A snapshot is all it takes to encode object locations into spatial memory

Harry H. Haladjian, Fabien Mathy

A snapshot is all it takes

to encode object locations into spatial memory

Harry H. Haladjian & Fabien Mathy

Research highlights:

- We compared spatial memory accuracy in visual short-term and short-term memory.
- Spatial memory did not improve when stimuli were viewed for longer than 200-ms.
- Spatial compression was reduced on longer displays, but overall error was not.
- No evidence for the grouping of nearby discs into clusters to aid performance.
- Results suggest a fast non-independent extraction of global spatial information.

A snapshot is all it takes to encode objects locations into spatial memory

Harry H. Haladjian [a,b] & Fabien Mathy [a]

[a] Scolarité de Psychologie, Université de Franche-Comté - UFR SLHS

30-32 rue Mégevand, 25030 Besançon Cedex, France

[b] School of Social Sciences and Psychology, University of Western Sydney

Bankstown Campus, Building 24, Locked Bag 1797

Penrith, NSW 2751, Australia

Corresponding Author (current address):

Harry H. Haladjian

University of Western Sydney, School of Social Sciences and Psychology,

Bankstown Campus, Building 24, Locked Bag 1797

Penrith, NSW 2751, Australia

E-mail: h.haladjian@uws.edu.au or haroutioun@gmail.com

Phone: +61.04.5258.0027

Fax: +61.02.9772.6757

Body text word count: 5,769

Running head: Localization and STM

Abstract

This study investigates the accuracy of spatial memory for sets of objects by comparing localization accuracy in short-term memory and visual short-term memory. Observers in the short-term memory condition viewed masked displays containing 1 to 10 discs (1-second per item), and then reported the locations of these discs. Compared to a previous study that presented the same stimuli briefly (50 ms or 200 ms for all items, exposures more typical of visual short-term memory tests), observers were—not surprisingly—better at reporting the correct number of discs but localization accuracy did not improve significantly. Additionally, responses were spread among different clusters and not focused on individual clusters in all exposure durations. These results indicate that spatial information for a set of objects is extracted globally and quickly, with little benefit from extended encoding durations that could favor some deliberative forms of chunking.

Keywords: spatial memory; short-term memory; visual short-term memory; subitizing; chunking; clustering analysis

Snapshot encoding of spatial information:

Location memory for visual-short-term- and short-term-memory exposures

1. Introduction

Visual short-term memory and visual attention both have been studied in order to identify the stages in which perception is faced with capacity limitations. The capacity limit in memory, for example, can affect the quality of simultaneous object representations, where less information about distinct objects can be retained when a greater number of objects must be remembered. A debatable limit of four "slots" for memory has been commonly cited in this literature (e.g., Cowan, 2001; Luck & Vogel, 1997; Zhang & Luck, 2008). The memory for visual information also may be limited by the capacity of attentional processes, such as an individuation mechanism in early vision that can "index" up to four items (Pylyshyn, 1989). This limit of four items is observed in enumeration studies, for example, where counting errors increase substantially for sets larger than four (e.g., Dehaene & Cohen, 1994; Revkin, Piazza, Izard, Cohen, & Dehaene, 2008; Trick, 2008). The fast and error-free counting of up to four items, called "subitizing", is thought to be facilitated by the visual individuation or indexing mechanism (Trick & Pylyshyn, 1989, 1993).

A recent study, however, found a higher subitizing range when observers reported numerosity by marking the locations of a set of items (Haladjian & Pylyshyn, 2011). In this study, observers were shown masked displays with 2-9 randomly-placed discs at very brief durations (50 to 350 ms). After the mask, observers marked the locations of each disc on a blank computer screen. In addition to measuring spatial memory for sets of objects, this reporting method provided a numerosity estimate. Enumeration performance was found to be high for displays with up to six items when using the localization method, but only around four items when using a conventional reporting method with Arabic numerals (the limit typically reported for subitizing). This increased capacity may be attributed to the reporting method, since the act of "pointing" to the locations of the discs may engage a memory involved in motor responses (e.g., Goodale & Milner, 2004). Another possible explanation for the increased capacity is the result of perceptual grouping, where nearby discs are grouped together for more efficient storage in visual short-term memory (Cowan, 2001; Mathy & Feldman, 2012). Effectively, the proximity of two discs can form a cluster and result in non-independent spatial information for the two discs that can be encoded in a more compact way into a single slot in memory, resulting in a greater number of available slots for the encoding of other discs. That information processing systems can overcome capacity limitations whenever a form of relational information can be computed in the data has received particular attention recently in the visual short-term memory literature (Alvarez & Cavanagh, 2004; Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Bays, Wu, & Husain, 2011; Brady, Konkle, & Alvarez, 2009, 2011; Fougnie & Alvarez, 2011; Magnussen, Greenlee, & Thomas, 1996; Olson & Jiang, 2002; Wheeler & Treisman, 2002; Xu, 2002). This hypothesis will be tested using clustering techniques in order to determine whether or not grouping is used by observers to encode more information about objects in the localization task.

A second related question is whether this putative organization in memory operates in a lossless (Mathy & Feldman, 2012) or a lossy manner (Brady & Alvarez, 2011), respectively by allowing the exact original data to be reconstructed from the compressed data or by computing ensemble statistics prior to compressing data (Cowan, 2001; Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999; Feldman, 1999; Jiang, Olson, & Chun, 2000; Mathy & Feldman, 2012; Sargent, Dopkins, Philbeck, & Chichka, 2010). By analogy, these two forms of compression would be popular file formats such as .png/.gzip or .jpeg/.mpeg. Effectively, there is a possibility that exact information about item locations can be compressed so that a greater number of items can be unpacked from a few chunks (Mathy & Feldman, 2012). As it is possible to recall a series of 50 numbers, such as 2-4-6-8-10-12-...100, by retaining the shorter description “pair numbers from 2 to 100”, it is possible to retain the location of several items organized with regularity (e.g., “• • • • • • •”) without loss of the original information. This lossless form of compression would imply a correct report of a limited set of items, with total loss of information for items that could not be encoded because of capacity limitations. An alternative hypothesis is that, due to a spreading of processing resources, the computation of ensemble statistics would tend to bias the representation of all individual items (Brady & Alvarez, 2011) with the prediction that a complete set of items is encoded globally and reported in a distorted fashion. These hypotheses will be tested in the following manner: a greater number of discs are expected to be reported in a less precise manner if the compression process operates in a lossy manner (with no accuracy distinction in the reports of all discs, as they all would be encoded using equal resource allocation), whereas fewer discs should be reported more precisely using a lossless compression process (with better recall of the first individual discs or the first group of discs for which a limitation in capacity would not be reached, and particularly bad recall of a subset of discs for displays that would exceed capacity).

A third question we addressed was whether or not a grouping of spatial information occurs on these displays for short and long exposures. We expected long viewing durations to enhance more deliberate chunking processes for encoding location information into short-term memory. Greater use of grouping with longer exposures would demonstrate that top-down factors contribute to improve the storage of spatial information in memory. Instead of constraining the spreading of processing resources at short durations, longer exposure durations could allow more flexible storage of items, especially for those items that form groups.

To investigate these questions, we examined chunking strategies in spatial memory using a task that required observers to remember object locations on displays containing randomly-placed discs, and we compared performance for visual short-term memory and short-term memory exposures. Observers viewed masked displays containing 1 to 10 discs using typical short-term memory exposures for measuring spans, that is, 1-second per item (for instance, used in the spatial span task by Lewandowsy, Oberauer, Yang, & Ecker, 2010, although their presentation of the stimuli is serial). We compared these results to the previous studies that presented the same stimuli for 50 ms or 200 ms for all items – exposures more typical of visual short-term memory displays (data from Haladjian & Pylyshyn, 2011; Haladjian, Singh, Pylyshyn, & Gallistel, 2010). We examined whether grouping of objects by proximity occurred and whether perceptive grouping or deliberative chunking strategies were used to create more compact representations. We expect that with longer viewing exposures, intentional chunking strategies may be used to both increase the accuracy of localization performance and to correlate with a more lossless form of compression than with shorter exposures.

Our results show that localization accuracy did not improve significantly with longer exposures. Additionally, responses were spread among different clusters and not focused on individual clusters. We discuss why this indicates that spatial information for a set of objects is extracted globally and quickly, with little benefit from extended encoding durations.

2. Methods

2.1 Participants

Thirty-six students and staff members from the Université de Franche-Comté were recruited for voluntary participation in this experiment during the summer trimester (2011); no payment was given. Two participants were removed from the analyses due to high variability in performance.

2.2 Apparatus

The experiment was programmed in MATLAB using Psychophysics Toolbox extension (Brainard, 1997). The stimuli were presented on a laptop running Windows XP, with a 38 x 22 cm LCD screen (1600 x 900 pixels & 60 Hz refresh).

2.3 Stimuli

The test stimuli were identical to those used in Haladjian & Pylyshyn, 2011, and comprised of 1-10 identical dark gray discs on a slightly lighter gray background (~1° viewing angle in diameter). The discs were randomly placed within a 22° x 13° region in the center of the laptop screen with a minimum distance of ~3° between discs (to minimize crowding). The discs were presented simultaneously for 1-10 seconds at a rate of 1-second per disc (e.g., a display with 2 discs was presented for 2 seconds, and a display with 9 discs was presented for 9 seconds). These low-contrast stimuli were designed to minimize after-images and optimize the effectiveness of the subsequent random-dot texture mask that was presented for 1 second.

2.4 Procedure

Each trial began with a 2-second presentation of a blank gray screen with a central fixation cross. The disc stimulus then appeared for the designated duration based on the number of discs on the stimulus (stimulus durations ranged from 1-10 seconds). This display was followed by a 1-second random-dot texture mask to limit viewing durations and prevent after-images. Finally, a blank gray screen appeared and observers used a computer mouse to place markers (“X”) on each of the perceived disc locations; then, they pressed the space bar to initiate the next trial at their own pace.

See Figure 1 for a schematic of a trial in this experiment. The program presented 10 trials for each of the display numerosities in a randomized order, for a total of 100 trials. The experimental session lasted less than 30 minutes.

2.5 Analyses

Enumeration accuracy was determined by comparing the number of discs on a stimulus display to the number of markers placed on the response display. For localization accuracy, stimulus-response pairs were established using Delaunay triangulation and nearest-neighbor methods in MATLAB. The triangulation process basically connects the nearest-neighbor discs in an optimal manner and creates a triangular mesh, where no discs fall inside another triangle and avoids long or extremely thin triangles (see Appendix Figure A for an example). Each of the response markers are individually added to this process to find the most likely stimulus disc and response marker correspondence (based on distance). Once a response marker was paired with a stimulus disc, the Euclidian distance between these two locations was computed to provide an estimate of the localization error. Also, the Delaunay triangulation procedure was performed on all the response coordinates, and the lengths of the triangle segments produced by this method were used to examine the localization errors in the responses (i.e., to identify spatial compression by comparing the average distances among the stimulus discs to the average distances among the responses). In all the analyses of variance (ANOVA) models reported, subject ID was included as a random factor to control for between-subject variability.

Since the current design only differed by using longer presentation durations than those used in previous similar studies (Haladjian & Pylyshyn, 2011; Haladjian, et al., 2010), we compared the results from all these studies to determine if increased exposure to the test stimuli improved task performance. That is, we wanted to measure the benefit of increasing the presentation durations from 200-ms or less to several seconds for displays with multiple discs.

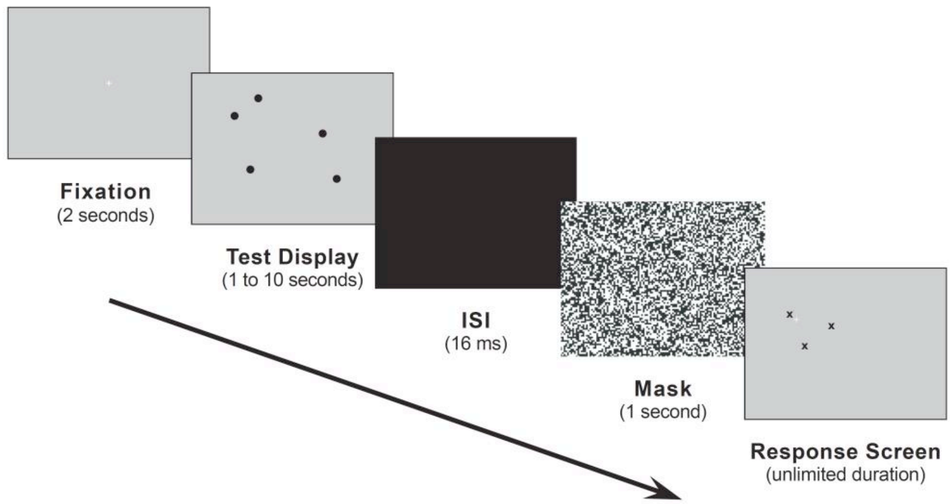

Figure 1. Schematic of the experimental design for the localization task.

3. Results

3.1 Enumeration Accuracy

The proportion of trials with the correct number of markers placed on the screen was very high in this experiment—an expected result since displays were presented long enough to allow observers to freely count all of the discs as they encoded their locations (in the debriefing, all participants reported that they counted the discs most of the time during the experiment). Nevertheless, there were more counting errors on displays with more than six items. These errors increased at a much lower rate than the results from the previous studies that used very brief presentation durations (Haladjian & Pylyshyn, 2011; Haladjian, et al., 2010). Figure 2 plots the proportion of trials that were enumerated correctly as a function of display numerosity for 50-ms displays (very short exposure), 200-ms displays (short exposure), and displays lasting for one-second or more (long exposure).

An ANOVA comparing average proportion of trials with correct responses was conducted, with duration and numerosity as fixed factors and the subject ID as a random factor (to correct for between-subject differences). The ANOVA results indicated significant main effects for numerosity (F(7,1737) = 428.3, $p < .001$, $\eta^2_p = .63$) and duration (F(1,152) = 263.4, $p < .001$, $\eta^2_p = .63$), with an interaction (F(7,1065) = 43.2, $p < .001$, $\eta^2_p = .22$). Additionally, for each numerosity, an ANOVA was performed to examine differences between the duration conditions. The ANOVAs for duration were significant for all numerosities except 3 to 5, with post-hoc analyses indicating significantly different performance among all three duration conditions for numerosities of 6 and greater (F's > 15, $p < .001$). These results indicate, not surprisingly, that longer viewing durations allow for more accurate recall of the number of items present on a stimulus display by providing observers enough time to count the number of discs while encoding locations. The primary interest of this study, however, is the performance on the localization of these discs, which will be the focus of the remainder of the analyses.

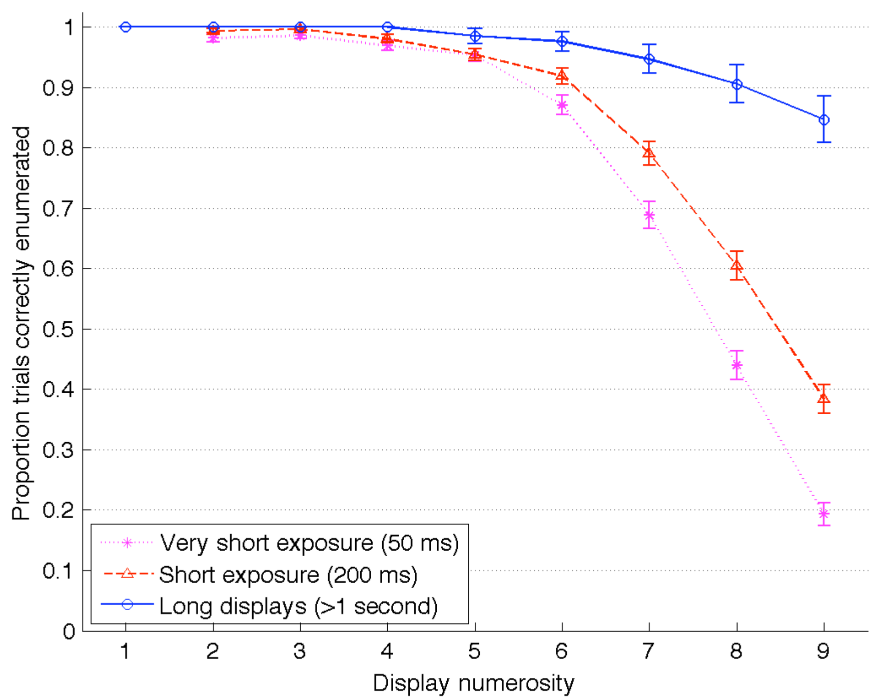

Figure 2. Proportion of trials with correct enumeration; long exposure ($N = 34$) and short /very short exposures (for both, $N = 152$; data collected in Haladjian & Pylyshyn, 2011 and Haladjian, et al., 2010). Note: Error bars represent 95% confidence intervals.

3.2 Localization Accuracy

Localization accuracy was measured as the distance between paired response markers and stimulus discs. Spatial information was expected to be better encoded for as many as four items given the expected short-term memory capacity during these longer viewing durations, but possibly for more than four items given the hypothesis that longer exposures would allow chunking to occur. Performance for longer exposures, therefore, should translate into a flat curve with lower localization errors for display numerosities of up to four items if no chunking occurred, or more items if some could be chunked. The results did not indicate such trends: there was an overall average error of 60 pixels (or ~2° viewing angle) and a regular logarithmic increasing magnitude of errors as the display numerosity increased. Figure 3 plots these results along with results from the previous studies with 50-ms and 200-ms exposures.

The ANOVA examining the magnitude of localization errors indicated significant main effects for numerosity ($F(7,1917) = 148.4$, $p < .001$, $\eta^2_p = .35$) and duration ($F(1,168) = 82.9$, $p < .001$, $\eta^2_p = .33$), but with no interaction ($F(7,1130) = 1.3$, $p = .254$, $\eta^2_p = .01$). ANOVAs comparing the duration for each numerosity separately found significant differences in performance for numerosities 2 to 7, with the very short duration (50-ms) being significantly worse than the other two durations for numerosities 2 to 6, and the short (200-ms) condition being slightly better than the other two durations for numerosity 7. In other words, the localization error was globally highest in the shortest exposure, but there were few differences between the short and long exposures. This indicates that errors in spatial memory depend more on the number of items that need to be processed rather than the duration of exposure to the stimuli, with localization accuracy almost at ceiling after 200-ms of exposure. These results suggest that, as long as a few hundred milliseconds are allowed for observing the displays, spatial information can be encoded quickly and globally, but not sequentially, since there was no increase in localization accuracy for the long viewing condition where observers had time to visit all locations individually.

The magnitude of localization errors also was examined only for the first response made on a display (Figure 4) in order to analyze localization errors without the cumulative increase in errors made with each subsequent click when multiple responses were required (i.e., any display with more than one item). This analysis particularly intended to test whether similar encoding quality for all discs occurs by spreading a global resource among them. The alternate prediction was that a deliberate form of chunking would result in a better allocation of the available resource for serially encoding the items, resulting in a lossless compression of the first chunks or the first items in memory. The ANOVA examining the magnitude of localization errors for the first response made on the display indicated a similar trend as above, with a significant main effect for numerosity ($F(7,1922) = 37.4$, $p < .001$, $\eta^2_p = .12$) and duration ($F(1,202) = 49.6$, $p < .001$, $\eta^2_p = .20$), with no interaction ($F(7,1182) = 0.6$, $p = .64$, $\eta^2_p = .01$). ANOVAs comparing the duration for each numerosity separately found significant main effects of duration for numerosities 2 to 8, with the very short 50-ms duration being significantly worse than the 200-ms condition for all numerosities and worse than the long exposure only for numerosity 2. There were fewer cases in the long exposure condition (data from only 34 subjects), and thus produced larger error bars and limits the interpretation of the results for the longer exposures, but the trend seems to be closer to the 200-ms than the 50-ms condition. These results again suggest that spatial information tends to be encoded quickly and globally (not sequentially) because there was no decrease in localization errors in the long viewing condition, which was also evident in the first response made in each trial.

Additionally, ANOVAs for each duration separately indicated a main effect of numerosity for all exposure durations, with post-hoc analyses indicating that the performance for numerosities of 4 and higher was significantly worse than for smaller numerosities. This also indicates that errors in spatial memory depended more on the number of items that needed to be processed rather than the duration of exposure to the stimuli, with localization accuracy almost at ceiling after 200-ms of exposure. Since this trend in errors was true even for the first response made on the screen, it suggests that the reproduction of locations for making these responses is obtained from an “ensemble” statistic or some sort of non-independent representation of disc locations.

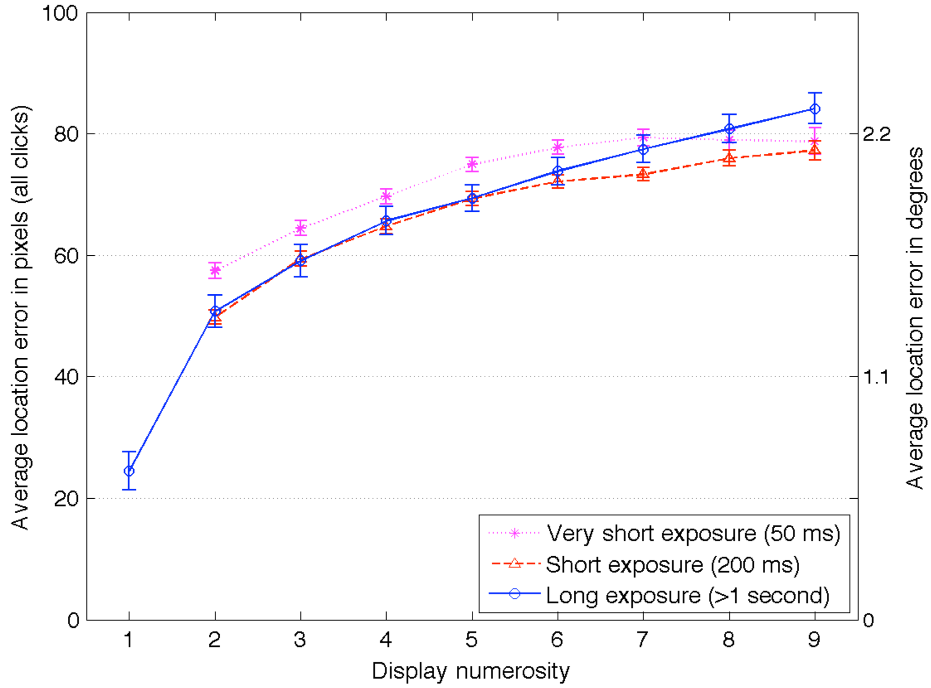

Figure 3. Average localization errors for all responses for long and short exposures. Note: Error bars represent 95% confidence intervals.

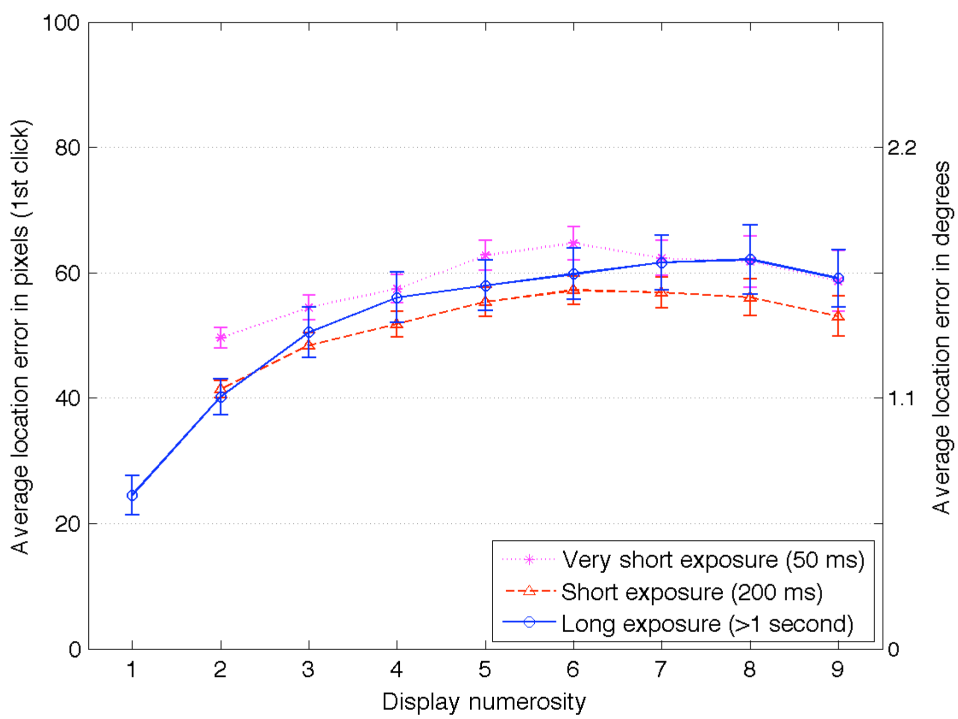

Figure 4. Average localization errors for the *first response* made on a display, for long and short exposures. Note: Error bars represent 95% confidence intervals.

3.3 Compression Effects

We observed less compression of spatial information for trials with longer encoding durations. In these trials with longer exposures, the responses did not show as much shortening in the lengths of the Delaunay triangle segments as in the shorter exposure trials. As previously described, the triangulation was performed for stimulus displays and subject responses (separately), which created an optimal triangular mesh of connections. Shorter Delaunay triangle segments in the responses (compared to the stimulus) would indicate that observers reported disc locations as being closer to each other than they actually were. This systematic compression of space was seen in the localization data from previous studies where the shortest exposure duration produced the most compression (Haladjian, et al., 2010). Figure 5 plots these compression effects as the average lengths of the Delaunay triangle segments on the stimulus displays *minus* the length of these segments on the response displays. A smaller value on Figure 5 indicates a more accurate representation and reproduction of the distances between objects. For displays with more than four items, we see a decrease in this compression as the viewing duration increases. In other words, the localization errors, though similar in magnitude, did not take the more systematic form of spatial compression in the longer viewing condition. The sort of error that did occur may be similar to a rigid shift of coordinates where distance between items was more accurate, but the items were misplaced on the screen. (See Appendix Figure A for an example of this compression in an observer's response.)

An ANOVA was performed on this compression measure and found significant main effects for numerosity ($F(6,1563) = 49.4$, $p < .001$, $\eta^2_p = .16$) and duration ($F(1,152) = 28.9$, $p < .001$, $\eta^2_p = .16$), with an interaction ($F(6,915) = 3.0$, $p < .01$, $\eta^2_p = .02$). Separate ANOVAs for each numerosity indicate significant main effects of duration for all numerosities except 7 and 9. Post-hoc analyses showed significantly less compression for the longer viewing duration for the numerosities 5 to 9. The compression-like errors tended to decrease with larger numerosities likely because displays with fewer discs have larger distances between them (and thus more room for error). Alternatively, observers may have a more accurate memory for the overall spread of items on the stimulus displays when viewed for longer durations, suggesting a slightly better encoding process when there is more time to view the displays (although still prone to the same amount of errors on average).

Overall, these results can be interpreted as an improvement in the spatial representation when the stimuli were viewed for longer durations, though this is a minor benefit (a decrease in compression of ~20 pixels on average, or 0.75 degrees, and a small effect size of 16% of the variation in performance due to the duration conditions). The observation that performance was both less compressed and similarly prone to localization errors in the longer duration seems paradoxical but can in fact be explained by a more exact report of the disc patterns (resulting in less compression), although the whole patterns were not correctly reported on the screen (resulting in localization errors). The average localization errors were therefore comparable in the short and long conditions, but qualitatively different. This supports the possibility that items recalled from memory with shorter encoding time are more prone to compression in memory, especially since extended response times affect the ability to maintain items in visual short-term memory and makes them more susceptible to such compression errors (Sheth & Shimojo, 2001).

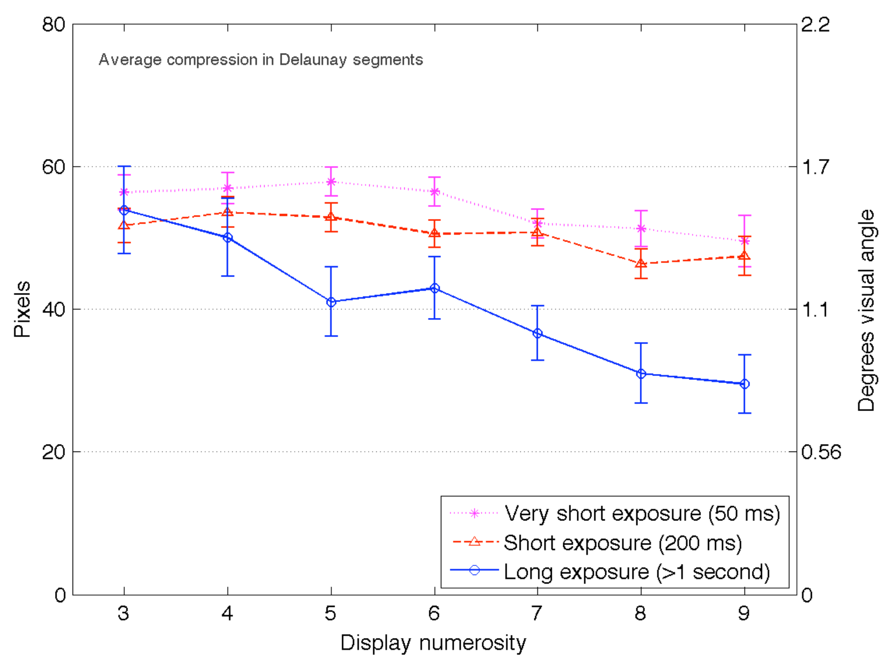

Figure 5. Compression errors based on the computation of Delaunay triangulation segments for long and short exposures. Note: The numbers indicate the average lengths of the Delaunay triangle segments

on the stimulus displays *minus* the length of these segments on the response displays; a smaller value indicates a more accurate representation. Error bars represent 95% confidence intervals.

3.4 Grouping Analysis

To examine whether or not grouping strategies were used to aid spatial memory, the discs on the stimulus displays were grouped into four regions using *k*-means clustering methods (for related topics, see Pothos & Chater, 2002). We set *k* to four, following the idea discussed in our introduction that working memory capacity is limited to four slots, and we hypothesized that the encoding process would likely rely on any available clusters that could be used to maximize encoding efficiency. This clustering method designated a disc's membership to one of four cluster regions on a display based on the mean distances between the discs – a process that identified the discs most likely to be grouped together based on proximity (we only analyzed trials containing five or more discs). Each observer response was then assigned to one of these four regions by associating it with the nearest centroid of a cluster within a region (see Appendix Figure B for an example of the clustering procedure). The results indicated that observers were highly likely to make at least one response in each of these four cluster regions, but some of the items within the regions could be missed (see Figure 6). The ANOVA on the proportion of trials with missed clusters indicated significant main effects for numerosity ($F(4,1246) = 11.2$, $p < .001$, $\eta^2_p = .04$) and duration ($F(1,160) = 29.8$, $p < .001$, $\eta^2_p = .16$), without an interaction ($F(4,639) = 2.2$, $p = .07$, $\eta^2_p = .01$). To further examine the effects of duration, ANOVAs for the individual numerosities were conducted and indicated significant main effects of duration for the numerosity conditions of 5, 6, and 7. Post-hoc analysis controlling for multiple comparisons only found a significant difference in the numerosity condition of 6, where the 50-ms condition had significantly more missed clusters than the 200-ms condition (mean difference = 0.04, $p < .001$).

Additionally, separate ANOVAs on the proportion of trials with missed clusters for each duration found main effects for numerosity in the very short ($F(4,638) = 8.9$, $p < .001$, $\eta^2_p = .05$), short ($F(4,639) = 5.7$, $p < .001$, $\eta^2_p = .03$), and long durations ($F(4,132) = 3.2$, $p = .02$, $\eta^2_p = .09$). Post-hoc analysis found significantly more missed regions: in the very short exposures for numerosity of 5 when compared to 7, 8, and 9, as well as numerosity of 6 compared to 8; in the short exposure, the numerosity of 5 had significantly more missed regions than all other numerosities; and in the long condition, there were more missed regions in numerosity of 5 than of 8. One possibility to keep in mind is a missed region may correspond to a missed item from the display (a miscounted trial) or may be the result of region assignment errors due to mis-localization errors. Nevertheless, although there may be some errors in the *k*-means clustering procedure, these errors should be uniform among the duration conditions and any differences in performance should still be evident.

These results indicate a global encoding of spatial locations instead of a sequential encoding of locations: in over 90% of trials with five or more objects on the screen, there was at least one response in all of the clusters. In other words, the observers noticed that something was present in each of the four cluster regions but, as we will discuss in the next section, they did not remember precisely the exact content within these regions. An opposite strategy favoring the intentional grouping of objects would have tended to show a greater proportion of missed regions given that the observer would have targeted a sequential report of the clusters, better encoding a first subset of objects, and having less room for the last objects given a hypothetical limitation in the number of slots available in short-term memory.

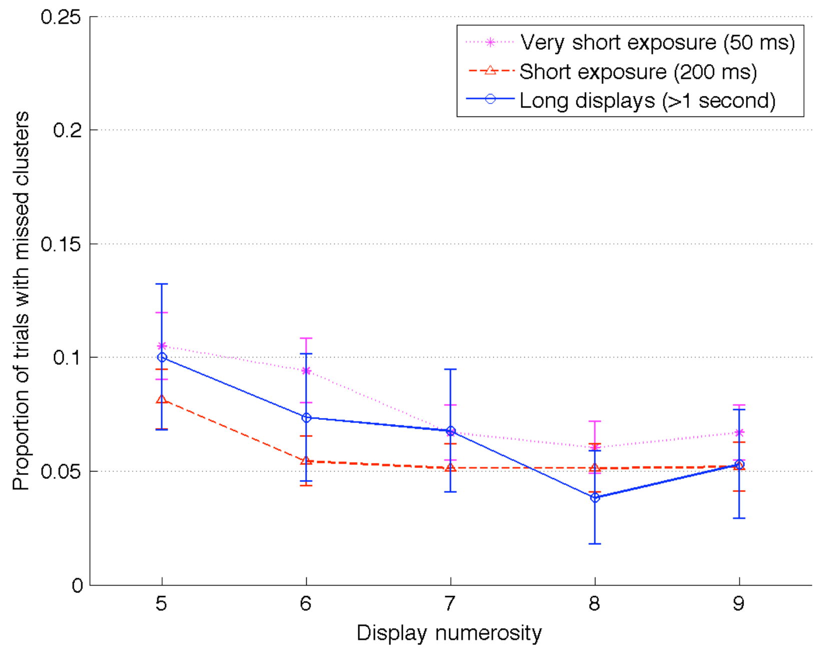

Figure 6. Proportion of trials with any cluster missed (by display numerosity on x-axis) for long and short exposures. Note: Error bars represent 95% confidence intervals.

4. Discussion

While some studies have focused on whether there exists two distinct processes for enumerating numerosities within and beyond the putative subitizing range of four items (e.g., Burr, Turi, & Anobile, 2010), the present study rather focused on whether there exists different processes for enumerating numerosities by examining longer exposure time to the stimuli. Because the present study described the accuracy of the spatial representations of a small set of discs in short-term memory with durations that offered the observers enough time to count the number of items in the field of the view, we focused our analysis on the memory for spatial information and we compared the new data to previous data where similar performance was measured on more rapid presentations typical of visual short-term memory experiments. Our analyses also examined how clustering can affect encoding and whether encoding is subject to a compression process that can become less lossy with greater encoding time. The rationale was that perceptual grouping allows items to be aggregated into larger structures (Feldman, 2007), which tends to free space in memory. Our hypothesis was that short-term memory displays (longer exposures) would enhance a sequential encoding of the available clusters, resulting in more precise memorization of spatial information, especially for the first encoded items. By opposition, we hypothesized that more rapid displays would encourage a global encoding of spatial information resulting in lossy compression of the spatial information.

The current findings provide no evidence for a sharp discontinuity in the encoding of spatial information between exposure durations. The results indicate that the encoding of spatial information occurs quickly and does not benefit from extended exposure to the stimulus. Contrary to our expectations, the longer viewing durations in this study, which provided ample opportunity to encode locations into short-term memory, did not improve localization performance from the previous studies using shorter viewing durations (Haladjian & Pylyshyn, 2011; Haladjian, et al., 2010). This suggests that there is no substantive optimization of spatial memory for longer durations that *a priori* offers more time for the effortful encoding of spatial information. Observers tend to encode the global spatial properties and not individual items or individual chunks of clusters – even when the number of items to recall was below the capacity of short-term memory (around four, if we refer to the theories mentioned in our introduction). This implies that the resource for processing spatial information is distributed across all items rather than divided into slots dedicated to encoding a few items almost perfectly. This result also supports the idea that the difficulty observed in the Corsi block task depends more on other factors than coding spatial locations (Gmeindl, Walsh, & Courtney, 2011), since spatial locations do not seem to be encoded serially and independently in short-term memory.

Additionally, the discs on the stimulus displays were not clearly clustered into subsets and encoded as separate groups into short-term memory. When dividing the stimulus displays into four regions using *k*-means clustering, it was evident that observers tended to report that something was present in all regions of the display, even when making errors as to the precise number of items present within each region. The results indicate that there is no grouping effect for short or long viewing durations, but rather there is a global encoding that is influenced by the overall spread of items on the display. Since the stimuli were designed to avoid crowding by maintaining a distance of at least 3° between the discs (Bahcall & Kowler, 1999; Intriligator & Cavanagh, 2001), as opposed to previous manipulations in which the items were organized in accordance with Gestalt grouping principles (e.g., Woodman, Vecera, & Luck, 2003), these results where some items were missed within regions cannot be attributed to crowding per se, but may be related to the attentional limits that occur under brief presentation durations.

One way to interpret the limits on localization accuracy in this study is to consider location as an object feature, and with more objects there are more features to remember. Localization accuracy might decrease as a function of the number of objects because of the required time for attention to move to each object and encode the relevant features. When the visibility of these displays is limited and there are more items than can be automatically individuated (e.g., via visual indexing), one may notice that there are an uncountable number of objects in addition to the four that were indexed. In order to accurately localize all the objects on such a display, one needs to encode each location into short-term memory and thus will face the limit imposed by the number of available “slots” in short-term memory. This may not be the case in our study, however, since localization accuracy was essentially the same whether the observer viewed the display for short durations (200 ms) or much longer durations (> 1000 ms). It does not seem that longer exposures to the stimuli can enhance memory for spatial information than what is encoded within the 200-ms exposures. Therefore, it is likely that an ensemble statistic is computed of the average spread (and possibly configuration) of the objects on the display rather than an encoding of individual locations. Additionally, it does not seem that a short-term memory limitation is at play in terms of “fixed slots” since there is an increase in localization errors for each numerosity within the four slot limit suggested by such theories. If each object location was assigned to a memory slot, there should be no performance difference within the capacity of short-term memory capacity (up to four items)—contrary to our results.

Overall, the current results do not seem to support a fixed slot theory of short-term memory (Zhang & Luck, 2008) for spatial memory because the encoding of spatial information is not clearly allocated for each viewed item or groups of items, but rather encoded on a more global scale (e.g., as relationships between objects). That localization accuracy seems to be near optimal in as little as 200-ms suggests that an overall snapshot of locations is extracted quickly and is used as the primary guide for localization. This seems to suggest that there is a hierarchical encoding of scene features (Brady & Alvarez, 2011), where the location of objects is a global property that is encoded first, separate from individual features. Further studies, however, are required to better understand the relationship between global and individual object representations, especially in terms of spatial memory.

A possible limitation of the current stimuli is that they were designed to prevent crowding, as mentioned earlier, with a minimum distance of ~3° between the discs. This may have reduced the presence of cluster-able groups of discs, and thus made it more difficult to employ a chunking strategy by the observers. It may be useful to conduct a future study that carefully controlled the appearance of clusters rather than relying on the random clusters created by the current experimental design. Additionally, a reason why localization may have failed to improve in this study is due to the lack of landmarks to aid spatial memory, which have been shown to increase localization accuracy in previous studies (e.g., Lee, Shusterman, & Spelke, 2006). This may also account for the increase in localization error found with each additional response made, as there is no stable frame of reference to constrain such errors. Additional versions of this experiment using landmarks to guide localization may reveal further useful information about spatial memory.

5. Conclusion

In a localization task with very short exposures to the stimuli, Haladjian & Pylyshyn (2011) found a subitizing range in visual-short-term memory that was close to the capacity of short-term memory (Miller, 1956). The current study simply questioned how spatial information would be encoded using a more appropriate timing for measuring short-term capacity (i.e., 1-second per item). In sum, the results from the current study suggest that the greater subitizing range observed in previous studies (Haladjian & Pylyshyn, 2011; Haladjian, et al., 2010) is not likely due to grouping strategies used to aid information processing capacity. Most spatial information is encoded globally in a "snapshot", and not from simply fitting the information for each separate item into a limited number of slots in short-term memory. The resulting lossy-type errors support the possibility that object locations are not encoded in representations independently of each other (see Brady & Alvarez, 2011). Since no evidence for chunking was found, further research is required to determine the cause of the increased subitizing range when reporting numerosity by pointing to locations.

Acknowledgments

This research was supported by a grant from the Agence Nationale de la Recherche (Grant # ANR-09-JCJC-0131-01) awarded to Fabien Mathy. This research was conducted during the summer of 2011 at the Université de Franche-Comté thanks to a postdoc grant awarded to Harry H. Haladjian by the Université de Franche-Comté.

Appendix

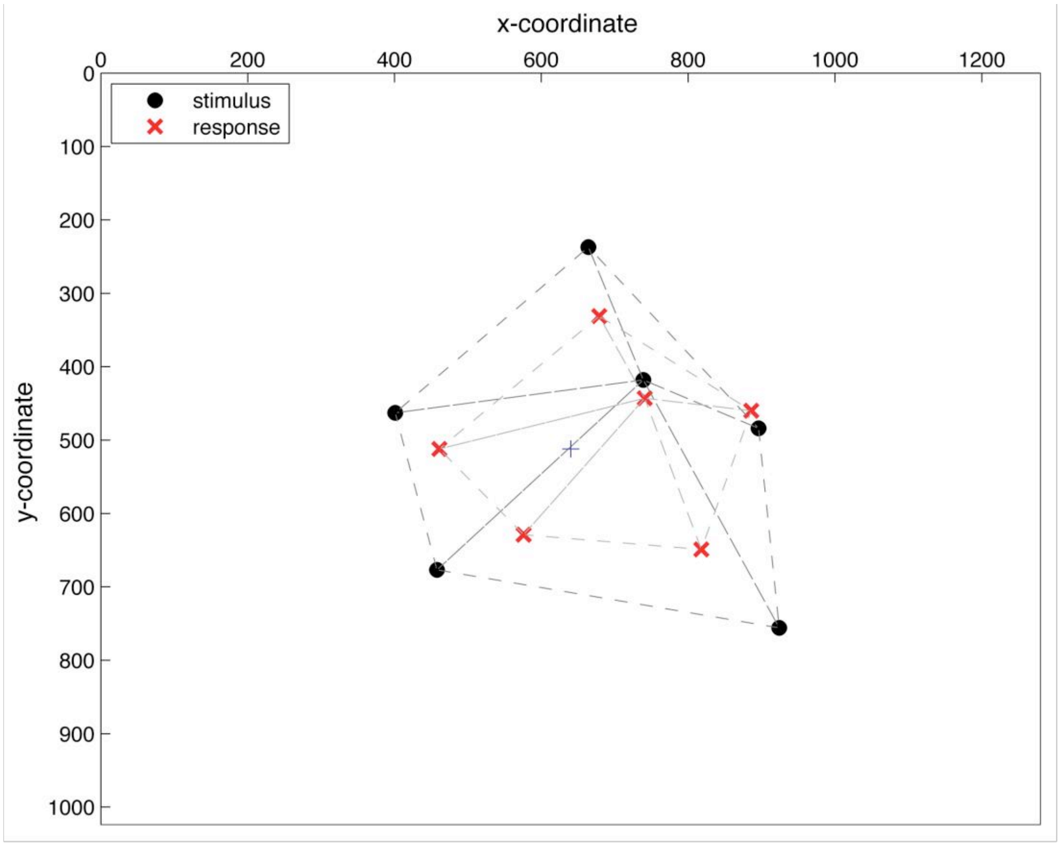

Figure A. Example of Delaunay triangle simplexes and compression effects. The black circles correspond to the stimulus discs and the lighter crosses correspond to the responses from a sample trial; connections between the circles or crosses are the segments identified from the Delaunay triangulation procedure. As this image shows, the distances between the responses are closer to each other than the distances between the stimulus dots.

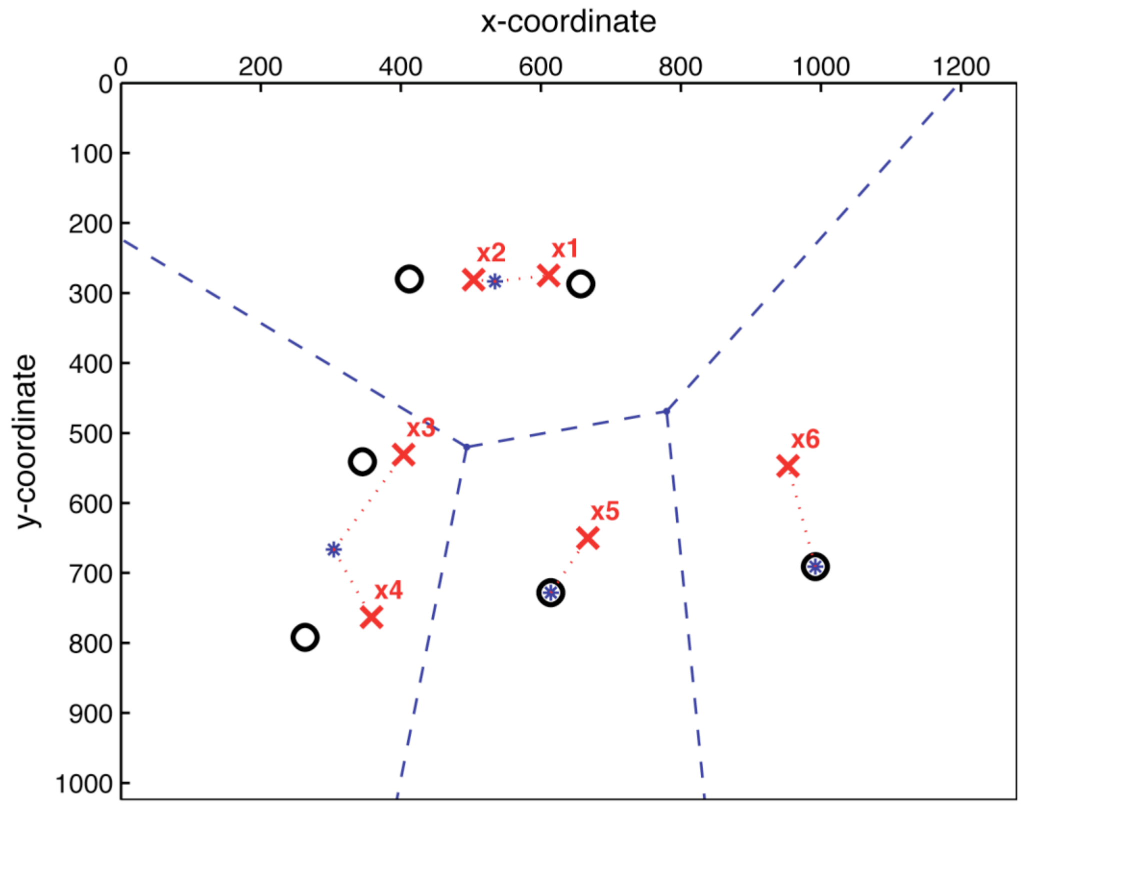

Figure B. Example of regions created by the *k*-means clustering for a single trial, where *k* = 4. The black circles represent the locations of the stimulus discs on a display from a sample trial (*x* and *y* axes correspond to screen dimensions). The grey crosses correspond to the observer's responses. To determine whether or not a response was made within a region, each response was paired with the nearest centroid of the cluster in a region (designated by an asterisk; the dashed line indicates to which region centroid a response was linked). In this trial, a response was made in all four regions.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*(2), 106-111.

Bahcall, D. O., & Kowler, E. (1999). Attentional interference at small spatial separations. *Vision Research, 39*(1), 71-86.

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(10), 7 1-11.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*(5890), 851-854.

Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia, 49*(6), 1622-1631.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science, 22*(3), 384-392.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138*(4), 487-502.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11*(5), 4.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Burr, D. C., Turi, M., & Anobile, G. (2010). Subitizing but not estimation of numerosity requires attentional resources. *Journal of Vision, 10*(6), 1-10.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences, 24*(1), 87-114; discussion 114-185.

Dehaene, S., & Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: Neuropsychological evidence from simultanagnosic patients. *Journal of Experimental Psychology: Human Perception & Performance, 20*(5), 958-975.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neuropsychologia, 37*(10), 1189-1199.

Feldman, J. (1999). The role of objects in perceptual grouping. *Acta Psychologica, 102*(2-3), 137-163.

Feldman, J. (2007). Formation of visual "objects" in the early computation of spatial relations. *Perception & Psychophysics, 69*(5), 816-827.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision, 11*(12).

Goodale, M. A., & Milner, A. D. (2004). *Sight unseen: An exploration of conscious and unconscious vision*. Oxford: Oxford University Press.

Gmeindl, L., Walsh, M., Courtney, S. M. (2011). Binding serial order to representations in working memory: A spatial/verbal dissociation. *Memory & Cognition, 39*(1), 37-46.

Haladjian, H. H., & Pylyshyn, Z. W. (2011). Enumerating by pointing to locations: A new method for measuring the numerosity of visual object representations. *Attention, Perception, & Psychophysics, 73*(2), 303-308.

Haladjian, H. H., Singh, M., Pylyshyn, Z. W., & Gallistel, C. R. (2010). The encoding of spatial information during small-set enumeration. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2839-2844). Austin, TX: Cognitive Science Society.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*(3), 171-216.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*(3), 683-702.

Lee, S. A., Shusterman, A., & Spelke, E. S. (2006). Reorientation and landmark-guided search by young children: Evidence for two systems. *Psychological Science, 17*(7), 577-582.

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. (2010). A working memory test battery for MATLAB. *Behavior Research Methods, 42*(2), 571-585.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279-281.

Magnussen, S., Greenlee, M. W., & Thomas, J. P. (1996). Parallel processing in visual short-term memory. *Journal of Experimental Psychology: Human Perception & Performance, 22*(1), 202-212.

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition, 122*(3), 346-362.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception & Psychophysics, 64*(7), 1055-1067.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26*(3), 303-343.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the first spatial-index model. *Cognition, 32*(1), 65-97.

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science, 19*(6), 607-614.

Sargent, J., Dopkins, S., Philbeck, J., & Chichka, D. (2010). Chunking in spatial memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 36*(3), 576-589.

Sheth, B. R., & Shimojo, S. (2001). Compression of space in visual memory. *Vision Research, 41*(3), 329-341.

Trick, L. M. (2008). More than superstition: Differential effects of featural heterogeneity and change on subitizing and counting. *Perception & Psychophysics, 70*(5), 743-760.

Trick, L. M., & Pylyshyn, Z. W. (1989). Subitizing and the finst spatial index model. *Bulletin of the Psychonomic Society, 27*(6), 490.

Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception & Performance, 19*(2), 331-351.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131*(1), 48-64.

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review, 10*(1), 80-87.

Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception & Performance, 28*(2), 458-468.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233-235.