that there was a significant difference between the horizontal and multiple vertical and between the single vertical and multiple vertical, but not between the horizontal and single vertical *display configurations*. Furthermore, as predicted by hypothesis H3, *display configuration* had a significant interaction with *group size* ($F_{4,18} = 3.21$, $p = 0.037$) in terms of reaction time during target-absent trials. Figure 5 shows the mean reaction times for each *display configuration* for each *group size*. While individuals were faster than pairs and groups of four with both the horizontal and single vertical display, they were significantly slower when working on the multiple vertical *display configuration*.

As is the case with attentional visual searching tasks, *number of distracters* had a strong effect on reaction time ($F_{1,9} = 59.59$, $p < 0.001$), with stimulus images containing 12 distracters having a mean reaction time of 5.95s and those with 16 distracters a mean reaction time of 7.03s.

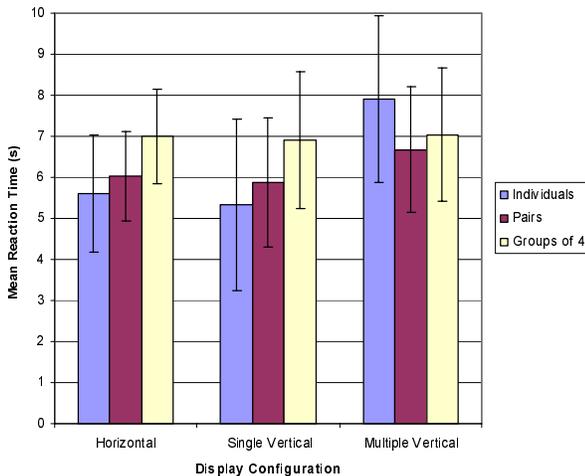


Figure 5. Mean target-absent reaction times for each display configuration for each group size.

3.7.1.3 Target Presence Reaction Time Comparison
 Interestingly, there appeared to be a significant interaction between *target presence* and *group size*, with larger groups having shorter reaction times during target-present trials than either pairs or individuals and visa-versa for target-absent trials. Figure 6 shows the mean reaction times for both *target presence* conditions for each *group size*.

Finally, there appears to be an interaction among *number of distracters*, *target presence*, and *group size*. Figure 7 shows the mean reaction times by *group size* for both *number of distracters* for each *target presence* condition. This type of figure is typical of visual search experiments, in part because the slopes of these lines indicate the efficiency with which a group handles extra distracter objects.

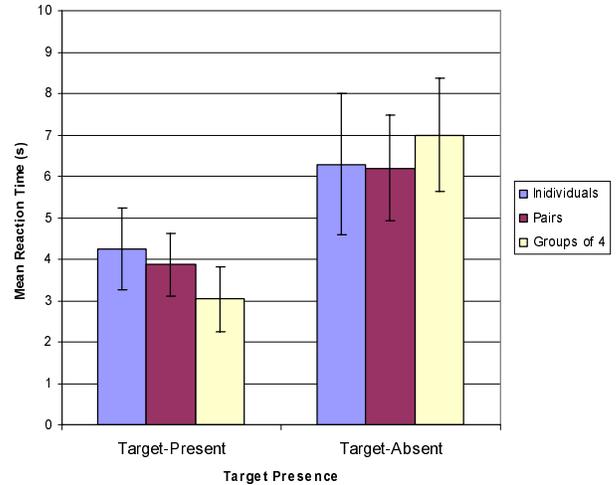


Figure 6. Mean reaction times for both target presence conditions for each group size.

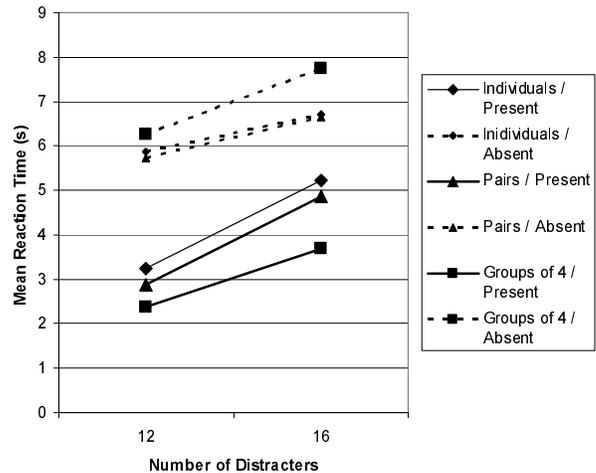


Figure 7. Mean reaction times for both number of distracters for both target presence conditions for each group size.

3.7.2 Error Rate Analysis

Again, because of the 10% target prevalence in our study, we analyzed target-present and target-absent trials independently. For a comparison between target presence conditions in terms of error rate, Figure 8 shows the mean error rates for both *target presence* conditions for each *group size*.

3.7.2.1 Target-present Error Rate

The order of *display configuration* presentation did not significantly affect error rate ($F_{2,9} = 0.271$, $p = 0.76$) and did not significantly interact with any of the independent variables, indicating that no asymmetrical learning effects occurred relative to error rate and that our analysis could safely include a within-participant comparison of *display configuration*.

Unlike with reaction time, *group size* had a significant effect on error rate in target-present trials ($F_{2,9} = 15.6$, $p = 0.001$) with groups of four committing fewer errors than pairs who in turn

committed fewer errors than individuals. This finding supports hypothesis H4. The mean error rates during target-present trials for each *group size* were 34.3%, 23.8%, and 16.7% for individuals, pairs, and groups of four respectively. A post-hoc pair-wise means comparison showed a significant difference between all possible pairs of *group size*. Figure 8 shows the target-present error rates for each group size.

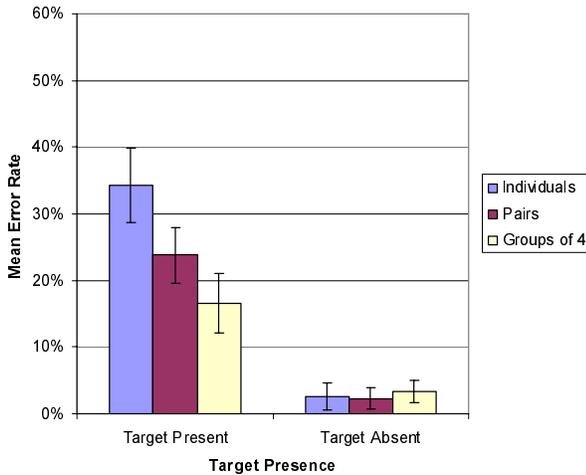


Figure 8. Mean error rates for each target-present condition for each group size.

Contrary to hypothesis H5, *display configuration* had no significant effect on error rate ($F_{2,9} = 0.49, p = 0.62$). Furthermore, our findings supported the rejection of hypothesis H6 as there was no significant interaction between *display configuration* and *group size* ($F_{4,18} = .64, p = 0.64$). Figure 9 illustrates the target-present error rates for each *display configurations* in our study.

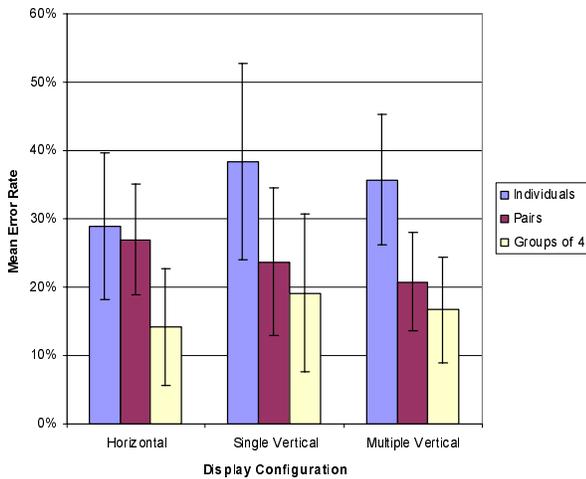


Figure 9. Mean overall target-present error rates for each display configuration for each group size.

Number of distracters had a significant effect on mean error rate for target-present trials ($F_{1,9} = 5.85, p = 0.04$). As one would expect, more distracter objects in a stimulus image led to higher error rates – knives and guns were more easily missed in

cluttered images. Interestingly, there was a significant interaction between *number of distracters* and *group size* ($F_{2,9} = 6.34, p = 0.02$). Individuals appeared less able to deal with additional distracter objects than either pairs or groups of four, whose error rates were nearly identical for stimulus images containing 12 and 16 distracter objects. Figure 10 shows the mean error rate for both *number of distracters* for each *group size*.

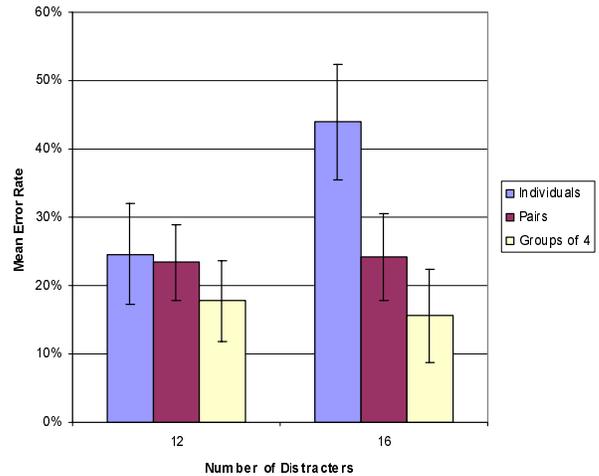


Figure 10. Mean error rate for target-present condition for each number of distracters for each group size.

3.7.2.2 Target-absent Error Rate

None of our independent variables had a main effect on the error rate of target-absent trials, nor were there any interactions among variables. In general, false positive errors were rare. There were slightly more false positive answers in the multiple display configuration than in the other *display configurations*, with mean error rates of 2.6%, 2.5%, and 3.2% for horizontal, single vertical, and multiple vertical respectively. Groups of four committed slightly more false positive errors than other *group sizes*, with means of 2.6%, 2.3%, and 3.3% for individuals, pairs, and groups of four, as shown in Figure 8.

4. DISCUSSION and OBSERVATIONS

A key finding in our experiment was that while larger groups did not lead to faster trials, they did lead to much lower error rates in target present trials. Groups of four averaged almost half as many errors as individuals.

4.1 Comparison to Previous Work

While Wolfe et al. [19] do not report the specific numeric error rates and reaction times for each target prevalence, an estimation of mean error rate can be made from the charts showing error rates and reaction times for stimulus images in the 10% prevalence condition containing 12 and 18 distracter objects. Individual participants in [19] committed errors in about 23% of target-present trials, a number close to half of our 38.4% error rate for individuals. The other major difference between the results of these two studies is apparent in the differences in reaction time. In Wolfe et al.'s study [19], individuals worked on a vertical display. Their mean reaction time averaged roughly 2-3 seconds for stimulus images containing 12-18 distracter

objects; however, our study’s individual participants averaged 5.3 seconds when working on a single vertical display. There are many potential differences between the two studies that might explain these differences, including differences in display quality, stimulus image resolution, number of repetitions, motivation, or simply differences between individuals. Most likely, the time penalty for missed targets that was included in our study design caused participants to move more slowly through the trials as they tried to avoid a lengthy 10 second penalty.

4.2 Single User Issues

Another striking finding was individuals’ poor performance in the multiple vertical display configuration. Individuals spent roughly 30% more time searching stimulus images in the multiple vertical display configuration than in either the single vertical or horizontal configurations (Figure 5). This increase in task time could be acceptable if it were accompanied by a similar decrease in error rate; however, no such decrease was observed (Figure 9). It is widely assumed that multi-monitor systems capable of displaying simultaneous views of, or simply more of, a dataset are beneficial to individuals, and this has indeed been the case in previous studies [1]; however, in this case it appears that the scanning of multiple views only added to the length of the task, not the accuracy with which an individual could complete it. Thus, a penalty is paid without any corresponding gain.

Individuals also seemed less able to contend with larger *number of distracter* objects (Figure 10). While pairs and groups of four had very similar error rates for stimulus images including 12 and 16 distracter objects, individuals’ error rate jumped with the additional clutter. However, as Figure 7 indicates, reaction times for all group sizes increase at about the same rate when the number of distracters goes from 12 to 16. This demonstrates that pairs and groups of four were able to overcome the additional clutter of extra objects (i.e., no error rate penalty) without incurring a corresponding penalty in terms of reaction time.

4.3 Group Collaboration

Prior to running the experiment, we expected that larger groups would perform faster than smaller groups partly because individuals in larger groups might gain confidence that their teammates would catch errors that they made. A few pairs into our experiment, however, it became obvious that our assumption would prove to be false – pairs and groups of four were talking constantly about the task, often pointing toward difficult to identify items in the stimulus images. Of course, no such discussion took place in the single user conditions. The effect of this discussion is illustrated in Figure 7, which shows that groups of four had the slowest reaction times for target-absent trials. It was during these trials that the most discussion about potential targets took place. Conversely, groups of four had the fastest reaction times for target-present trials. Once a group member identified a target object, they entered their answer, announced their finding to the team, which immediately moved onto the next task.

We looked at the order in which participants in groups of four entered their answers, thinking that within each group there may be those individuals who continuously led and/or those who were deferential to the group’s consensus. We found little variation among the average answering order of members within

each of our groups, indicating a relatively equal contribution from each member to the team and no obvious “leaders” or “followers/freeloaders”.

While it was not reflected in the reaction times (Figure 5), we observed that groups of four talked much less during a trial in the multiple vertical than in either the horizontal or single vertical display configurations. Discussion during the multiple vertical conditions tended to happen at the end of a trial, after each group member had spent some amount of time searching the image on the display immediately in front of him/her. While audio from the experiment sessions was not recorded, making a numerical analysis of the amount of talking occurring amongst participants impossible, we were able to look at the spread of reaction times over the members of a group. A shorter duration between the first and last participant’s answers would indicate a high-level of group cohesion, and a longer duration would indicate more independence amongst group members. Figure 11 shows the mean reaction time spread for groups of four for each *display configuration*. An ANOVA using display configuration as a within-participant variable shows a significant difference among the three setups ($F_{1,5} = 9.42, p = 0.028$) for groups of four. A post-hoc pair-wise means comparison shows a significant difference between the single and multiple vertical display configurations, indicating that groups of four answered more closely in time when working on a single display. One could speculate that this temporal cohesion indicates a greater level of group cohesion in the single display condition.

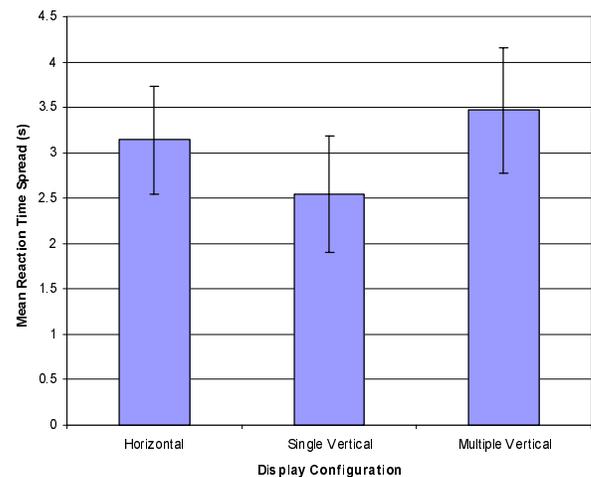


Figure 11. The spread of mean reaction times for groups of four for each display configuration.

Another interesting observation was that the amount of pointing at the screen and the language used to describe objects on the screen changed for groups of two and four across different display configurations. In the horizontal display configuration, the screen was easily within reach of all participants, and participants often pointed at potential targets and asked group members what they thought. This was such a common practice, that the display was covered with fingerprints by the end of each session. Communication among group members was similar in the single vertical condition, although reaching the display was difficult for the two individuals in a group of four sitting on the outside ends of the row of chairs. In this case, participants often referred to object by name as they pointed from afar. We also

observed the use of reference objects to verbally describe locations on the screen, similar to those strategies presented in [11]. Two groups of four came up with their own names for some of the common distracter objects so that they could reference another object on the screen with the distracter name and a direction (e.g. “look under the pigtail”). Problems arose in the multiple vertical display condition in which directions such as “below” had different meanings on different displays. This lack of shared context may explain the differences in the amount of discussion.

Table 1 shows the mean reaction times and error rates for each display configuration for each group size, organized by the continuum of information content and continuum of shared context presented in section 3.3. What is interesting about organizing the results in this way is that it emphasizes that greater information content did not necessarily lead to faster reaction times or lower error rates, especially in the case of individuals who seemed overwhelmed by the information content provided in the multiple vertical display condition.

4.4 Independence Among Group Members

A natural extension to the study presented in this paper is to compare the results of co-located teams, remotely distributed teams, and groups working independently. At first glance, one might expect the chance of a pair working independently to miss a target object to be the square of the probability of an individual missing it. Based on the results from our study, the target-present error rate of an ideal pair would be $(34.3\%)^2 \approx 11.7\%$, and four completely independent individuals would have an error rate of $(34.3\%)^4 \approx 1.4\%$. However, this model assumes independence among events, an assumption that is not met in this case as all group members are searching the same image. Indeed, a look at the mean error rate for each target-present image shows a wide range of difficulties. Even so, the results of our experiment showed far higher error rates than these ideals, suggesting a penalty caused by some quality of the group.

To further explore this issue, we combined the results from our first four individual participants into a *virtual team of four*. While a statistical comparison between virtual groups and actual groups is not possible given the number of individual participants in our study, a quick look at the relative scores provides some insight into the effects of the presence of group members on an individual’s performance. Overall, the virtual

group had a higher error rate for *target-present* trials than our actual groups, with mean error rates of 20.0% and 16.7% for virtual and actual groups of four respectively. This larger error rate was due mainly to errors in the multiple vertical *display configuration*, in which the virtual group had an error rate of 35.0% as opposed to the actual groups’ 16.7% error rate. The virtual group of four committed fewer errors in the single vertical and horizontal *display configurations* than the actual group did (15% vs. 19% and 10% vs. 14% for single display and horizontal display respectively).

Although it is clear that a group is less likely to commit an error than an individual, it appears as though an individual may be more likely to commit an error when they are part of a group than when working on their own. The optimal use of a group may be as independent screeners offering concurrent, separate opinions. Alternatively, remotely distributed teams that have limited means of communication may avoid some of the drawbacks of co-located groups when performing this type of task. Clearly, a more detailed inspection of this issue is needed.

5. CONCLUSION AND FUTURE WORK

While our results should be of interest to anyone involved in designing systems for group searching tasks, important differences exist between the simulated baggage screening task that our study’s participants completed and other visual searching tasks that small groups perform. For example, a team of doctors examining a medial image bring a huge amount of contextual knowledge to the table. Information about the patient’s medical history and symptoms, as well as preconceived diagnoses must influence the group’s search. Similarly, intelligence analysts looking at satellite images will often view a series of images of the same location, leading to a high correlation of features among the images. The searching of images presented later in the sequence is certainly influenced by previously viewed images. While recruiting expert participants for these types of experiments is difficult, a study designed around the specific constraints of the tasks described above would be a valuable piece of future work.

One interesting question that arose during the design of this experiment was that of adding interaction to the search. A baggage screener wishing to get another view of an x-ray might benefit from being able to rotate the picture by 90 degrees, and invert or otherwise filter the image to better examine objects of interest. Similarly, mammograms are often taken from both the

Table 1: Mean reaction times and error rates for each group size organized by the continuum of information content and by the continuum of shared context.

	Amount of Information Content, Target-Present Trials			Amount of Information Content, Target Absent Trials		
	Single Display	Horizontal	Multiple Vertical	Single Display	Horizontal	Multiple Vertical
1	3.8s (38.3%)	3.9s (28.9%)	5.0s (35.7%)	5.3s (2.7%)	5.6s (3.7%)	7.9s (1.4%)
2	4.2s (23.7%)	3.9s (26.9%)	3.5s (20.8%)	5.9s (1.3%)	6.0s (1.9%)	6.7s (3.7%)
4	3.2s (19.2%)	2.7s (14.2%)	3.2s (16.7%)	6.9s (3.4%)	7.0s (2.1%)	7.0s (4.4%)

	Amount of Shared Context, Target-Present Trials			Amount of Shared Context, Target Absent Trials		
	Multiple Vertical	Horizontal	Single Display	Multiple Vertical	Horizontal	Single Display
2	3.5s (20.8%)	3.9s (26.9%)	4.2s (23.7%)	6.7s (3.7%)	6.0s (1.9%)	5.9s (1.3%)
4	3.2s (16.7%)	2.7s (14.2%)	3.2s (19.2%)	7.0s (4.4%)	7.0s (2.1%)	6.9s (3.4%)

top and side providing a doctor with multiple views with which to make a diagnosis. An analyst working with a high-resolution photograph may choose to enlarge an interesting portion of a image or may filter it digitally. The question remains, how might collaboration and performance be affected by the addition of interaction?

While our design included a *horizontal display* condition, most research into *tabletop* interfaces have examined larger display areas than that used in our study. Our aim in using identical monitors in every display configuration was to reduce the complexity of the user study design. Display size and resolution will likely affect group performance. It is interesting to note the similarity of performance for individuals working with the horizontal and single vertical display conditions. In this case, the orientation of the display itself made little difference to perception. It would be a valuable piece of work to investigate larger display sizes used in both a horizontal and vertical orientation. A large horizontal workspace may emphasize the foreshortening and distortion of images that has been shown to reduce performance in visual searching tasks [10].

As part of our ongoing research, we plan to address these and other issues as we continue to explore small group collaboration.

6. ACKNOWLEDGMENTS

This work was partially supported by the Advanced Research and Development Activity (ARDA) and the National Geospatial-intelligence Agency (NGA) under Contract Number HM1582-05-C-0028. The views, opinions, and findings contained in this report are those of the author(s) and should not be construed as an official Department of Defense position, policy, or decision, unless so designated by other official documentation.

The authors would like to thank our study participants, without whom this type of research would not be possible. Additionally, we would like to thank our research colleagues at MERL for their feedback and help in drafting this paper.

7. REFERENCES

- [1] Ball, R. and North, C. Effects of tiled high-resolution display on basic visualization and navigation tasks. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA, April 02 - 07, 2005), 1196-1199.
- [2] Biederman, I. and Gerhardstein, P.C. Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 1162-1182.
- [3] Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, G. and Starkweather, G., Toward Characterizing the Productivity Benefits of Very Large Displays. in *Proceedings of Human-Computer Interaction - INTERACT 2003*, (Zürich, Switzerland, 2003), 9-16.
- [4] Darley, J.M., and Latane, B., Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377-383.
- [5] Grudin, J., Partitioning Digital Worlds: Focal and Peripheral Awareness in Multiple Monitor Use. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (Seattle, Washington, USA, 2001), 458-465.
- [6] Hemera Images. <http://www.hemera.com/>
- [7] Hutchings, D., Smith, G., Meyers, B., Czerwinski, M. and Robertson, G., Display space usage and window management operation comparisons between single monitor and multiple monitor users. in *Proceedings of the working conference on Advanced visual interfaces*, (Gallipoli, Italy, 2004), ACM Press, 32-39.
- [8] Janis, I. Groupthink. *Psychology Today*, 5(6). 43.
- [9] Jolicoeur, P. The time to name disoriented natural objects. *Memory & Cognition*, 13, 289-303.
- [10] Lawson, R., Humphreys, G. and Jolicoeur, P. The Combined Effects of Plane Disorientation and Foreshortening on Picture Naming: One Manipulation or Two? *Journal of Experimental Psychology: Human Perception and Performance*, 26 (2). 568-581.
- [11] Mainwaring, S. D. Tversky, B., Ohgishi, M. and Schiano, D. J. Descriptions of simple spatial scenes in English and Japanese. *Spatial Cognition and Computation*, 3, 3-42.
- [12] North, C. and Shneiderman, B., Snap-together visualization: a user interface for coordinating visualizations via relational schemata. in *Proceedings of the working conference on Advanced visual interfaces*, (Palermo, Italy, 2000), ACM Press, 128-135.
- [13] Roberts, J.C., On Encouraging Multiple Views for Visualization. in *Proceedings of IEEE Symposium on Information Visualization InfoVis'98*, (Research Triangle Park, NC, USA, 1998), IEEE Computer Society, 8-13.
- [14] Ryall, K., Forlines, C., Shen, C. and Morris, M.R., Exploring the effects of group size and table size on interactions with tabletop shared-display groupware. in *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, (Chicago, Illinois, USA, 2004), ACM Press, 284-293.
- [15] Shen, C., Vernier, F., Forlines, C. and Ringel, M., DiamondSpin: an extensible toolkit for around-the-table interaction. in *Proceedings of the SIGCHI conference on Human factors in computing systems*, (Vienna, Austria, 2004), ACM Press, 167-174.
- [16] Steiner, I. *Group Process and Productivity*. Academic Press, New York, NY, 1972.
- [17] Tan, D. and Czerwinski, M., Effects of Visual Separation and Physical Discontinuities when Distributing Information across Multiple Displays. in *Proceedings of OZCHI 2003 Conference for the Computer-Human Interaction Special Interest Group of the Ergonomics Society of Australia*, (Brisbane, Australia, 2003), 184-191.
- [18] Wolfe, J.M. What Can 1,000,000 Trials Tell Us About Visual Search? *Psychological Science*, 9 (1). 33-39.
- [19] Wolfe, J.M., Horowitz, T.S. and Kenner, N. Rare items often missed in visual searches. *Nature*, 435. 439-440.