

Does disruption of a scene impair change detection?

Kazuhiko Yokosawa

Department of Psychology, University of Tokyo,
Tokyo, Japan



Hidemichi Mitsumatsu

Department of Psychology, University of Tokyo,
Tokyo, Japan



When we view a scene, we generally feel that we have a rich representation of that scene. Recent research has shown, however, that we are unable to detect relatively large changes in scenes, which suggests an inability to retain the visual details from one scene view to the next. In the present study, we investigated whether we can retain and make use of global and semantic information from a scene in order to efficiently detect changes from one scene to the next. Results indicated that change detection was practically independent of scene disruption with one exception. Better performance in the meaningful scenes was observed only in the whole-scene presentation condition where the participants knew that the stimulus was extracted from the meaningful scene.

Keywords: scene recognition, visual memory, change blindness

Introduction

When we view a scene, we are generally under the impression that we have a rich representation of that scene; [Potter \(1976\)](#) has confirmed that we are able to rapidly identify objects within a scene. It is also known, however, that we are unable to detect relatively large changes between two images during saccades ([Grimes, 1996](#)). Our inability to detect change is called change blindness.

[Rensink, O'Regan, and Clark \(1997\)](#) have developed a new paradigm referred to as the flicker paradigm and have shown that change blindness is not caused by a saccade-specific mechanism. They presented two scene images alternately with brief blanks in between and repeated this until participants reported a change. Their participants found it difficult to note any changes between the two images, which suggests that unless people make an effort to store visual representations to short-term memory, the details of visual representations are easily lost presumably because these details are replaced by the substitution of subsequent images.

So far, there have been two main approaches to research on change detection. Some studies have used scene stimuli (e.g. [Grimes, 1996](#); [Rensink et al., 1997](#)), while others have used non-scene stimuli such as random dot matrices ([Phillips, 1974](#)), letters ([Pashler, 1988](#); [Smilek, Eastwood, & Merikle, 2000](#)), and common objects ([Simons 1996](#)). For example, [Phillips \(1974\)](#) investigated the contribution of iconic memory to change detection by manipulating inter-stimulus interval (ISI), and demonstrated that iconic memory contributes to change detection only within short ISI (100 ms). [Smilek et al. \(2000\)](#) argued that unattended changes play a functional role in guiding focal attention, and [Simons](#)

(1996) reported that changes in spatial configuration are identified more accurately than those of identity.

Although valuable with regard to their specific topics, these studies provided little evidence as to whether the mechanism that mediates change detection may be identical between studies using scene stimuli and those using non-scene stimuli. Using scene stimuli, [Rensink et al. \(1997\)](#) manipulated the targets' prominence and reported that changes in targets of central interest were detected quicker than those of marginal interest. Prominent targets serve as key objects to represent the scene. Thus, it is suggested that changes in key objects are detected faster than changes in non-key objects. In contrast, [Hollingworth and Henderson \(2000\)](#) manipulated the semantic relationship between the target and the background that provides the scene. A target inconsistent with its background (for example, a fire hydrant in a living room) was detected faster than a consistent target (a chair in a living room).

The studies of [Rensink et al. \(1997\)](#), and [Hollingworth & Henderson \(2000\)](#) suggest that change detection depends on the relationship between the target and the scene. That is, if the targets are key objects in the representation of the scene or if these targets are semantically inconsistent with the scene, detection of those targets is facilitated. However, there remains the question as to whether disruption of the scene, regardless of target prominence or consistency, affects change detection. In [Rensink et al.'s \(1997\)](#) and [Hollingworth & Henderson's \(2000\)](#) studies, participants were always able to represent the scene, so it is unclear whether disruption of the scene makes a difference in change detection. The scene elicits the observers' expectations of what and where objects are likely to appear in a scene. Such expectations might be sufficient to enable an efficient search for change throughout the image. Thus, it is likely that the global and semantic information of a scene

facilitates change detection. However studies of change blindness reveal that subjects have a very limited ability to integrate visual information from one view to the next and highlight our inability to integrate the global and/or semantic description of the scene between successive views. If people cannot retain this description in successive views, it is difficult to establish detection of changes between scenes.

In the present study, we manipulated the global and semantic information of a scene and showed that disruption of the scene did not impair change detection in almost all conditions.

Experiment 1

Biederman (1972), and Biederman, Glass, and Stacy (1973) have reported that objects are recognized more accurately and quickly when they are presented in normal rather than in jumbled images. The jumbled images in these studies were created by dividing normal images into six sections and then rearranging them randomly. As a result, the jumbled images disrupted the scene. These studies suggest that representation of the scene facilitates object representation. In experiment 1 of the present study, we attempted to determine whether similar effects occur with regard to change detection as a result of jumbled images. The RTs of change detection were measured using a flicker paradigm (Rensink et al., 1997).

Methods

Participants

Seventeen young adults (mean age, 21.4) participated. All reported normal or corrected-to-normal vision.

Apparatus and Stimuli

Presentation of the stimuli and recordings of responses were controlled by a Macintosh G3 computer. Stimuli were displayed on a 19-inch color monitor. The following experiments used the same apparatus. Stimuli (23° x 18°) were color photographs of common scenes (e.g. photographs of parks, shopping streets). Of the 100 images created, 99 were used. Three experimental conditions were employed. Under the normal condition, normal photographs were used. Under the jumble 6 condition, photographs were divided into six sections, and the six sections were rearranged, with only the section that included the change region remaining in its original position. Under the jumble 24 condition, photographs were divided into 24 sections, and the 24 sections were rearranged, with only the section which included the change remaining in its original position. Under all conditions, black lines were drawn along the boundary regions of the 24 sections. This condition was introduced to control for the effect of boundary lines under the jumble condition (See Figure 1). Modified

images were created by adding one change to the original images. The extent of change was limited so as not to exceed one section created by dividing the images into 24 sections. There were three types of changes made (color change, positional change, and absence of object). Modifications of images were made so that changes were clearly visible once participants noticed them. We avoided subtle changes as much as possible. The degree of interest was tested for every object being changed in 99 of the 100 scenes used in all three experiments plus one scene only in experiments 2 and 3. Interest was determined via an independent pilot experiment in which five naïve participants provided a brief verbal description of each scene. Following the example of Rensink et al. (1997), central interests were defined as objects and areas mentioned by three or more observers. As a result, 17 percent of 100 objects turned out as of central interest, if we used a broad criterion. However, the change for objects of central interests was limited to a relatively small part, because the extent of change did not exceed one section created by dividing the images into 24 sections. For example, participants selected a motorbike to be of central interest, which occupied many parts of 24 sections in a scene, but the changing part was its rearview mirror, which no one selected. If we used a narrow criterion, any object would not reach the status of having central interest.

Procedure

Participants were seated in front of a computer monitor, and the viewing distance was fixed at 57 cm by a chin and forehead rest. Each trial was started by pressing the computer's mouse button. After participants started the trial, the original image (250 ms) repeatedly alternated with a modified image (250 ms), with a brief blank (250 ms) inserted between the images. The blank stimulus was painted black. The participants' task was to press the mouse button as soon as they saw the change and then point with the mouse to the section that included the changing region. The pointing task served as confirmation of correct change detection. Each pair of images in all trials included one changing region. A few trials (< 2%) were eliminated from the analysis because of incorrect change detection. Participants had to respond within a time limit of 1 minute. Response times over 1 minute were recorded as being of 1-minute duration. In total, 99 trials were conducted; thus, there were 33 trials under each condition. The stimulus set was randomly divided into three sets at the start of each experiment, each including 33 pairs of images. Each set was assigned to one of three conditions. In each stimulus set, three types of change (color, position, presence or absence of object) were included. The three stimuli set were counterbalanced across conditions. The number of trials for each types of change was not controlled, since the type of change was not the main concern for the present study (interested readers can refer to Aginsky & Tarr, 2000;



Figure 1. Examples of stimuli under each jumbled condition. (a): normal condition. (b): jumble 6 condition. (c): jumble 24 condition. In these stimuli, two wheels of a baby carriage disappeared in the modified images.

Rensink et al, 1997; Shore & Klein, 2000). The participants did not know what type of change they were searching for during the trials, though they did know there were three types of change in this experiment.

Each participant was subjected to all three conditions. Three conditions were mixed in a block. Each image was used in only one trial.

Results and Discussion

The mean error rate was 1.7 % and did not differ between three conditions, $F(2,16)=0.2$, $MSE=7.7E-5$, $p>.7$. Mean RTs were calculated by averaging the median RT of each participant. The following experiments also used this calculation. RTs of the three conditions are shown in Figure 2. ANOVA revealed no significant difference between the three conditions, $F(2,32)=0.3$, $MSE=46434.0$, $p>.74$.

These results indicate that representation of the scene did not facilitate change detection. However Biederman et al. (1973) demonstrate that objects are recognized more quickly when presented under the normal condition rather than under the jumble condition. The difference between the present study and that of Biederman et al. could be attributed to the difference in the tasks. In the object-identification task, participants had to make a

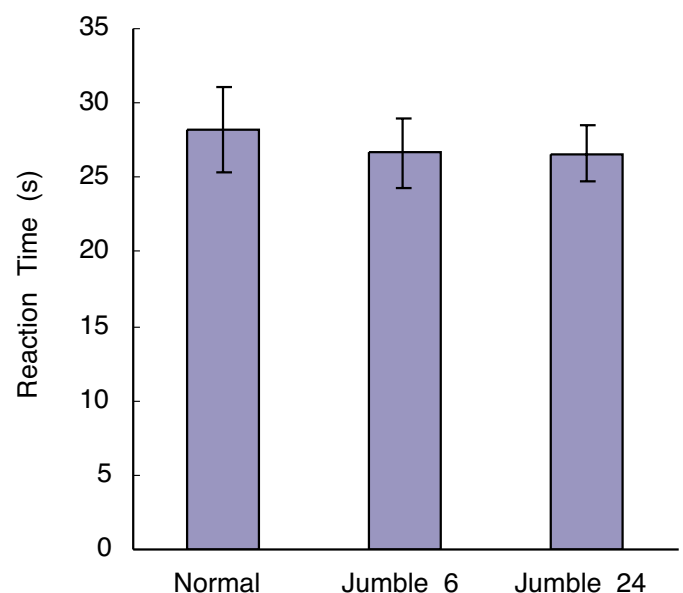


Figure 2. Results of Experiment 1. Error bars indicate a standard error.

semantic judgment, while in the change-detection task they had to focus primarily on physical properties such as color and position rather than semantic information. The successive views provide semantic information about what and where objects are likely to appear in a scene. Disruption of the change detection task might have been attenuated because of its requirement for physical-property processing.

Our results agree with those based on the flicker paradigm of Shore & Klein (2000), who reported that when scene images are inverted change detection RTs do not differ from those obtained when scene images are upright. Inverting scene images is similar to jumbling in that it makes it difficult to understand the meaning of the scenes.

Experiment 2

In experiment 1, disruption of the scene by jumbling appeared not to impair change detection. However, jumbling is not the sole available method for disrupting the scene. Several studies describe how they have eliminated parts of scenes or objects (Antes, Penland, & Metzger, 1981; Bar & Ullman, 1996; Boyce, Pollatsek, & Rayner, 1989). If this elimination is spread throughout the entire image, representation of the scene becomes difficult.

However, partially eliminated images have a smaller drawing area than the original images (Boyce et al., 1989) and so one cannot legitimately compare the two. The studies of Bar & Ullman (1996), and Antes et al. (1981) were able to circumvent this problem, by not directly comparing the accuracy between the two. Instead, Bar & Ullman (1996) examined the effects on object recognition of the object's spatial configuration in partially eliminated images without displaying the original images, showing that better performance was observed in the original configuration condition.

Thus, in experiment 2, we investigated how eliminating parts of scenes might effect the disruption of change detection. Scene stimuli were divided into 24 sections and we manipulated the number of the image sections displayed. The numbers of sections displayed were 3, 10, 17, and 24. Comparison was made between the search slopes to circumvent the problem of displayed area differences. The search slope was defined as RT difference between the consecutive numbers of sections displayed. The manipulation of the numbers of sections displayed paralleled that of set size in visual search studies except that the 24-sections condition had a special status in the present experiment. If partial elimination impairs change detection, one would expect change detection to become more efficient as the numbers of displayed sections increased and the scene becomes more representable. Otherwise, search efficiency would be constant, regardless of the number of sections displayed.

Methods

Participants

Eighteen young adults (mean age, 21.8) participated. All reported normal or corrected-to-normal vision.

Stimuli

Normal and jumbled images were used as stimuli. These stimuli were divided into 24 sections and we manipulated the number of the image sections displayed. The numbers of sections to be displayed were 3, 10, 17, or 24. Spaces between displayed sections were painted black. A sample of normal stimuli is shown in Figure 3.

Procedure

Ten participants were subjected to the normal condition and 8 the jumble condition. The participants' task was the same as in experiment 1. Each participant participated in one block consisting of 100 trials. The numbers of sections to be displayed were 3, 10, 17, or 24. For each image, the number of the sections displayed differed across the participants. For example, an image was used in 3-sections condition for some participants, and the same image was used 10, 17, or 24-sections condition for other participants. There were 25 trials for each condition. Trials of these numbers of sections were intermixed in a block. Thus, analysis was conducted in a within-participants design.

Results and Discussion

Mean RTs of the normal condition as a function of the number of sections are shown in Figure 4. ANOVA revealed that there were differences in RTs between four conditions ($F(3,27)=31.6$, $MSE=30075$, $p<.0001$). Post-hoc analysis (Student-Newman-Keuls analysis) showed that there were difference in RTs between 3 sections and 7 sections, and between 7 sections and 10 sections, and between 10 sections and 17 sections ($p<.05$). However, there was no difference in RTs between 17 sections and 24 sections.

Mean RTs of the jumble condition are shown in Figure 5. ANOVA revealed that there were differences in RTs between four conditions ($F(3,21)=16.4$, $MSE=58849$, $p<.0001$). Post-hoc analysis showed the same as that in the normal condition. There was no difference in RT between 17 sections and 24 sections, although the other three differences were significant ($p<.05$).

Under both conditions, there might be a smaller increase in RT from 17 sections to 24 sections, compared to the linear RT increase from 3 sections to 17 sections. This might be explained as a kind of ceiling effect, because 24-section (that is, the whole scene) condition

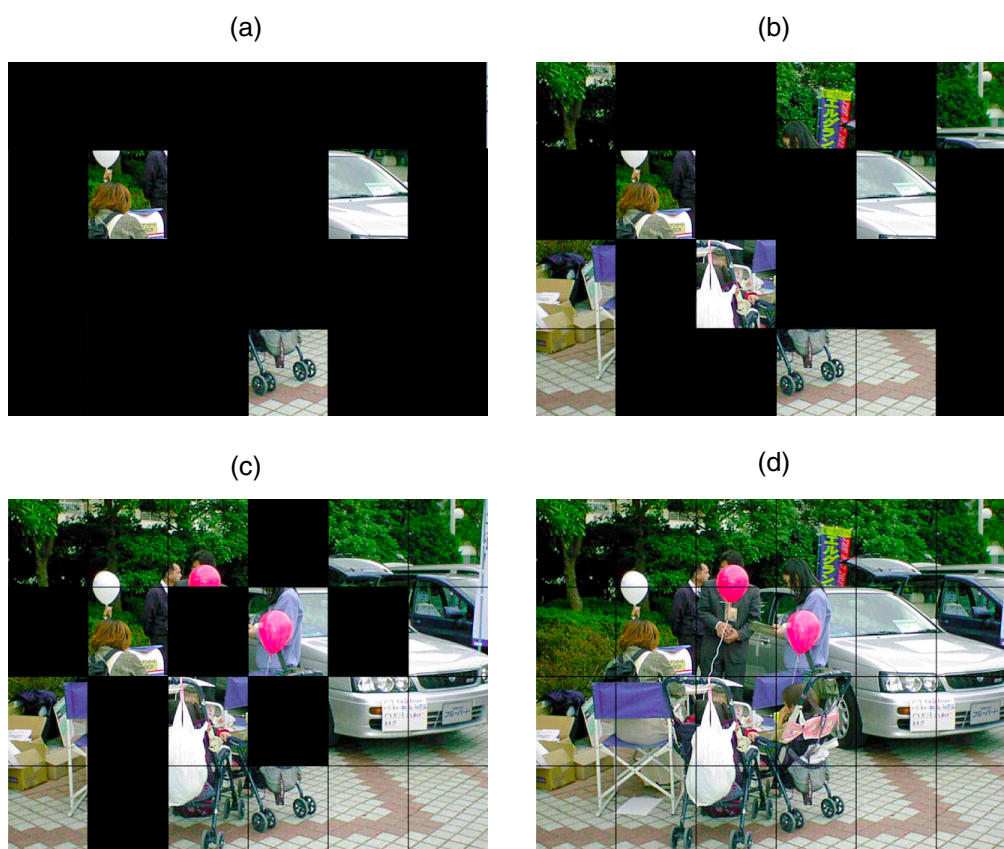


Figure 3. Examples of normal stimuli under each partial elimination condition. (a): 3 sections. (b): 10 sections. (c): 17 sections. (d): 24 sections.

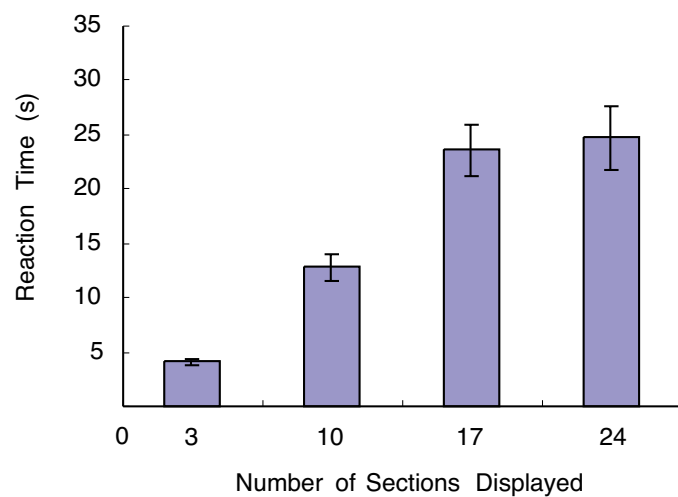


Figure 4. Results of the normal condition in Experiment 2. Error bars represent a standard error.

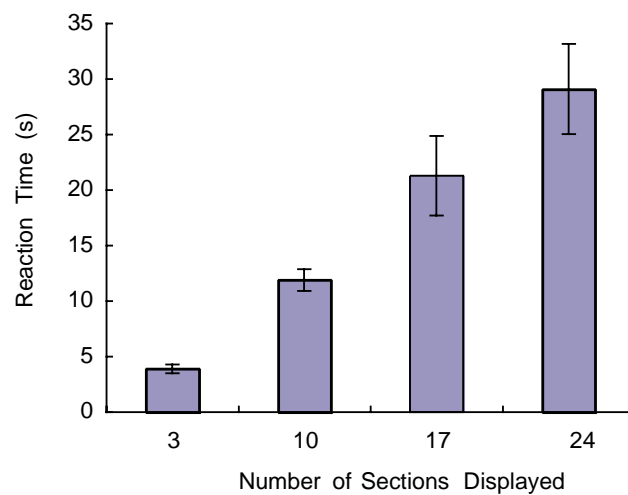


Figure 5. Results of the jumble condition in Experiment 2. Error bars represent a standard error.

had a special status as a visual search display. However, a further ANOVA revealed that the difference in slopes was significant only in the normal condition, $F(2,18)=4.1$, $MSE=62345.9$, $p<.05$. The slope from the 17 to 24-section condition was lower compared to the other two slopes ($p<.05$; Student-Newman-Keuls analysis). All slopes in the jumble condition showed a linear increase in RTs but the difference was not significant. These different statistical results suggested that the completion of a scene by displaying the whole set of 24 sections facilitates change detection a little and that images with a partial set of sections displayed do not have a facilitating effect even under the 17-sections condition. This set size effect in the jumble condition is consistent with the linear increase of RTs in change detection as the numbers of letters and digits also increase (Smilek, Eastwood, & Merikle, 2000).

In experiment 1, no difference in change detection was observed when comparing between normal and jumbled images. Experiment 2 showed that the efficiency may be slightly high when the whole set of sections was displayed, compared to when the partial set was displayed.

Jumbling and partial elimination were introduced to disrupt the scene. However, jumbling had no effect on change detection, whereas partial elimination slightly impaired change detection. One important difference between the two experiments was that normal and jumbled conditions were intermixed in experiment 1, whereas, in experiment 2, blocked design was used. If the critical factor for efficient change detection that emerged in 24-sections condition in experiment 2 was the blocked design of normal images, then no efficient change detection will emerge when the mixed design is introduced. To examine this possibility, experiment 3 was conducted in which the 17-sections condition and 24-sections condition were introduced both in normal and jumbled images. Those normal and jumbled images were mixed in one block. If the blocked design of normal images was critical for the emergence of efficient change detection, RTs of the 24 section condition of the normal images will no longer be as fast as those of the 17 sections condition of the normal images when the mixed design was introduced.

Experiment 3

Methods

Participants

Nine young adults (mean age, 21.5) participated. All reported normal or corrected-to-normal vision.

Procedure

The participants' task was the same as in experiments 1 and 2. Unlike in experiment 2, both normal and

jumbled images were used. The numbers of sections to be displayed were 17, or 24. Each participant participated in one block consisting of 100 trials. In half of the 100 trials, normal images were presented. In the other half, jumbled images were presented. Half of the 50 trials of normal images, and half of the 50 trials of jumbled images, were composed of the 17-sections condition. The remaining 50 trials of normal and jumbled images were composed of the 24-sections conditions. Trials of these numbers of sections were intermixed in a block. Thus, analysis was conducted in a within-participants design.

Results and Discussion

Mean RTs of both the normal and jumble conditions as a function of the number of sections are shown in Figure 6. RTs of the 24 sections condition were longer than those of the 17 sections condition both in normal and jumbled images, $F(1,8)=13.5$, $MSE=15635.1$, $p<.01$, $F(1,8)=34.3$, $MSE=5803.1$, $p<.001$. Two-way ANOVA (image type and the number of sections as factors) revealed that the main effect of the number of sections was significant, $F(1,8)=56.7$, $MSE=14184.4$, $p<.0001$.

Experiment 3 was conducted to examine whether the mixed design of normal and jumbled images results in an increase in RT from the 17 to 24-sections condition in normal images. The result confirmed a RT increase in both the normal and jumble conditions. These results indicate that the critical factor for efficient change detection lies in whether the design was mixed or blocked.

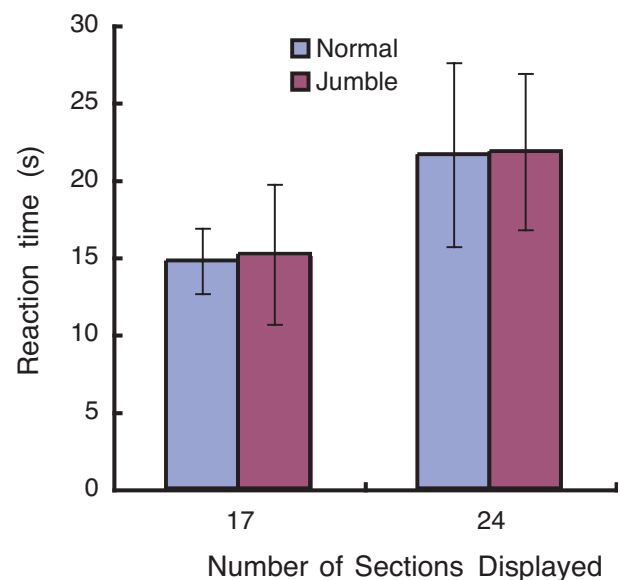


Figure 6. Results of Experiment 3. Error bars represent a standard error.

General Discussion

This study demonstrates that global and semantic information is not related to change blindness. Surprisingly, the disruption of the scene by jumbling and elimination hardly impaired change detection.

Nakayama (1990) described a theoretical framework in which a wide spatial distribution is equal to a visual analysis of global scene characteristics, whereas a narrow spatial focus is invariably tied to local visual analysis. Based on his framework, the attentional focus for a new scene always begins at global and moves to local as required. This position finds empirical support in the work of Biederman (1972) and Biederman, Glass, and Stacy (1973), where they demonstrate the inferiority of object detection in jumbling images. There is a possibility that object detection (Biederman, 1972) is a different visual process from object change detection, where participants have to focus primarily on physical properties rather than semantic information.

However, in a change detection task, Austen & Enns (2000) manipulated the detail level of display items by using compound letters. The results showed that when attention was distributed among multiple items, changes at the global level were detected more rapidly and accurately than changes at the local level. Austen & Enns (2000) suggest that the global level of representation may be a default for the visual system. Our results did not support this view in change detection. Because no differences in RTs between the normal and jumble conditions were found in experiments 1 and 3, it is likely that the visual system conducted the same sequential search in both the normal and jumble conditions using some sort of mechanism independent of the scene. This difference between our study and that of Austen & Enns (2000) might be because the global level predominance is not available for the whole scene, but only for whole objects like large letters formed by arranging smaller letters.

In experiment 2, however, the blocked design was introduced, resulting in efficient change detection in the 24-sections condition compared to the 17-sections condition, while there was linear increase in RTs from the 3 to 17-sections conditions. Although the statistical analysis did not support this tendency strongly, the blocked design might have enabled a high degree of top-down control. When a mixed design is introduced, a high degree of top-down control becomes difficult, resulting in a sequential search of scene-independent mechanism. This was confirmed by the results of experiment 3, resulting in sequential search in the 24-sections condition of normal images. Thus, when the participants were uncertain whether coherent images were presented in consecutive trials, only a sequential search was performed. Austen & Enns (2000) showed that change detection depends critically on the expectancy of the observer for

both the focused and distributed attention condition. Future studies will be needed to address the effect of this expectancy for the scene.

Conclusions

The present study investigated whether humans could utilize global and semantic information to detect the change effectively. This study has demonstrated that this kind of information does not help improve change blindness. Surprisingly, the disruption of the scene by jumbling and elimination hardly impaired change detection. When the meaningless scene due to jumbling was mixed in a block, the participants conducted sequential search even when the meaningful scene was presented. However, when they knew that the stimulus was extracted from the meaningful scene, the fact that they were able to have a grasp of the scene, slightly facilitated the change detection.

Acknowledgments

This research was supported by a Grant-in-Aid for Scientific Research No.13224021 awarded to Kazuhiko Yokosawa from the Japan Society for the Promotion of Science. We would like to express our thanks to Reiko Suzuki for her help with this research.

Commercial relationships: None.

References

- Aginsky, V., & Tarr, M.J. (2000). How are different properties of a scene encoded in visual memory? *Visual Cognition*, 7, 147-162.
- Antes, J. R., Penland, J.G., & Metzger, R. L. (1981). Processing global information in briefly presented pictures. *Psychological Research*, 43, 277-292. [PubMed]
- Austen, E. & Enns J. T. (2000). Change detection: Paying attention to detail. *Psyche*, 6, 11.
- Bar, M., & Ullman, S. (1996). Spatial context in representation. *Perception*, 25, 343-352. [PubMed]
- Biederman, I. (1972). Perceiving real-world scenes. *Science*, 177, 77-80. [PubMed]
- Biederman I., Glass, A.L., & Stacy E.W. Jr. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology*, 97, 22-27. [PubMed]
- Boyce, S. J. Pollatsek, A., & Rayner, K. (1989). Effect of background information on object Identification. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 556-566. [PubMed]

- Grimes, J. (1996). On the failure to detect changes in scenes across saccade. In K. Akins (Ed.), *Perception* (Vancouver Studies in Cognitive Science, Vol. 5, pp. 89-110). New York: Oxford University Press.
- Hollingworth, A., & Henderson, J. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition*, 7, 1/2/3, 213-235.
- Nakayama, K. (1990). The iconic bottleneck and the tenuous link between early visual processing and perception. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 411-422). Cambridge, UK: Cambridge University Press.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, 44, 369-378. [PubMed]
- 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 & Memory*, 2, 509-522. [PubMed]
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive change in scenes. *Psychological Science*, 8, 368-373.
- Shore, D. I., & Klein, R. M. (2000). The effects of scene inversion on change detection. *The Journal of General Psychology*, 127, 27-43. [PubMed]
- Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science*, 7, 301-305.
- Simons, D. J. & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Science*, 1, 261-267.
- Smilek, D., Eastwood, J. D., & Merikle, P. M. (2000). Does unattended information facilitate change detection? *Journal of Experimental Psychology: Human Perception and Performance*, 26, 480-487. [PubMed]