Detecting and remembering pictures with and without visual noise

Ming Meng
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Mary C. Potter
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Objects in a scene are often partially occluded without causing the viewer any problem: the occluded parts are apparently represented via amodal completion. To evaluate human ability to perceive and remember partially occluded pictures, we showed sequences of pictures using rapid serial visual presentation (RSVP) for durations of 53 ms, 107 ms, 213 ms, or 426 ms/picture. Participants either attempted to detect a named target (e.g., "businessmen at table") or were given a yes–no recognition memory test of one item. In Experiment 1, with as much as 30% of the picture area covered, detection and recognition were both well above chance. More interestingly, occlusion significantly affected recognition memory but not target detection. In Experiment 2, when pictures were inverted, occlusion impaired detection as severely as recognition. For target detection, the interaction between occlusion and inversion was significant. By contrast, taking away color information did not significantly reduce detection's tolerance of occlusion (Experiment 3). Finally, Experiment 4 showed that with 40% of the picture area occluded, detection performance was impaired. These results support the hypothesis that contextual gist information facilitates visual processes that tolerate occluding noise. Although inversion and color were tested in particular, the presented paradigm can also be used to investigate the role of other factors in gist representation.

Keywords: target detection, recognition memory, RSVP, occlusion, scene perception, amodal completion

Citation: Meng, M., & Potter, M. C. (2008). Detecting and remembering pictures with and without visual noise. *Journal of Vision, 8*(9):7, 1–10, http://journalofvision.org/8/9/7/, doi:10.1167/8.9.7.

Introduction

One remarkable characteristic of human visual perception is its tolerance to visual noise. For example, partial occlusion is ubiquitous in normal visual experience, yet observers usually can reconstruct and recognize partially occluded objects through amodal completion (e.g., Kanizsa & Gerbino, 1982; Sekuler & Palmer, 1992). In a picture such as that in Figure 1A, the woman's body is partially occluded by the rock, but her head and feet are not perceived as separated; instead, the perception of a continuous torso behind the rock is not impeded by the absence of direct sensory input. In Figure 1B, the need for completion is more evident. Despite the many disks covering the picture, however, understanding of the occluded scene does not seem particularly difficult. This seemingly effortless process of perceptual completion reflects the constructive nature of visual representation. Characterizing the visual system's ability to tolerate occluding noise is one way of investigating the constructive mechanisms of visual processing.

The purpose of the current study is to evaluate human ability to perceive and remember partially occluded pictures of natural scenes. Previous studies of visual completion mainly focused on pictorial occlusion of simple objects like circles, triangles, and rectangles (e.g., Bruno, Bertamini, & Domini, 1997; Guttman, Sekuler, & Kellman, 2003; Joseph & Nakayama, 1999; Rauschenberger, Liu, Slotnick, & Yantis, 2006; Rauschenberger & Yantis, 2001; Sekuler & Palmer, 1992). However, natural scenes have much more complex statistical properties than such simple objects and they are more closely linked to our everyday experiences. Can observers identify and recognize a partially occluded scene in a brief presentation? How may the availability of information about conceptual gist in a given task, and the demands of that task, interact with the effects of occlusion? Answers to these questions may help us to understand how the visual system can represent natural scenes in a constructive manner, allowing it to tolerant partial occlusion.

It is known that detection of a target picture presented briefly is more accurate than recognition memory for the picture (Potter, 1976). Human observers have little difficulty in detecting a named target when viewing scenes presented in a rapid sequence for about 113 ms each. However, about 300 ms per picture of further processing is normally required to form a memory representation of a scene that can resist conceptual masking from a following scene (Potter, 1976; Potter & Levy, 1969). In addition to

doi: 10.1167/8.9.7 Received October 16, 2007; published July 16, 2008 ISSN 1534-7362 © ARVO

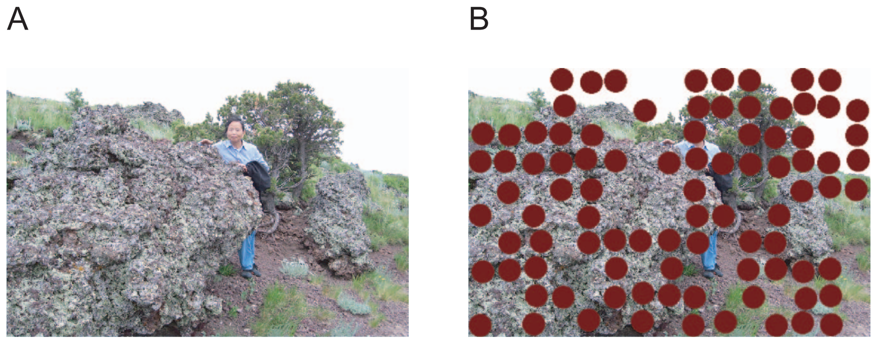

Figure 1. Example of occlusion. (A) Without occluding dots. (B) With occluding dots.

the fact that recognition memory requires extra processing, one important difference between target detection and recognition is that gist information about the target is provided in the detection task. This gist description may facilitate understanding of the scene (Potter, 1976; for reviews, see Intraub, 1999; Oliva, 2005; Potter, 1999). Here we predict that gist information may also help observers to tolerate the effects of occlusion, when partially occluded natural scenes are presented in a rapid sequence.

In Experiment 1, we used rapid serial visual presentation (RSVP) to investigate whether occlusion impairs an observer's ability to detect and remember natural scene pictures presented for between 53 and 426 ms per picture. Experiment 2 tested whether inverting those pictures increases the difficulty of detection and recognition of the occluded pictures. Previous research has shown that inverted pictures are more difficult to process than upright pictures (Evans & Treisman, 2005; but see also Rousselet, Mace, & Fabre-Thorpe, 2003). It is unknown whether occlusion interacts with inversion. For target detection, Experiment 3 further tested whether occlusion interacts with whether the picture is presented in color versus in grayscale. Experiment 4 tested the effect of increasing the percent occlusion on target detection. The general method used in all of the experiments is described in detail in Experiment 1.

Experiment 1: Detecting and remembering pictures with and without occlusion

Rapid serial visual presentation (RSVP) has been widely used to study the visual system's capacity for processing a series of words, objects, or pictures (e.g., Chun & Potter, 1995; Evans & Treisman, 2005; McKeeff, Remus, & Tong, 2007; Potter, 1975; Potter & Levy, 1969). Here, by comparing RSVP trials with or without occlusion, we tested whether, with limited processing time, occlusion impairs an observer's ability to detect and remember pictures of natural scenes. In this experiment, disks occluded 30% of the area of each picture. Participants were tested with presentation durations as brief as 53 ms per picture, less than the duration previously thought to be required for object amodal completion (Murray, Sekuler, & Bennett, 2001; Ringach & Shapley, 1996; Sekuler & Palmer, 1992). Previous studies of unoccluded scene perception have used presentation durations varied across a wide range (Potter, 1976; Potter & Levy, 1969). To make our results potentially comparable with these studies, longer presentation durations of 107, 213, and 426 ms/picture were also tested.

Methods

Participants

Thirty-two volunteers from the Massachusetts Institute of Technology community received payment for participation. All reported normal or corrected-to-normal visual acuity and were naïve to the purpose of the experiment. Sixteen participants were assigned to each of two groups: a detection group and a recognition memory group. No participants took part in more than one of the experiments in the present study, and none had been in earlier studies with pictures in this laboratory.

Apparatus, materials, and procedure

The pictures used in the experiment were 1296 color photographs with widely varying content, including indoor and outdoor scenes (size = 300 × 200 pixels, visual angle about 12.9 × 8.6 degrees when viewed at the normal viewing distance of 45 cm). The pictures were given short descriptive titles or names in a previous study, and these names were used in the detection condition. Sequences of pictures were shown using rapid serial visual presentation (RSVP) for durations of 53, 107, 213, or 426 ms/picture, within subjects. Each sequence contained 8 pictures. No picture was repeated in the experiment. In half of the trials, all 8 pictures in the sequence were partially occluded, whereas in the other half they were presented in the original form. Partially occluded pictures were covered by 90 red disks (size = 16 pixels in diameter) that occluded 30% of the picture area. The disks, which we will term dots, appeared randomly in 90 of the 150 virtual squares, 20 × 20 pixels, that made up the 300 × 200 pixel picture. Within the square, the dot was randomly placed in a 19 × 19 pixel inner region, so that no dot directly touched another dot and the dots were not perfectly aligned either vertically or horizontally. The position of the dots varied from picture to picture in a sequence. The same random order of trials was used for every participant. The SOA and whether the sequence in the trial was presented normally or partially occluded were counterbalanced within subjects and within trials between subjects.

All stimulus presentations were controlled using MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) driven by an Apple Macintosh G3 computer. Stimuli were presented in the center of a 17-in. CRT monitor (refresh rate = 75 Hz, resolution = 1024 × 768 pixels). Pictures were shown against a black background, which was present throughout. The room was dimly illuminated. For each participant, there were 144 trials and 8 practice trials. Participants were allowed to rest between trials, at any point.

For the detection task, participants were instructed to look for a named target picture in the RSVP sequence in each trial. After a button-press to initiate the trial, a descriptive title of the target (e.g., "businessmen at table") was presented in the center of the screen for 2 seconds. Then, a fixation cross was presented for 300 ms, followed by a blank of 200 ms and the picture sequence. Following the RSVP sequence, a white rectangle was presented for 500 ms to signal the end of the trial. The participant was instructed to press a key labeled "YES" or "NO" on the keyboard to indicate whether he/she had seen a picture that fit the target description. The correct answer was "YES" on 2/3 of the trials. In those trials, the serial position of the target picture was random but never the first or last in the 8-item sequence.

For the recognition group, the same RSVP sequences of pictures were presented to participants, but without any target description given in advance. The picture that was the target in the detection condition was tested for recognition. For the trials with no target, another picture not in the sequence (matching the name in the no-target detection condition) was presented. After the white rectangle following the RSVP sequence, the test picture (without dots) was presented for 0.5 s. The relatively short duration of the test picture encouraged the participant to make a rapid decision. The participant was instructed to press a key labeled "YES" or "NO" on the keyboard to indicate whether he/she remembered seeing that test picture in the sequence. The correct answer was "YES" on 2/3 of the trials. In those trials, the serial position of the tested picture (the target picture in the detection condition) was random but never the first or last in the sequence. When the correct answer was "NO," the test picture came from the same source as all other pictures but had never been presented to the participant.

Data analyses

In this and the following experiments, we used A′ as a criterion-free measure to evaluate the participant's ability to detect/recognize the target pictures. A′ is a nonparametric signal detection measure, which also has greater sensitivity than *d′* when data contain extreme values (Donaldson, 1992; Stanislaw & Todorov, 1999). To compare the mean A′ in each condition, analyses of variance (ANOVAs) were performed. All results significant at the .05 level or better are reported.

Results and discussion

Figure 2 shows the group mean A′ for each condition in the detection task (blue lines) and in the recognition task (red lines). Solid lines represent the intact picture condition; dashed lines represent the occluded condition. Error bars represent ±1 *SEM*. In both tasks, performance was significantly above chance even at the shortest duration (53 ms), regardless of occlusion. As expected, in the ANOVA the main effect of SOA was highly significant for both the detection task, $F(3, 120) = 31.68$, $p < 0.001$, and the recognition task, $F(3, 120) = 14.78$, $p < 0.001$. Most interestingly, the effect of occlusion was not significant for the detection task, $F(1, 120) = 3.07$, $p = 0.082$, but highly significant for the recognition task, $F(1, 120) = 15.21$, $p < 0.001$. The interaction between

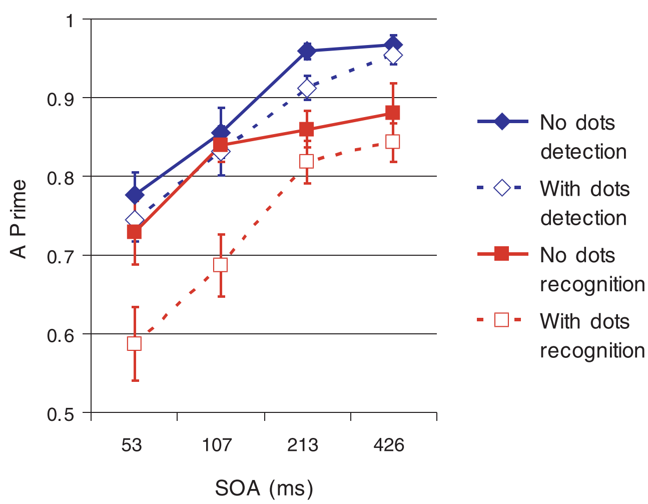

Figure 2. Performance as measured by A′ in detecting a specified picture (blue lines) and recognizing a presented picture (red lines) as a function of SOA. Solid lines represent the intact picture condition; dashed lines represent the occluded condition. Error bars represent ±1 *SEM*.

occlusion and SOA failed to reach significance either for detection, $F(3, 120) = 0.18$, $p = 0.91$, or recognition, $F(3, 120) = 1.71$, $p = 0.17$.

A mixed model ANOVA comparing the two tasks revealed that the participants were worse overall at the recognition task than at the detection task, $F(1,240) = 42.98$, $p < 0.001$. The main effects of SOA, $F(3,240) = 40.14$, $p < 0.001$, and occlusion, $F(1,240) = 17.73$, $p < 0.001$, were also significant. More interestingly, the interaction between task condition and the effect of occlusion was significant, $F(1,240) = 5.01$, $p < 0.05$; none of the other interactions were significant. These results support the notion that occlusion greatly impairs recognition memory but not target detection. Note that in the detection task participants were given a title that contained gist information about the target picture. Contextual gist information is known to be important for top-down facilitation of picture processing (Bar, 2004; Chun, 2000; Henderson & Hollingworth, 1999; Intraub, 1997). In the present detection task, gist information may also have helped participants to tolerate the visual noise that was induced by occlusion, whereas no such information was available in the recognition task, accounting for the larger effect of occlusion in the latter case.

The results showing that participants are better at detection than recognition are consistent with previous research (Potter, 1976). For the unoccluded condition (solid lines in Figure 2), the main effect of SOA was significant, $F(3,120) = 15.70$, $p < 0.001$; the performance difference between detection and recognition was also significant, $F(1,120) = 10.04$, $p < 0.01$. The difference appears to have been minimal at the two shorter SOAs, although the interaction between SOA and task failed to reach significance, $F < 1.0$. The relatively good performance in the recognition task is consistent with evidence that recognition memory is good for a picture tested immediately after presentation, although it declines with delay or with tests of more than one picture (Potter, Staub, Rado, & O'Connor, 2002). It has been shown that pictures in RSVP are often understood momentarily but require further processing before being consolidated into conceptual short term memory (for a review, see Potter, 1999).

Experiment 2: Detecting and remembering inverted pictures with and without occlusion

Experiment 1 supports the hypothesis that contextual gist information can help the visual system tolerate noise. It is also possible that our visual system can sometimes rely on local features such as color and contour collinearity to detect a target. For example, to detect a beach scene, participants might achieve better than chance performance by simply looking for white patches of sand and blue patches of ocean. Even in the occluded condition, residual white and blue color information is still available; therefore, participants could still use this local feature strategy to detect the beach scene. On the other hand, spatial arrangement of the colors (e.g., blue ocean over white sand) rather than the color information alone may be crucial for detection. Inversion interrupts spatial relationship but not local color features. Experiment 2 used inverted pictures to test whether and to what extent our visual system might rely on local features to tolerate occlusion.

Inverting a picture preserves many orientation-invariant local features such as contrast, spatial frequency, color, and contour collinearity. On the other hand, inversion greatly affects configural spatial layout, which is an important component of scene gist (Oliva, 2005; Oliva & Torralba, 2001; Schyns & Oliva, 1994; Torralba & Oliva, 2003; Wolfe, 1998). Previous RSVP research has shown that detection of inverted pictures is more difficult than detection of upright pictures (Evans & Treisman, 2005; but see also Rousselet et al., 2003). Recognition memory for inverted pictures is also generally less good than memory for upright pictures (e.g., Diamond & Carey, 1986; Yin, 1969). We asked whether the inversion effect interacts with interference from occluding dots differentially for detection and recognition.

Our hypothesis is that, for upright pictures in Experiment 1, having a descriptive title allowed viewers to detect picture gist, thereby minimizing the effect of occlusion. It has been proposed that a "gist" may include an inventory of objects, a scene category label, the spatial frequency distribution, and the global layout of objects (Fei-Fei, Iyer, Koch, & Perona, 2007; Oliva, 2005; Oliva & Torralba, 2001; Wolfe, 1998). Although inversion does not impair all components of the gist, it interferes with holistic representation of the gist, particularly by changing the global spatial layout of the picture. If our hypothesis is correct, then inversion will significantly reduce tolerance of occlusion in the detection task. However, if having a title allows viewers to use orientation-invariant local features to detect a target despite inversion and occlusion, then the negative effect of occlusion plus inversion should be less marked in the detection task than in the recognition task.

Methods

Participants

Thirty-two people from the same source as Experiment 1 participated in Experiment 2.

Apparatus and procedure

The stimuli and experimental procedure were the same as Experiment 1, except that the RSVP pictures were inverted. In the recognition task, the test picture was presented upright. Thus, the testing stimuli were exactly the same as in Experiment 1; only the RSVP sequence was inverted.

Results and discussion

Figure 3 shows participants' A′ performance for inverted pictures in the detection task (blue lines) and in the recognition task (red lines). Solid lines represent the intact picture condition; dashed lines represent the occluded condition. Error bars represent ±1 *SEM*. As expected, the effect of SOA was highly significant for both detection, $F(3,120) = 24.81$, $p < 0.001$, and recognition, $F(3,120) = 8.33$, $p < 0.001$. More interestingly, the effect of occlusion was also highly significant for both the detection task, $F(1,120) = 15.89$, $p < 0.001$, and the recognition task, $F(1,120) = 11.1$, $p < 0.01$. A mixed model ANOVA comparing the two tasks revealed that the participants were worse at the recognition task than at the detection task, $F(1,240) = 23.40$, $p < 0.001$. Overall, the main effects of SOA, $F(3,240) = 27.26$, $p < 0.001$, and occlusion, $F(1,240) = 25.78$, $p < 0.001$, were also significant. Unlike for upright pictures in Experiment 1, the interaction between task condition and the effect of occlusion was far from significant, $F < 1.0$. None of the other interactions were significant either.

We also conducted mixed model ANOVAs comparing performance in Experiment 2 (inverted) with that in Experiment 1 (upright), separately for detection and recognition. For the detection task, there were highly significant effects of occlusion, $F(1,240) = 17.91$, $p < 0.001$, and inversion, $F(1,240) = 25.03$, $p < 0.001$, as well as SOA, $F(3,240) = 53.78$, $p < 0.001$. Interestingly, the interaction between the inversion effect and the occlusion effect was significant, $F(1,240) = 5.00$, $p < 0.05$, with a larger effect of occlusion in the inverted condition. For the recognition task, there were again highly significant effects of occlusion, $F(1,240) = 25.51$, $p < 0.001$, and inversion, $F(1,240) = 10.84$, $p < 0.01$, as well as SOA, $F(3,240) = 20.76$, $p < 0.001$. However, there was no interaction between inversion and occlusion, $F < 1.0$. A further analysis comparing detection and recognition tasks in the two experiments showed that the main effects of SOA, $F(3,480) = 63.96$, $p < 0.001$, inversion, $F(1, 480) = 31.58$, $p < 0.001$, occlusion, $F(1,480) = 43.25$, $p < 0.001$, and different tasks, $F(1,480) = 61.73$, $p < 0.001$, were all significant; the three-way interaction among inversion, occlusion, and task was not significant, $F(1,480) = 1.71$, $p = 0.19$.

Comparing the results in Experiment 2 with Experiment 1, the inverted pictures are more difficult to process. Most notably, inversion significantly increased the occlusion effect on detection, making it as large as the occlusion effect on recognition. This finding cannot be easily explained if detection of occluded pictures typically relies on local features that are little affected by inversion. Instead, our results favor the hypothesis that detection is facilitated by configurational information that is interfered with by inversion but is relatively insensitive to occlusion alone. By inverting the pictures, here we show that when configural information is weakened, detection and recognition have similar difficulties with visual noise in the form of occlusion.

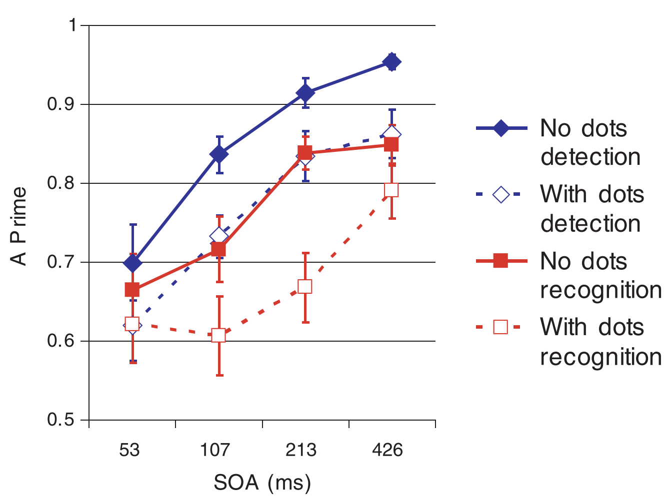

Figure 3. For inverted pictures (Experiment 2), performance in detecting a specified picture (blue lines) and recognizing a presented picture (red lines) as a function of SOA. Solid lines represent the intact picture condition; dashed lines represent the occluded condition. Error bars represent ±1 *SEM*.

Experiment 3: Detecting gray scale pictures with and without occlusion

Previous research has reached mixed conclusions about the importance of color in object perception and scene processing (e.g., Biederman & Ju, 1988; Davidoff & Ostergaard, 1988; Delorme, Richard, & Fabre-Thorpe, 2000; Fei-Fei, VanRullen, Koch, & Perona, 2005; Goffaux et al., 2005; Oliva & Schyns, 2000; Ostergaard & Davidoff, 1985; Price & Humphreys, 1989; Wurm, Legge, Isenberg, & Luebker, 1993). Does color information help to mitigate the effects of occlusion in the present study? In Experiment 3, we tested whether gray-scale versions of the pictures (with or without red dots like those used in the previous experiments) increased sensitivity to occlusion in the detection task. Unlike the inverted pictures in Experiment 2, gray-scale pictures generally leave contextual gist information intact. If the gist information plays a critical role in enabling viewers to tolerate occluding noise, removing color should have little effect on detection. However, if the visual system uses color information to mitigate the effects of occluding noise, target detection with occluded gray-scale pictures should be more difficult than with occluded color pictures.

Methods

Participants

Participants consisted of 16 people from the same source as Experiments 1 and 2; none had participated in those experiments.

Apparatus and procedure

The apparatus and procedure were the same as in the detection condition of Experiment 1, except that the stimuli were gray-scale pictures instead of colored pictures.

Results and discussion

Figure 4 shows participants' A′ performance on the detection task, together with the results of Experiment 1. Black lines represent performance with gray scale pictures (solid = intact picture condition; dashed = occluded condition). Light blue lines are data from Experiment 1 for colored pictures. Error bars represent ±1 *SEM*. For gray-scale pictures, the main effect of SOA, $F(3,120) = 44.55$, $p < 0.001$, and the main effect of occlusion, $F(1,120) = 6.97$, $p < 0.01$, were both significant. The interaction between occlusion and SOA was not significant, $F < 1.0$. More interestingly, detection of gray-scale pictures largely replicated detection of colored pictures in Experiment 1. In a mixed model ANOVA comparing detection in Experiments 1 and 3, neither the overall effect of color nor the interaction between color and occlusion was significant, $F < 1.0$. The main effect of SOA, $F(3,240) = 74.33$, $p < 0.001$, and the main effect of occlusion, $F(1,240) = 9.36$, $p < 0.01$, were both significant. None of the other interactions were significant, $F < 1.0$.

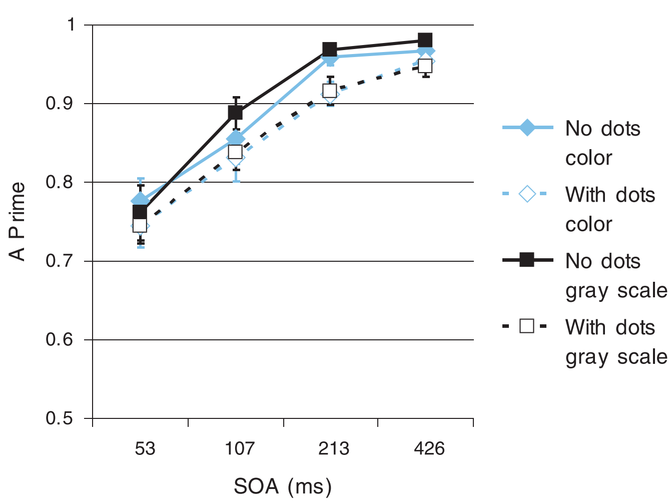

Figure 4. Performance in detecting a specified picture in the grayscale condition (Experiment 3, black lines) compared with the original color condition (Experiment 1, light blue lines) as a function of SOA. Solid lines represent the intact picture condition; dashed lines represent the occluded condition. Error bars represent ±1 *SEM*.

These results suggest that with conceptual gist information available, color information does not seem critical for identifying RSVP pictures. It may appear that occlusion was more of a problem with gray-scale than with colored pictures (see Figure 4), inasmuch as occlusion had a significant effect on detection in Experiment 3 but not Experiment 1. However, the relevant interaction between Experiments 1 and 3 was not significant.

Experiment 4: Detection when occluding dots cover 40% of the picture

Experiments 1 and 3 showed that the visual system has considerable tolerance for occluding noise in the detection task. Quite counterintuitively, detection performance is far above chance and, indeed, little impaired for pictures in which 30% of the area is covered by dots and presentation is as brief as 53 ms/item. However, there must be a limit to how much occlusion the visual system can tolerate. The purpose of Experiment 4 was to determine at what level occlusion significantly affects detection, as the percentage of occlusion increases. The number of occluding dots was increased from 90 to 120, covering 40% of the picture area.

Methods

Participants

Participants consisted of 16 people from the same source as previous experiments; none had participated in the earlier experiments.

Apparatus and procedure

The apparatus and procedure were the same as in the detection condition of Experiment 1, except that in the occluded condition pictures were covered by 120 red dots instead of 90; thus, there was a dot within a random 120 virtual squares out of the 150, with each virtual square consisting of 20 × 20 pixels. The dots covered 40% of the pictures instead of 30%.

Results and discussion

Figure 5 shows participant's performance on the detection task, together with the results from Experiment 1. Black lines represent data of Experiment 4, with occluding dots cover 40% of the pictures in the occluded condition (solid = intact picture condition; dashed = occluded condition). Light blue lines are the results of Experiment 1 shown for comparison. Error bars represent ±1 *SEM*. In the analysis of Experiment 4, the main effect of SOA, $F(3,120) = 25.98$, $p < 0.001$, and the main effect of occlusion, $F(1,120) = 32.97$, $p < 0.001$, were both highly significant. The interaction between SOA and occlusion approached significance, $F(3,120) = 2.24$, $p = 0.087$, suggesting that with longer SOAs performance may begin to recover even from 40% occlusion.

A mixed model ANOVA comparing the results of Experiment 4 and the detection condition in Experiment 1 revealed that, overall the main effects of SOA, $F(3,240) = 56.40$, $p < 0.001$, and occlusion, $F(1,240) = 39.44$, $p < 0.001$, were significant. The main effect of the difference between Experiment 1 and 4 was not significant, $F < 1.0$. However, the interaction between Experiment and occlusion was significant, $F(1,240) = 10.59$, $p < 0.01$; as expected, occlusion had a larger effect in the 40% condition. Pooling the data from Experiment 4 and the detection condition in Experiment 1, we had three levels of occlusion: no occlusion (Experiments 1 and 4), 30% occlusion (Experiment 1), and 40% occlusion (Experiment 4). A general linear model analysis comparing these three levels revealed that overall the main effect of SOA, $F(3,106) = 55.99$, $p < 0.001$, and the main effect of occlusion, $F(2,30) = 17.44$, $p < 0.001$, were both highly significant. The interaction between SOA and occlusion was not significant, $F(6, 90) = 1.63$, $p = 0.15$. These results match an intuitive prediction that more occluding dots would increase visual processing difficulty.

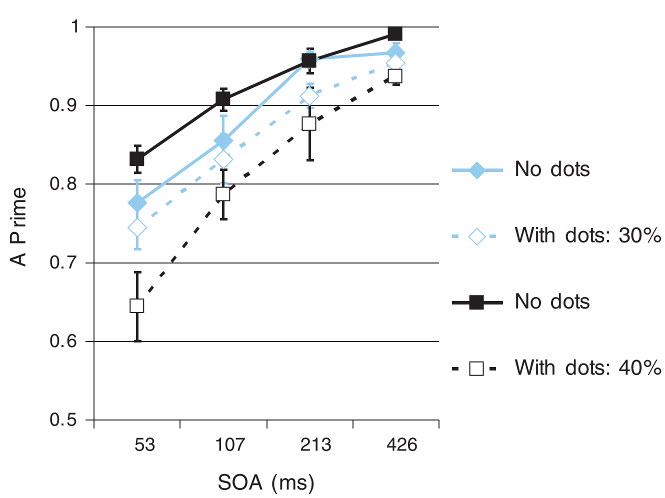

Figure 5. Performance in detecting a specified picture as a function of SOA. Black lines represent data of Experiment 4 with occluding dots covering 40% of the pictures in the occluded condition (solid = intact picture condition; dashed = occluded condition). Light blue lines are the results of Experiment 1 with occluding dots covering 30% of the pictures in the occluded condition. Error bars represent ±1 *SEM*.

General discussion

Overall, the visual system tolerates occluding noise to a great extent even when processing time is very brief. At the four tested presentation rates (53, 107, 213, and 426 ms per item), with as much as 30% of the picture area covered, target detection performance and recognition memory performance were both well above chance (Experiment 1). Interestingly, occlusion affected recognition memory significantly more than detection for upright pictures. However, when pictures were inverted, occlusion impaired detection as severely as recognition (Experiment 2). In contrast, taking away color information had only a small effect on detection's tolerance of occlusion, suggesting that the visual system does not rely on color information to detect a target in the presence of occluding noise (Experiment 3). Finally, Experiment 4 showed that with 40% of the picture area occluded, occlusion impaired detection significantly.

What mechanisms account for the visual system's remarkable ability to tolerate occluding noise? We think contextual gist information may play an important role in the processes that allow detection despite visual noise. With gist information provided in the target detection task, participants were significantly better at tolerating occlusion than in the recognition memory task when no advance gist information was available. However, for target detection, the interaction between occlusion and inversion was significant. Inversion made gist understanding difficult; therefore, with upside down pictures, participants in the detection task had as much difficulty in tolerating occlusion as did those in the recognition task. In contrast, when color information was taken away in Experiment 3, contextual gist was preserved, and participants' tolerance for occlusion in the detection task was not changed significantly. Clearly, inversion does not necessarily destroy all the gist representation. For example, nonface objects are relatively insensitive to inversion (Yin, 1969). Nonetheless, the inventory of objects may also be an important part of a gist (Wolfe, 1998). We hypothesized that only effects that impact gist representation would reduce noise tolerance. Our experiments particularly tested inversion and color; the same paradigm could also be used to investigate the role of other effects on gist representation and their relation to tolerance of occlusion.

The important role of gist in picture processing concurs with results of previous psychophysical and neuroimaging studies. Immediate memory for pictures incorporates information about gist, as shown by the ability to recognize a gist title and the tendency to falsely recognize a picture with the same gist as one just seen (Potter, Staub, O'Connor, & Potter, 2004). Contextual information has long been known to be highly influential in object perception (Biederman, Mezzanotte, & Rabinowitz, 1982; Henderson & Hollingworth, 1999; Palmer, 1975). Moreover, boundary extension (Intraub, 1997) and change blindness experiments (Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997) have shown that gist encoding can greatly bias visual representation of pictured scenes. A recent neuroimaging study has found that a blurred oval-shaped stimulus, if placed where a face was expected to be, can activate a face-selective area in the fusiform gyrus, presumably due to top-down contextual modulation (Cox, Meyers, & Sinha, 2004). A cortical network including parahippocampal sulcus, prefrontal cortex, and retrosplenial cortex is further proposed to mediate the effects of contextual associations on visual perception and cognition (Bar, 2004).

Amodal completion may be also involved in tolerating occlusion during rapid picture processing, particularly when a large percentage of the picture area is covered. Previous studies have used simple illusory contour stimuli to investigate object amodal completion and have suggested that the completion procedure may take place within 100–200 ms (Gold & Shubel, 2006; Ringach & Shapley, 1996). Consistent with this latency, Experiment 4 showed a trend for a drop in detection performance at the highest presentation rates (53–107 ms/item; see Figure 5) when dots covered 40% of the picture. However, scenes have much more complex statistical properties than simple objects and therefore picture processing may be different from object processing (Braun, 2003; Kayser, Körding, & König, 2004; Li, VanRullen, Koch, & Perona, 2002). Scene processing also activates different areas in the brain compared to object processing (Epstein & Kanwisher, 1998). Further investigation is needed to find out the exact relationship between amodal completion and how the visual system tolerates occlusion in scene processing.

Being able to tolerate visual noise can offer a major survival advantage for an intelligent organism. In the wild, predators may be hidden in bushes, partially occluded by leaves and branches. Rapidly detecting such danger in less than perfect viewing conditions can make a life or death difference. Similarly, predators benefit from withstanding visual noise in pursuit of prey. The present experiments show how successful our visual system is in moderating the effect of visual noise. Information about gist may facilitate constructive processing, enabling the detection of occluded pictures in RSVP. Combining this top-down facilitation and amodal completion, the visual system appears to excel at tolerating occlusion noise.

Acknowledgments

Part of the present research was presented at Vision Sciences Society 6th Annual Meeting in 2006. This work was supported by grant MH47432 from the National Institute of Mental Health.

Commercial relationships: none.

Corresponding author: Ming Meng.
Email: mmeng@mit.edu.
Address: 46-4089, 77 Massachusetts Ave., Cambridge MA 02139, USA.

References

Bar, M. (2004). Visual objects in context. *Nature Reviews, Neuroscience, 5,* 617–629. [PubMed]

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology, 20,* 38–64. [PubMed]

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143–177. [PubMed]

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. [PubMed]

Braun, J. (2003). Natural scenes upset the visual applecart. *Trends in Cognitive Sciences, 7,* 7–9. [PubMed]

Bruno, N., Bertamini, M., & Domini, F. (1997). Amodal completion of partly occluded surfaces: Is there a mosaic stage? *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1412–1426. [PubMed]

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4,* 170–178. [PubMed]

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 109–127. [PubMed] [Article]

Cox, D., Meyers, E., & Sinha, P. (2004). Contextually evoked object-specific responses in human visual cortex. *Science, 304,* 115–117. [PubMed]

Davidoff, J. B., & Ostergaard, A. L. (1988). The role of colour in categorial judgements. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 40,* 533–544. [PubMed]

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: A study in monkeys and humans. *Vision Research, 40,* 2187–2200. [PubMed]

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107–117. [PubMed]

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General, 121,* 275–277. [PubMed]

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601. [PubMed]

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1476–1492. [PubMed]

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7*(1):10, 1–29, http://journalofvision.org/7/1/10/, doi:10.1167/7.1.10. [PubMed] [Article]

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition, 12,* 893–924.

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition, 12,* 878–892.

Gold, J. M., & Shubel, E. (2006). The spatiotemporal properties of visual completion measured by response classification. *Journal of Vision, 6*(4):5, 356–365, http://journalofvision.org/6/4/5/, doi:10.1167/6.4.5. [PubMed] [Article]

Guttman, S. E., Sekuler, A. B., & Kellman, P. J. (2003). Temporal variations in visual completion: A reflection of spatial limits? *Journal of Experimental Psychology: Human Perception and Performance, 29,* 1211–1227. [PubMed]

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50,* 243–271. [PubMed]

Intraub, H. (1997). The representation of visual scenes. *Trends in Cognitive Sciences, 1,* 217–222.

Intraub, H. (1999). Understanding and remembering briefly glimpsed pictures: Implications for visual scanning and memory. In V. Coltheart (Ed.), *Fleeting memories: Cognition of brief visual stimuli* (pp. 47–70). MIT Press: Cambridge, MA.

Joseph, J. S., & Nakayama, K. (1999). Amodal representation depends on the object seen before partial occlusion. *Vision Research, 39,* 283–292. [PubMed]

Kanizsa, G., & Gerbino, W. (1982). Amodal completion: Seeing or thinking? In J. Beck (Ed.), *Organization and representation in perception* (pp. 167–190). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kayser, C., Körding, K. P., & König, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Current Opinion in Neurobiology, 14,* 468–473. [PubMed]

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 9596–9601. [PubMed] [Article]

McKeeff, T. J., Remus, D. A., & Tong, F. (2007). Temporal limitations in object processing across the human ventral visual pathway. *Journal of Neurophysiology, 98,* 382–393. [PubMed]

Murray, R. F., Sekuler, A. B., & Bennett, P. J. (2001). Time course of amodal completion revealed by a shape discrimination task. *Psychonomic Bulletin & Review, 8,* 713–720. [PubMed]

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Encyclopedia of neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41,* 176–210. [PubMed]

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42,* 145–175.

Ostergaard, A. L., & Davidoff, J. B. (1985). Some effects of color on naming and recognition of objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 579–587. [PubMed]

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory and Cognition, 3,* 519–526.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. [PubMed]

Potter, M. C. (1975). Meaning in visual search. *Science, 187,* 965–966. [PubMed]

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 509–522. [PubMed]

Potter, M. C. (1999). Understanding sentences and scenes: The role of conceptual short term memory. In V. Coltheart (Ed.), *Fleeting memories: Cognition of brief visual stimuli* (pp. 13–46). Cambridge, MA: MIT Press.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81,* 10–15. [PubMed]

Potter, M. C., Staub, A., O'Connor, D. H., & Potter, M. C. (2004). Pictorial and conceptual representation of glimpsed pictures. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 478–489. [PubMed]

Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly-presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1163–1175. [PubMed]

Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 41,* 797–827. [PubMed]

Rauschenberger, R., Liu, T., Slotnick, S. D., & Yantis, S. (2006). Temporally unfolding neural representation of pictorial occlusion. *Psychological Science, 17,* 358–364. [PubMed]

Rauschenberger, R., & Yantis, S. (2001). Masking unveils pre-amodal completion representation in visual search. *Nature, 410,* 369–372. [PubMed]

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8,* 368–373.

Ringach, D. L., & Shapley, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision Research, 36,* 3037–3050. [PubMed]

Rousselet, G. A., Macé, M. J., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision, 3*(6):5, 440–455, http://journalofvision.org/3/6/5/, doi:10.1167/3.6.5. [PubMed] [Article]

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges—Evidence for time-scale-dependent and spatial-scale-dependent scene recognition. *Psychological Science, 5,* 195–200.

Sekuler, A. B., & Palmer, S. E. (1992). Perception of partly occluded objects—A microgenetic analysis. *Journal of Experimental Psychology: General, 121,* 95–111.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1,* 261–267.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31,* 137–149. [PubMed]

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network, 14,* 391–412. [PubMed]

Wolfe, J. M. (1998). Visual memory: What do you know about what you saw? *Current Biology, 8,* R303–R304. [PubMed]

Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 899–911. [PubMed]

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 141–145.