Time Window from Visual Images to Visual Short-Term Memory: Consolidation or Integration?

Yuhong Jiang

Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract. When two dot arrays are briefly presented, separated by a short interval of time, visual short-term memory of the first array is disrupted if the interval between arrays is shorter than 1300–1500 ms (Brockmole, Wang, & Irwin, 2002). Here we investigated whether such a time window was triggered by the necessity to integrate arrays. Using a probe task we removed the need for integration but retained the requirement to represent the images. We found that a long time window was needed for performance to reach asymptote even when integration across images was not required. Furthermore, such window was lengthened if subjects had to remember the locations of the second array, but not if they only conducted a visual search among it. We suggest that a temporal window is required for consolidation of the first array, which is vulnerable to disruption by subsequent images that also need to be memorized.

Key words: visual short-term memory, visual integration, short-term consolidation

Introduction

A recent study conducted by Brockmole, Wang, and Irwin (2002) convincingly demonstrated that a long time window, on the order of 1300–1500 ms, is needed to represent a visual image in a stable form in visual short-term memory (VSTM). This finding was revealed in a clever, temporal integration task (Di Lollo, 1980). On a 5 by 5 grid, Brockmole et al. presented two dot arrays, each occupying a different set of 12 cells, such that when overlaid on the grid, the arrays would cover the entire grid except one cell. Upon the presentation of the second dot array, subjects were asked to report which cell was the empty one. Subjects rarely made errors in mistaking the second array as the empty cell, and such errors, when committed, were unaffected by the interstimulus interval (ISI) between the two arrays. However, subjects often mistook cells of the first array as the empty cell. These errors were infrequent when the ISI was zero, but increased dramatically for ISI 100 ms or greater, and gradually leveled off for ISI beyond 1300–1500 ms. Performance asymptoted at around 1500 ms. Such a long temporal window before asymptote is surprising, given that visual sensory memory of the first array lasts only about 100 ms (Phillips, 1974; Sperling, 1960). A process that requires a long temporal lag must not be tied to sensory memory, but what is it? Brockmole et al. believed that this process was the temporal integration between the two visual arrays. That is, it takes up to 1500 ms to temporally integrate one visual image with another.

In this study we investigate an alternative hypothesis that the temporal window is needed for consolidating array 1 into VSTM. In our approach, we modified Brockmole et al.'s task into a probe task that required no integration between visual arrays. We then tested VSTM of array 1 as a function of the ISI between the two arrays. If a substantial time window is required for the probe task, then there is no need to explain the time window in terms of integration.

The consolidation hypothesis is supported by studies on visual memory of scenes and pictures (Potter, 1976), and from a phenomenon called the "attentional blink" (Raymond, Shapiro, & Arnell,

The current study is supported by a Helen Hay Whitney Research Fellowship. I would like to thank Molly Potter for insightful discussion and Chris Baker, Mike Mangini, Lakshminarayan Srinivasan, and Laura Wagner for helpful comments.

DOI: 10.1027//1618-3169.51.1.45

1992). In her pioneering studies using the rapid serial visual presentation (RSVP) technique, Potter presented a sequence of visual scenes to subjects and probed their recognition for the old images following the sequence. She found that the consolidation of a scene into visual STM took 200–300 ms. Representation of the scene was vulnerable to disruption by subsequent images. Attention enhances consolidation and reduces disruption. If subjects knew which scene to look for in advance, they could attend to the target scene and reduce interference from subsequent images. Potter proposed that scenes and pictures need to be transformed into a conceptual short-term memory, which is vulnerable to disruption before the image is fully consolidated.

The temporal lag for consolidation is not limited to scenes and pictures that are meaningful and conceptual. It is also shown for simple visual displays that are devoid of conceptual information. In studies of the attentional blink, when two targets are embedded among an RSVP sequence, detection and recognition of the second target is severely impaired if the temporal interval between the two targets is within 500 ms. Here again, the second target is processed but its consolidation cannot be complete when resource is devoted to the first target (Chun & Potter, 1995).

We propose that a temporal window of 300 to 500 ms is needed for consolidation of a visual image into a stable representation in VSTM. During this time window, processing of the image is vulnerable to disruption. Its vulnerability is reduced if memory resource can be fully devoted to this image, but is increased if the following image also needs to be processed. In tasks that do not necessitate the integration across two visual images, the temporal integration account predicts that memory for the first image should not be affected by the ISI between images. In contrast, the consolidation hypothesis predicts that an ISI of up to 500 ms is needed for memory of the first image to be fully consolidated. The following experiments provide evidence for the consolidation hypothesis.

Experiment 1: Temporal Window of Disruption by a Masking Array

In this experiment we completely removed the necessity to integrate the first array with the second. The first dot array contained black dots that were positioned in 12 cells of a 5 by 5 grid. A variable ISI was inserted before a second dot array was presented. The second array contained white dots that were positioned in all 25 cells of the grid. One of the cells was marked by a small red probe. The task was to determine whether the probed cell was initially occupied by the first dot array or not. Figure 1 is a schematic sample of a trial. If a long temporal window, longer than the duration of sensory memory (100 ms), is required for performance to reach asymptote in the current experiment, then mechanisms other than temporal integration should be responsible for such a long time course.

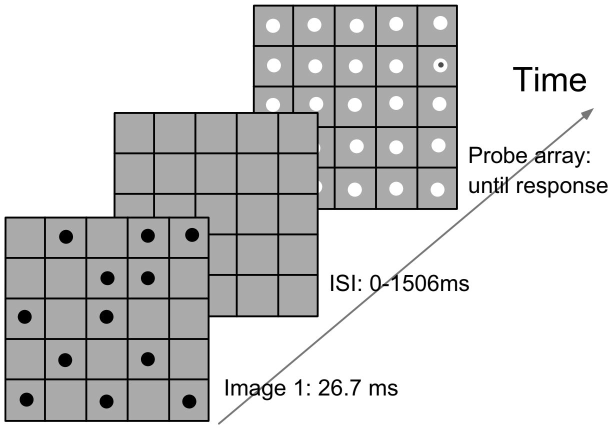

Figure 1. Schematic sequence of a trial tested in Experiment 1. The task is to say whether the cell probed by the red dot was initially occupied by the black dots.

Method

Participants

Seven participants (18–27 years old) from the Boston area volunteered for the experiment. They all had normal or corrected-to-normal visual acuity and normal color vision.

Equipment

Participants were individually tested in a room with dim lighting. They viewed a computer screen from an unrestricted distance of about 57 cm. The experiment was programmed using MacProbe 1.88A (Hunt, 1994).

Trial Sequence

Each trial started with a 5 by 5 grid (9.4 × 9.4°) for 400 ms, followed by a black dot array (each dot was 0.5 × 0.5°) that occupied a random set of 12 cells. The dot array was presented for 26.7 ms and erased, leaving the grid on the display. After a variable interval (ISI) another array was presented containing a white dot-array that occupied all 25 cells. One of the cells on the second array was probed by a red dot (0.2 × 0.2°). Subjects were asked to press the key labeled "same" if the probe was at one of the memory locations, and to press the key labeled "different" if it was previously blank. Upon their response, the probe array was erased and accuracy feedback in the form of a smiling or a sad face icon was presented for 200 ms. One second later the next trial started. Subjects took a break every 70 trials.

Design

One factor – ISI – was manipulated in this experiment. It had seven levels: 0, 40, 107, 200, 506, 1000, or 1506 ms. There were 30 trials at each ISI; half were "same" trials and the other half "different" trials. Each subject completed 14 practice trials and 210 experimental trials.

Data Analysis

In this study we calculated A' at each ISI as a measure of subject's sensitivity (Grier, 1971). We also measured percent correct and d'. The overall pattern of results were the same whether A', percent correct, or d' were used. A' was preferred as a measure of memory accuracy (Donaldson, 1993). Mean hit and false alarm rates are shown in the Appendix.

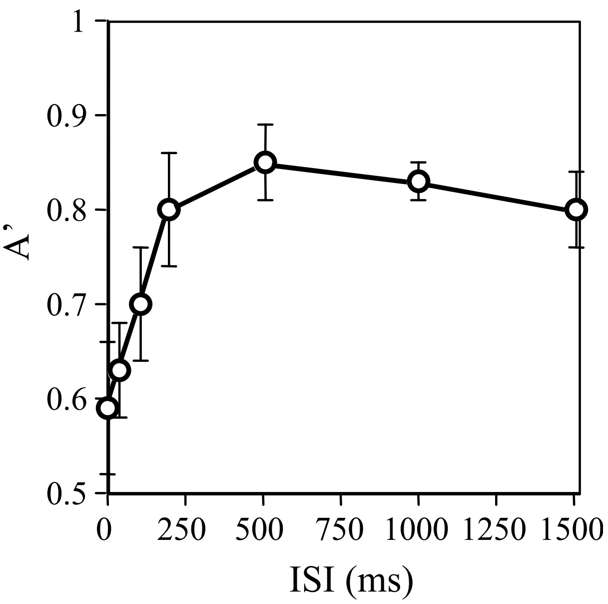

Figure 2. Results from Experiment 1: Performance (A') as a function of ISI between the two images. A' can range from 0.5 to 1.0, where 0.5 is chance performance and 1.0 is perfect performance. Standard error bars indicate inter-subject variance.

Results

Figure 2 shows the average A' measured across subjects. Chance performance corresponds to an A' of 0.5, and perfect performance to 1.0. A repeated measures ANOVA with ISI as the within-subject variable revealed a highly significant main effect of ISI, $F(6, 36) = 5.24$, $p < .001$. Performance increased gradually within the first 500 ms interval, reached maximum at 500 ms, and then leveled off. Pairwise comparison between adjacent ISIs failed to reveal statistical significance at any adjacent pairs, but performance was significantly higher when the ISI was 500 ms than when it was 0, 40, or 107 ms (all p values $< .05$). Note that performance was poor when the ISI was zero. This was because the second image contained white dots, which directly masked the locations of the black dots. Masking allowed no opportunity for sensory memory of image 1 to persist.

There was no evidence for a speed-accuracy tradeoff in subjects' performance. Mean RT for correct trials was 1207, 1048, 1004, 1022, 1082, 1035, and 1051 ms for the 7 ISIs, $F(6, 36) = 3.23$, $p < .012$. RT for the first interval (ISI = 0) was significantly slower than those for the other conditions, which did not differ from one another ($F < 1$). Because speed-accuracy tradeoff isn't an issue in any of the other experiments, we will no longer report RT in later experiments.

Experiment 2: Temporal Window and Memory for the Second Image

The first experiment showed that when integration was removed, a temporal window, on the order of about 200–500 ms, was entailed for VSTM of the first image to reach asymptote. This suggests that integration is not necessary to explain the intermediate temporal lag observed by previous researchers (Brockmole et al., 2002). Instead, we suggest that the consolidation of a visual image takes approximately 200–500 ms for an image to receive stable representation. Such consolidation is disrupted by the presentation of a masking array within the time window.

In this experiment, we further test whether a temporal window of consolidation is required when the second image does not mask the first. That is, when the second image does not overlap with the first, it does not produce sensory masking. The second array may still produce substitution masking (Di Lollo, Enns, & Rensink, 2000), although strictly speaking the two images do not share common onsets and thus substitution masking of the first image should be negligible at ISIs greater than 100 ms. Even under such conditions, the presentation of a second image can disrupt the consolidation of the first, especially if the second image also needs to be consolidated in memory.

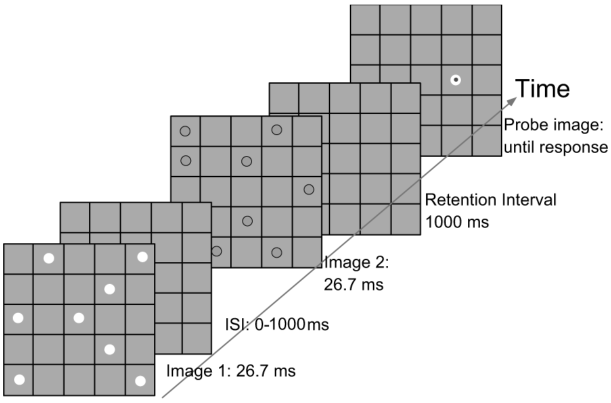

Figure 3. Schematic trial sequence tested in Experiment 2. In different blocks, subjects were asked to remember the first array only, the second array only, or both arrays. The probe was a white dot in the pure-image 1 block, a green dot in the pure-image 2 block, and either a white or a green dot in the mixed block.

Figure 3 shows a schematic sample of a trial sequence tested in Experiment 2. The first image was an array of white dots that occupied 8 of the 25 cells in the 5 by 5 grid. After a variable ISI, the second image, an array of green dots, occupied another set of 8 cells. One second later a probe image was presented, with one cell occupied. Subjects were tested in three blocks.

In the pure-image 1 block, the probe was always a white dot with a red center. The task was to report whether or not this cell was among the first array of white dots.

In the pure-image 2 block, the probe was always a green dot with a red center. The task was to report whether this cell was among the second array of green dots.

Finally, in the mixed block, the probe was a white dot with a red center on 50% of trials, and a green dot with a red center on the other 50% of trials, randomly intermixed. When the probe was white, subjects were to compare the probe with the first array of white dots; when the probe was green, subjects were to compare the probe with the second array of green dots.

We were interested in the memory performance for the first array. Because the second array did not overlap with the first, it should not create sensory masking of the first array. Will there be a temporal window for memory of the first array to reach asymptote? More importantly, will the duration of such a window be affected by whether the second image was ignored, as in the pure-image 1 block, or was attended, as in the mixed block?

Method

Participants

Seven participants from the same subject pool were tested in this experiment.

Procedure

Each subject was tested in three blocks of tasks: pure-image1, pure-image2, and mixed block. In all blocks, each trial started with a blank grid of 400ms followed by an array of white dots (n = 8) randomly presented on a 5 by 5 grid for 26.7 ms. Then the array was erased leaving the grid on the screen for a variable interval (ISI). A second array of green dots (n = 8) were then presented on the grid occupying another 8 cells for 26.7 ms. The array was then erased leaving the grid on the screen for 1 s. Finally, a probe dot was presented. It was always a white dot with a red center in the pure-image 1 block, a green dot with a red center in the pure-image 2 block, and either a white or a green dot with a red center in the mixed block. Subjects were instructed to remember array 1 and ignore array 2 in pure-image 1, to remember array 2 and ignore array 1 in pure-image 2, and to remember both arrays in the mixed block. Which array was probed was randomly selected from trial to trial in the mixed block.

When array 1 was probed, the probe was among one of the array 1 locations on 50% of the trials (“same” trials), and among one of the 9 locations not filled by either array on the other 50% of trials (“different” trials). Similarly, when array 2 was probed, the probe was among one of the array 2 locations on 50% of the trials, and among one of the 9 locations not filled by either array on the other 50% of trials.

Subjects were tested in 5 practice trials in a mixed-block design. Then they completed 160 trials in pure-image 1, 160 trials in pure-image 2, and 320 trials in the mixed block. Five levels of ISIs were used: 0, 107, 200, 506, and 1000 ms. The other aspects of the experiment were similar to Experiment 1.

Results

Figure 4 plots A’ as a function of ISI and block type separately for memory for array 1 and array 2. Un-

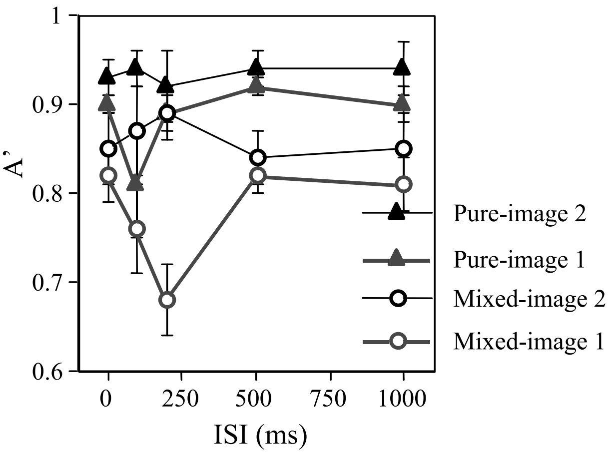

Figure 4. Results from Experiment 2: A' as a function of ISI between images, separated for the pure blocks (image 1 only or image 2 only) and mixed blocks.

like Experiment 1, the second array did not overlap spatially with the first array, so it did not produce sensory masking. Performance was therefore high when the ISI was 0 in the current experiment. The following important results were revealed. First, performance was higher for both array 1 and array 2 during the pure blocks than the mixed blocks. That is, if subjects had to memorize both arrays, performance for each array was poorer than if they only had to memorize one array. Second, performance for array 2 was not affected by the ISI, but performance for array 1 was. Third, the effect of ISI on array 1 depended on whether array 2 was ignored or memorized. When array 2 was ignored, array 1 could be disrupted within a time window of about 0–200 ms. But when array 2 was memorized, array 1 could be disrupted within a time window of 0–500 ms. That is, memorizing the second array prolonged the time window during which array 1 could be disrupted.

These results were confirmed by a repeated-measures ANOVA with two factors, block type (pure or mixed) and ISI (0–1000 ms), separately for A' of array 1 and array 2. For array 1, there was a significant main effect of block type, with higher accuracy for the pure block, $F(1, 6) = 17.10$, $p < .006$, a significant main effect of ISI, $F(4, 24) = 3.58$, $p < .02$, and a significant interaction between block type and ISI, $F(4, 24) = 2.88$, $p < .044$. Performance was poorest when the ISI was 107ms and largely recovered by 200 ms if array two was ignored. But if array 2 was remembered performance was poorest when ISI was 200 ms and did not recover until 500 ms. For array 2, there was a significant main effect of block type, with higher accuracy for the pure block, $F(1, 6) = 59.71$, $p < .001$. The main effect of ISI was not significant, nor was the interaction between ISI and block type, both F values < 1.40, $p > .25$.

Further evidence that memory consolidation characterizes performance came from an analysis on the mixed condition when accuracy was collapsed for array 1 and array 2. The combined A' was 0.83, 0.83, 0.80, 0.83, and 0.83 for ISIs of 0, 107, 200, 506, and 1000 ms, respectively. These values were not significantly different from one another, $F < 1$. This suggests a surprising degree of constancy for memory capacity of the two arrays.

Experiment 3: Temporal Window and Attention to the Second Image

Experiment 2 suggested that the temporal window needed for consolidation of an image is modulated by whether the subsequent image is ignored or remembered. The window was about 200 ms if the subsequent image was ignored, but increased to about 500 ms if the subsequent image was to be remembered. But what led to the increase in the temporal window of disruption? Is it because the second image was attended or because it needed to be remembered? To find out, in this experiment we replaced the second image with a visual search display. Instead of trying to remember the locations of an array of green dots, subjects needed to conduct a visual search from an array of "L"s. The target was a "T", either upright or inverted, and the distractors were "L"s. The task was to report which target – upright or inverted "T" – was presented. Compared with Experiment 2, the visual search task was attention-demanding but did not require memory for array 2. If attention to a second image is sufficient to lengthen the time window during which consolidation of image 1 can be disrupted, we should find the same pattern of results as Experiment 2. However, if memory of the second image is necessary for the lengthening of the temporal window, then we should not observe a change in the temporal property of image 1 consolidation.

Method

Seven participants were tested in this experiment in three blocks: pure-image 1, pure-image 2, and a mixed block. Image 2 now contained a visual search display of 7 "L"s ($0.9 \times 0.9°$) and one "T" that was either upright or inverted. When image 2 was probed, a question "upright or inverted T?" was presented at the center of the display. When image 1 was probed, a white dot with a red center was presented, just like Experiment 2. Subjects were asked to remember the white dots in pure-image 1, search for a “T” among “L”s in pure-image 2, and to do both in the mixed block. All other aspects of the experiment were the same as in Experiment 2.

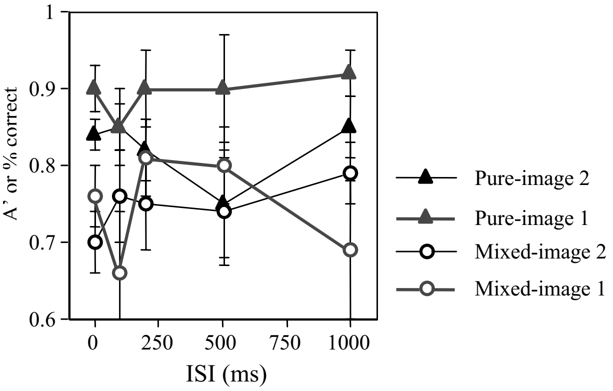

Figure 5. Results from Experiment 3: Performance as a function of ISI and block type. A’ was the dependent measure for image 1 performance, and percent correct was the dependent measure for image 2 performance, which involved a 2-alternative-forced-choice task. For both types of measures, 0.5 corresponds to chance and 1.0 corresponds to perfect performance.

Results

Figure 5 shows mean A’ for array 1, and mean percent correct for array 2 as a function of ISI and block type (pure or mixed). An ANOVA on block type and ISI showed that for performance on array 1, there was a significant main effect of block type, $F(1, 6) = 20.39$, $p < .004$, showing better performance for the pure blocks. There was also a significant main effect of ISI, $F(4, 24) = 3.77$, $p < .016$, showing a dip in performance when ISI was 107 ms. However the interaction between block type and ISI was not significant, $F(4, 24) = 1.15$, $p > .35$. Thus, even though the need to attend and search from image 2 reduced memory for array 1, the temporal window of the consolidation of array 1 was unaffected by attention.

For performance on array 2, there was a significant main effect of block type, $F(1, 6) = 9.19$, $p < .023$, but no effect of ISI or interaction, F values < 1.10, $p > .35$.

General Discussion

In this study we asked whether the long temporal window during which a visual image can be disrupted by subsequent images was necessarily due to temporal integration between the two images. Using a probe task that removed the necessity to integrate visual images, we found that a long temporal window of 200–500 ms was still required for memory of an image to reach asymptote. This suggests that a process of slow temporal integration is not needed to account for the time window. Instead, a simpler account, a slow consolidation for array 1, is sufficient to explain such temporal properties.

Specifically, following Potter’s conception that a relatively long consolidation period is needed for a visual image to receive stable representation (Potter, 1976), we suggest that the representation of a visual image is vulnerable to disruption within a temporal window of 200–500 ms. It can be disrupted by a masking display that terminates sensory memory of the first image (Experiment 1), or by a non-overlapping display that needs to be memorized (Experiment 2). The temporal window for disruption is reduced to about 200 ms if the subsequent image is non-overlapping and ignored (Experiment 2), or if it is attended in a search task but not memorized (Experiment 3). We believe that the need to memorize both images significantly lengthened the consolidation period for the first image. The combined memory capacity for the two arrays observed in Experiment 2 varied little across ISIs, suggesting that there is a constant capacity limit on the amount of information being consolidated. The visual search task of Experiment 3 did not lengthen the temporal window, perhaps because only one bit of information – target identity – competes for memory with array 1. Future studies should vary memory set size for array 2 to see how it affects the temporal window of the consolidation of array 1.

Finally, we note that this study does not necessarily refute the possibility that temporal integration across images takes a substantial amount of time. In fact, the consolidation hypothesis and the temporal integration hypothesis (Brockmole et al., 2002) are not exclusive. Both consolidation and integration may take a long time. The consolidation hypothesis is preferred here because it is simpler and generalizes to tasks that require no integration. Future studies are needed to isolate temporal integration and to test whether it also requires a long temporal duration, and whether it lengthens the temporal window during which the processing of an image can be disrupted (see Jiang & Kumar, in press).

References

Brockmole, J. R., Wang, R. F., & Irwin, D. E. (2002). Temporal integration between visual images and visual percepts. *Journal of Experimental Psychology: Human Perception & Performance, 28,* 315–334.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual pre-

sentation. *Journal of Experimental Psychology: Human Perception & Performance, 21,* 109–127.
Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General, 109,* 75–97.
Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129,* 481–507.
Donaldson, W. (1993). Accuracy of d' and A' as estimates of sensitivity. *Bulletin of the Psychonomic Society, 31,* 271–274.
Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75,* 424–429.
Hunt, S. M. J. (1994). MacProbe: A Macintosh-based experimenter's workstation for the cognitive sciences. *Behavior Research Methods, Instruments, & Computers, 26,* 345–351.
Jiang, Y., & Kumar, A. (in press) Visual short-term memory for two sequential arrays: One integrated representation or two separate representations? *Psychosomatic Bulletin and Review.*
Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16,* 283–290.
Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2, 509–522.*
Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18,* 849–860.
Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74,* 1–29.

Received December 11, 2002
Revision received February 13, 2003
Accepted February 13, 2003

Yuhong Jiang

33 Kirkland Street, Room 820
Cambridge, MA 02138
USA
Tel.: +1 617 496 4486
Fax: +1 617 495 3728
E-mail: yuhong@wjh.harvard.edu

Appendix A:

Mean hits and false alarms obtained from each experiment.

Experiment	Conditions: Hit/False alarm						
Exp1: ISI	0	40	107	200	506	1000	1506
Hits/FA	.46/.35	.48/.31	.45/.22	.69/.22	.79/.21	.79/.32	.81/.35

Experiment					
Exp2: ISI	0	107	200	506	1000
Mixed-image 1	.70/.28	.67/.33	.67/.43	.73/.26	.76/.30
Mixed-image 2	.70/.09	.73/.10	.71/.19	.75/.20	.77/.11
Pure-image 1	.77/.11	.69/.20	.75/.13	.78/.06	.83/.16
Pure-image 2	.86/.12	.87/.08	.81/.07	.86/.06	.87/.08
Exp3: ISI	0	107	200	506	1000
Mixed-image 1	.59/.20	.68/.43	.70/.25	.80/.33	.68/.36
Pure-image 1	.79/.13	.78/.24	.86/.15	.83/.11	.85/.11

