12

Multiple Object Permanence Tracking: Maintenance, Retrieval and Transformation of Dynamic Object Representations

Jun Saiki
Kyoto University
Japan

1. Introduction

Our visual world is composed of multiple dynamic objects with various visual features. For efficient interaction with the world, the visual system needs to keep binding of object features and update them as their dynamic changes. Given severe limitation of our visual short-term memory (VSTM) (Luck & Vogel, 1997; Pashler, 1988), it is a challenge to understand how the visual system deals with this binding problem in dynamic environment. In this chapter, I will review research on this issue, mainly focused on experimental studies using the paradigm called "multiple object permanence tracking" (Imaruoka et al., 2005; Saiki, 2002, 2003a, 2003b, 2007; Saiki & Miyatsuji, 2007, in press).

Transformation of object representations in dynamic environment has been investigated mainly using multiple object tracking task (MOT) (Pylyshyn & Storm, 1988; Shcoll & Pylyshyn, 1998). In MOT, a dozen of identical objects (dots) are randomly moving around on the display, and observers required to track a subset of these objects. Although research with MOT revealed various properties of object representations used by visual cognition mechanisms, the issue of binding various object features into an object representation remains unclear, because MOT only manipulates spatiotemporal location of objects, not other features. To address the issue of feature binding in dynamic environment, multiple object permanence tracking (MOPT) task used objects with different colors and shapes, and investigated how these objects' features are bound together in dynamic displays.

This chapter will describe five topics investigated with MOPT paradigm. First, how feature binding is maintained over dynamic movement of multiple objects? A series of experiments revealed that our ability of keeping binding of objects' color, shape and their spatiotemporal locations was significantly impaired when objects move (Saiki, 2003a, 2003b). Importantly, object motion was quite slow and predictable, so that the impairment was not due to failure of tracking of objects per se. Second, memory for feature binding was evaluated more strictly (Saiki & Miyatsuji, 2007). Switch detection task used in previous work showed that task performance was quite good when objects were stationary. However, simple switch detection task may overestimate our ability, and a more strict test revealed that even if objects were stationary, our ability of maintaining feature binding was much more limited than previous studies suggested. Third, is memory maintenance, or memory retrieval responsible for the performance impairment in MOPT task? To test this, I used retrieval cues in novel paradigms that directly evaluate the memory for triple conjunctions; type identification and relevant-feature switch detection, in comparison with a simple change-detection task (Saiki & Miyatsuji, in press). We found that a retrieval cue provided no benefit with the triple conjunction tasks, but significant facilitation with the change-detection task, suggesting that low capacity estimates of object file memory in VSTM reflect a limit on maintenance, not retrieval. Fourth, are these findings specific to arbitrary combination of shape and color, or more general including ordinary objects? In other words, how does prestored knowledge on color-shape binding in everyday objects affect feature binding in VSTM? To address this issue, I used everyday object such as lobster and frog, and showed that there is still significant impairment in memory for feature binding (Saiki, 2007). At the same time, there were significant differences in observer's performance. Finally, we have investigated neural correlate of feature binding in VSTM. An fMRI experiment using MOPT task revealed that in addition to frontoparietal network known to be involved in various attention related tasks, we have found significant activation of anterior prefrontal cortex, suggesting that maintenance of feature binding in visual working memory requires additional processing in anterior prefrontal cortex, which is markedly different from activation observed with the simple MOT task (Imaruoka et al., 2005).

Based on these findings, unlike the widely accepted view that VSTM has the capacity of 3-5 feature bound object representations, our ability to keep feature binding in VSTM is more limited. Relationship between MOPT experiments and other studies with various tasks, theoretical implications, and possible implications for human interface and other human factor applications will be discussed.

2. Multiple Object Permanence Tracking: Experimental Paradigm

2.1 General design

Multiple object permanence tracking (MOPT) task is conceptually a mixture of MOT and change detection tasks used in studies on VSTM. It deals with dynamic display as in MOT, but use objects defined with multiple features as in change detection tasks. Observers require to maintain feature bindings of objects, and to update as they move. The task can be decomposed into three aspects: objects to be maintained, spatiotemporal dynamics of objects, and behavioural tasks. I will summarize these aspects following the general description of stimulus configuration.

2.2 General description of stimulus configuration

Stimuli were a number of objects (usually defined by color and/or shape) configured on an imaginally circle, usually at eccentricity of 4 deg in visual angle. The objects were occluded by a gray windmill-shaped occluder, and the background was black. Objects and/or a windmill-shaped occluder smoothly rotated with constant angular velocities, so the sequence alternated visible and invisible periods regularly. Each stimulus sequence began with the visible state, followed by several alternations of visible and invisible periods. A switch event occurred at one period in the middle of the sequence (Figure 1).

2.3 Objects to be maintained

Each object was defined by color and shape. Many experiments used color manipulation alone with the same shape (disk). When shape was manipulated, either simple geometric shapes (disks, square, triangle, and pentagon) or natural objects (lobster, frog, banana, and violin) were used (Figure 2). Colors were usually four equiluminant colors. Combination of color and shape was arbitrary and randomly combined in each trial, except for the experiment in section 3.4 using natural objects. All objects in a single trial have different shape and color. The number of objects in a single trial was usually four, except for experiments in section 3.1 varying between two to six. Experiments manipulating the number of objects investigated the capacity of binding memory. A switch event (color, shape or color-and-shape) occurred at one period in the middle of the sequence.

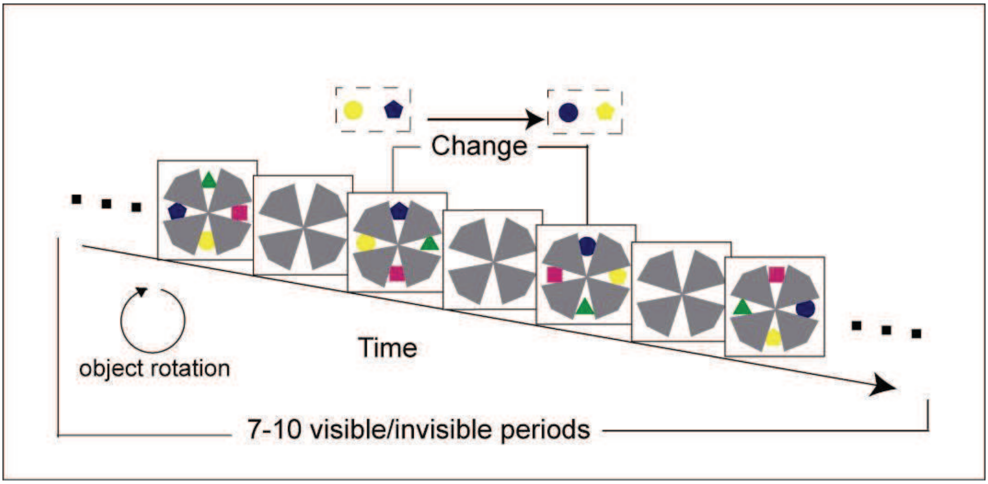


Figure 1. Schematic illustration of spatiotemporal stimulus configuration

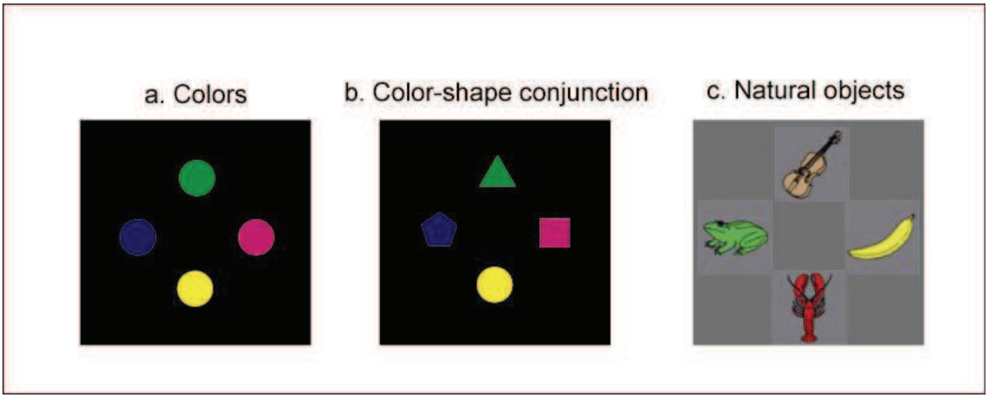


Figure 2. Sets of objects used in MOPT studies. Colors are not exactly matched to those used in experiments

2.4 Spatiotemporal dynamics of objects

Objects and/or a windmill-shaped occluder smoothly rotated with constant angular velocities, so the sequence alternated visible and invisible periods regularly. Each stimulus sequence began with the visible state, followed by several alternations of visible and invisible periods. Two rotation directions of the pattern (clockwise and counterclockwise) were used. The angular velocity of objects, from 0°/s (i.e., static) to 125°/s was manipulated by the relative motion of the objects and occluder, which kept the exposure and occlusion durations constant (Figure 3). Note that the fastest angular velocity was still much slower than the maximum velocity of approximately 360°/s in the simple location tracking task (Verstraten et al., 2000). Furthermore, regular rotation was completely predictable, unlike the standard MOT task (Pylyshyn & Storm, 1988). The occlusion duration was manipulated by the width of the occluder opening, such that the wider the opening, the longer the visible period (Figure 3).

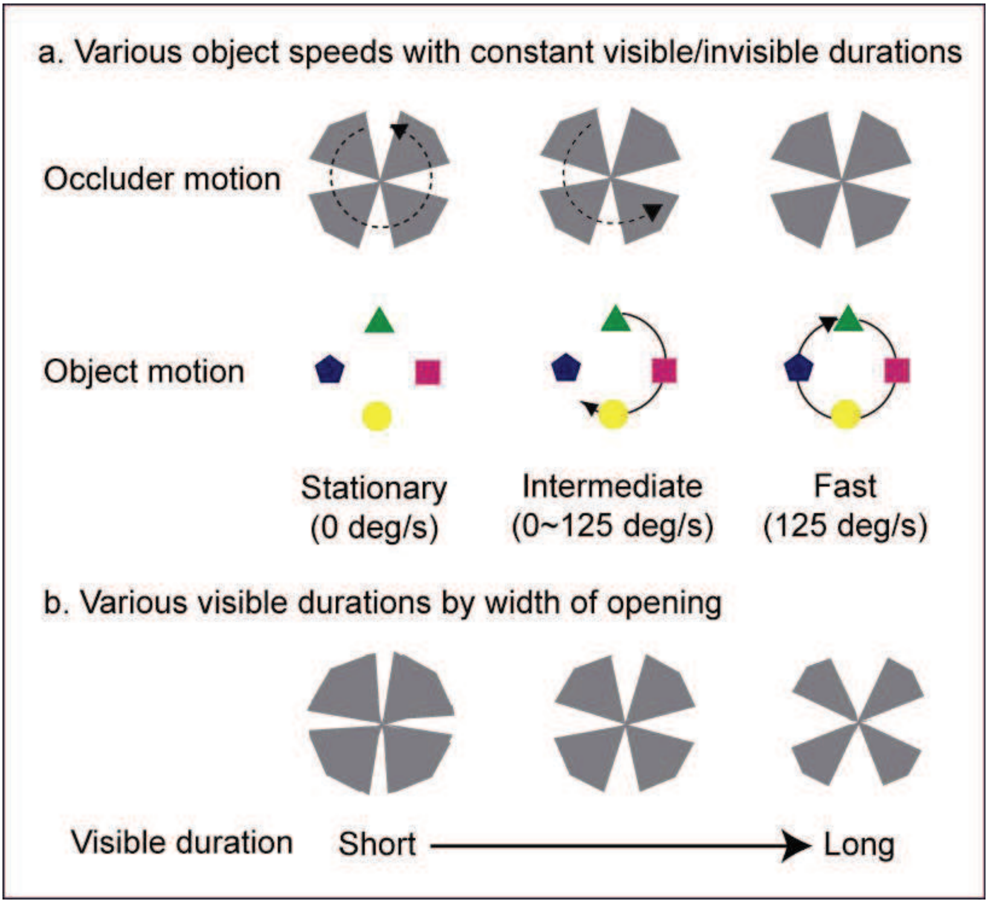
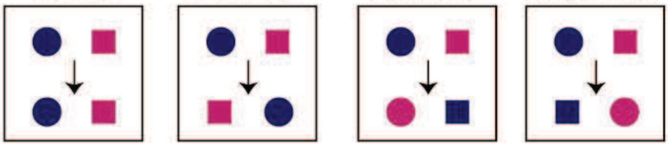


Figure 3. Illustration of spatiotemporal dynamics of objects

2.5 Behavioral task

Observers were asked to try to pay attention to the whole pattern throughout a trial, but the fixation was not monitored. They made various judgments by a key press, without correct feedback. There was no time pressure to make a response, and observers did not have to wait until the sequence ended, to make a response. To avoid verbal encoding, articulatory suppression was used. Each trial began with a beep that prompted articulatory suppression. Afterwards, the first frame of the sequence appeared on the screen and remained stationary. Five hundred ms later, the motion sequence began. Before experimental trials, observers had a block of several practice trials to familiarize themselves with the procedure.

The behavioral task is judgment about an event in the middle of MOPT sequence. When an object is defined by color and shape, four events are possible. Suppose a red square and blue circle make a change (Figure 4). The four possible change types are: no change (red square and blue circle); color change (blue square and red circle); shape change (red circle and blue square) and both change (blue circle and red square). Three different tasks were simple change detection, type identification, and relevant-feature switch detection. The simple change detection requires to judge yes when any switch occurs, which is the same as typical change detection task widely used in the literature. The type identification task requires participants to identify which event occurs in the stimulus sequence, as discrimination among four alternatives. In the relevant-feature switch detection task, the participant was instructed to monitor either color or shape, and required to judge whether the stimulus sequence included a switch event on the prespecified feature dimension. This task had only two response alternatives, as in a simple change-detection task, but required distinction between color- and shape-switch events. Figure 4 summarizes the mapping of events and responses, and I should note that the type identification and relevant-feature switch detection tasks evaluate memory for feature binding more strictly as detailed below.



Event Type	No Switch	Both Switch	Color Switch	Shape Switch
Task				
Simple Change Detection	No	Yes	Yes	Yes
Type Identification	None	Both	Color	Shape
Relevant-feature Switch Detection (color)	No	Yes	Yes	No
Relevant-feature Switch Detection (shape)	No	Yes	No	Yes
	Response Mapping			

Figure 4. Mapping of event types and responses in various tasks used in MOPT experiments

2.6 Specific manipulations

The basic paradigm of MOPT was modified in various ways to investigate characteristics of binding in VSTM. Besides those described above, two manipulations are worth mentioning here. One is the number of switching events, with which we investigated spontaneous strategies of selective attention. The other is cueing a target object (precue and postcue) to investigate whether performance impairment reflect memory maintenance or retrieval.

3. Studies using MOPT

3.1 Maintenance of feature binding over dynamic movement of multiple objects

The first series of experiments with MOPT (Saiki, 2003a; 2003b) investigated basic spatiotemporal characteristics of dynamic updating of multiple object representations. Saiki (2003a) primarily investigated dynamic updating of three objects with apparent motion displays. Unlike prototypical MOPT display as described in section 2, participants were shown sequences of 10 frames depicting a triangular pattern of three colored disks rotating by a certain angle per frame. The sequence was either regular clockwise or counterclockwise rotation throughout, containing one frame in which the locations of two colors were switched (color-switch), or containing one frame in which a new color was replaced with an old one (color-replacement). Participants were required to judge whether a sequence was regular or irregular without identifying its type (color-switch or color-replacement). Notice that detection of a color-switch needs memory for the conjunction of each disk's color and spatiotemporal location, whereas detection of a color-replacement does not. Thus, the performance for the color-switch condition is the critical measure of memory for the binding of color and spatiotemporal location in this paradigm. A pilot experiment with the equilateral triangle pattern rotating 60° per frame showed that color switch detection was difficult, while color replacement detection was extremely easy.

The difficulty in the color switch detection in the pilot experiment may not be due to color-location binding, but simply to failure in tracking pattern rotation. Saiki (2003a) used stimuli without such ambiguity in pattern motion, and investigated whether the difficulty in color switch detection could be overcome simply by making pattern motion unambiguous. Experiments 1 and 2 disambiguated the motion correspondence by using bilateral triangles, and smooth and continuous motion, respectively. The ambiguous condition was used as a baseline to evaluate the effect of disambiguation of motion correspondences. Overall, switch detection performance did not show any significant improvement in the unambiguous motion conditions over the ambiguous motion conditions. The difficulty in color switch detection is not due to the problem of perceiving colors with moving objects, because color replacement detection was almost perfect. Disambiguation of objects' motion by pattern configuration, smooth and continuous motion, and elimination of abrupt onset and offset is insufficient for successful color switch detection.

A second series of experiments in Saiki (2003a) investigated the effect of rotation angle on the switch detection performance. Experiment 3 showed that a reduction of the interframe rotation angle substantially improved the color switch detection performance. Within 45° interframe rotation, the hit rate for color switch was significantly better than the ambiguous baseline condition of 60° interframe rotation. Experiment 4 examined whether the effect obtained in Experiment 3 was due to the amount of spatial displacement, or the amount of angular displacement by enlarging the distances between the disks of a pattern, showing that this facilitatory effect was due to interframe rotation angle, not due to interframe spatial displacement. Experiment 5 further examined whether the effect obtained in the previous experiments was mediated by angular velocity or angular disparity by manipulating frame duration, suggesting that rotation angle, not angular velocity, determined the performance. Finally, Experiment 6 showed that the spatiotemporal predictability of locations is necessary for the facilitatory effects of reduced rotation angle. Note that these results are not simply reflecting the success or failure in object tracking, because Experiment 7 showed that object tracking was quite successful in this task setting.

A series of experiments revealed that even when the motion correspondences are unambiguous by the use of pattern configurations and continuous motion, and object tracking is successful, color switch detection performance is difficult; there was no significant improvement compared with the situations where motion correspondences were inherently ambiguous. At the same time, it has been revealed that color switch detection performance is critically dependent on the inter-frame rotation angle, and that a facilitatory effect occurred only when spatiotemporal predictability was satisfied.

As an extension of Saiki (2003a), Saiki (2003b) investigated the spatiotemporal characteristics of dynamic updating using smoothly moving multiple objects with an occluder. Objects were colored disks, and the binding of the object's color and its location should be dynamically updated. Experiment 1 investigated the effects of angular velocity of the pattern, which manipulated the objects' rotation speed with constant visible and invisible durations. Unlike Saiki (2003a) with all moving objects, this experiment parametrically varied object movement from stationary to a substantial speed (125°/s). Even within the range of successful object tracking and color perception, angular velocity strongly affected the observers' performance of color switch detection in the MOPT task, suggesting that color-location binding is quite difficult when objects are moving. The ROC analysis revealed that this effect was not due to response biases.

Experiment 2 examined the effect of occlusion duration to evaluate the "life-span" of object working memory in dynamic situations. Overall, both angular velocity and occlusion duration had significant effects on color switch detection performance, and these two factors are largely independent. The cost of object motion was significant even with a minimum occlusion period (40-ms), suggesting that object motion impairs color switch detection regardless of the length of the invisible period. The results of Experiment 2 suggest that in visual working memory, the transformation cost and retention cost, manipulated by object motion and occlusion duration, respectively, are largely independent.

Experiment 3 evaluated the "capacity" of dynamic object working memory in terms of the number of objects and the relationship between retention and processing costs. There were six objects, and observers were asked to track the color switch between target objects (2, 3, 4 or 6) prespecified at the beginning of each trial by flashing (Pylyshyn & Storm, 1988). A color switch occurred either between the target objects, or between the non-target objects, and observers were asked to ignore any color switches between non-targets. In this experiment, processing and retention costs were manipulated by angular velocity and the number of targets, respectively. Overall, the processing and retention costs were independent, and a significant effect of angular velocity was observed even in the 2- and 3-target conditions. The results of Experiments 2 and 3 showed that effects of motion were observed regardless of the retention costs, which is inconsistent with the view that motion affects general processing resource of visual working memory, and at least for the visual working memory measured by the MOPT paradigm, processing and retention are largely independent. Recently, some research on working memory has suggested the independence of processing and retention (Towse et al., 2000).

Because previous experiments investigated only color-location binding, it is unclear whether the findings reflect object level or single feature level representations. The final experiment used multidimensional objects defined by shape and color to investigate the dynamic updating of multidimensional feature binding. Objects had different shapes, as well as colors, and switch occurred with either color alone (color switch), shape alone (shape switch), or color and shape (object switch). The task was simple switch detection, and comparison of accuracy among different switch types can dissociate different hypotheses. If triple conjunction representations for objects are formed (object token hypothesis), there should be no difference among different switch types, because all these switch types involve the same amount of change in triple conjunction representations. In contrast, if a set of single conjunctions (color-location and shape-location) is formed (feature-location binding hypothesis), object switch detection will be more accurate than shape and color switches if both color-location and shape-location coding are fully available, because an object switch involves a switch of both bindings, whereas others involve only one of them. Moreover, if the availability of two types of conjunction coding is reduced due to object motion or other factors, the advantage of object switch will be reduced. Overall, results were consistent with the feature-location binding hypotheses. In the stationary conditions, object switch detection was significantly better than the color switch detection and the shape switch detection. In contrast, there was no advantage for object switch detection in the moving condition.

Our ability to maintain episodic representations of multiple objects in a completely predictable dynamic situation is limited. Objects' features are not bound together in a dynamic situation, even when their motion is quite slow and completely predictable and well within the range of ordinary object motion. This finding strongly suggests that previous findings obtained with static displays (Luck & Vogel, 1997; Vogel et al., 2001) and a dynamic multiple-object tracking task (Pylyshyn & Storm, 1988) may not reflect the function of common high level episodic representations such as object files. The dynamic maintenance of features has been used as an important hallmark of objectness in object-based attention literature (Tipper et al., 1994; Chun and Cavanagh, 1997; Valdes-Sosa et al., 1998); thus, the failure in dynamic updating of object features casts doubts on the proposal that visual working memory is object-based in a strong sense.

These results are largely consistent with recent evidence that the system of visual cognition works with much less memory than we previously believed (Ballard et al., 1997; Horowitz & Wolfe, 1998; Rensink et al., 1997). Unlike previous demonstrations, this work provides an experimental paradigm enabling parametric investigations of spatiotemporal characteristics of visual working memory, revealing some important findings. The present work has some implications for the issue of feature binding in visual cognition (Treisman, 1999). The extremely short life-span and limited capacity of memory for dynamic feature-location binding suggest that such binding is quite transient. It is well known that the binding problem is computationally quite difficult, especially in the case of multiple objects. The present findings may indicate that the visual system functions without solving a multiple-object binding problem. Instead of holding integrated representations of multiple objects, the visual system may bind perceptual features of a single object by attentional processing only when necessary (Rensink, 2000). Rensink (2000) reviewed the literature of change blindness and related phenomena, and proposed the notion of virtual representation, which provides only a limited amount of coherent structure, but provides it whenever requested, making it appear as if all the detailed, coherent structure is present simultaneously. Such representation is a "just in time" system, which is an inherently dynamic process. Although such architecture presupposes quite efficient attentional mechanisms, it has the advantage that the short life-span of feature binding avoids crosstalk among multiple binding. This simple serial binding architecture may be enough to deal with real life dynamics.

3.2 Strict evaluation of feature binding memory

In the initial studies on MOPT (Saiki, 2002, 2003a, 2003b), switch detection was markedly impaired as motion speed increased. Although performance level is rather high when objects are stationary, it is unclear how much feature bound memory can be maintained. The multidimensional MOPT experiment suggests fairly limited capacity for feature binding memory. Moreover, the results of some studies using a change detection task suggest that our capacity for object representation in visual memory is more limited than previously believed (Alvarez & Cavanagh, 2002; Bahrami, 2003; Olson & Jiang, 2002; Wheeler & Treisman, 2002; Xu, 2002). The literature is currently equivocal regarding the capacity of memory for feature binding. Saiki and Miyatsuji (2007) utilized a new experimental paradigm that appears suited for evaluating binding memory, and analyzed performances using mathematical models analogous to those used in perceptual feature binding studies.

Two necessary conditions must be met to properly evaluate the use of feature conjunctions. First, to eliminate possible contributions from simple feature information, the stimulus set should use identical sets of features in different combinations. Saiki (2002, 2003a, 2003b) and Wheeler and Treisman (2002) satisfied this condition. The second condition is the use of a task able to evaluate the representation of feature combination. One task satisfying this condition is the perceptual identification task used in perceptual feature binding. Change detection tasks used in visual memory obviously fail to satisfy this condition. Because the task is simply detecting a change, representational schemes other than feature combination, such as simple stimulus salience (Itti & Koch, 2000), can account for correct change detection.

Unlike perceptual binding, however, the simple identification task also displays problems. In visual memory, a variety of simple identification tasks tend to underestimate memory capacity, as exemplified by a classic study of the partial report paradigm by Sperling (1969). Moreover, even if subjects are simply asked to report an object with change, cognitive load in response mapping is significant, and is quite likely to affect memory performance. Furthermore, direct identification forces participants to transform visual information into verbal form, which can also compromise visual memory performance.

To avoid problems with both detection and simple identification tasks, a type identification task was devised. The type identification task requires participants to identify which event occurs in the stimulus sequence, as discrimination among four alternatives. Correct identification of change type requires memory for feature combinations. At the same time, unlike simple identification, cost in response mapping is negligible. These characteristics are crucial, particularly when using the wide varieties of colors and shapes seen in most visual memory tasks. If several colors and shapes are used, type identification based solely on salience is almost impossible, and cost in terms of response mapping in single identification becomes prohibitive. Compared with change detection tasks, the type identification task can thus extract important additional information regarding binding memory.

Using type identification with multidimensional MOPT, the role of object motion and number of switching events in binding memory for multiple objects was investigated. A unique property of the MOPT paradigm is the ability to evaluate memory for feature binding in dynamic situations. One of the hallmarks of the objectness is the maintenance of feature binding across spatiotemporal changes (i.e., motion), so the MOPT provides important information about the properties of object representations. Unlike previous studies (Saiki, 2002, 2003a, 2003b), the type identification paradigm could eliminate effects from saliency-based mechanisms, and extract the effect of feature binding more strictly. The second factor to be evaluated was the number of switches. MOPT in previous works involved two switching events in each trial, with the switched state returning to the initial state in the next occlusion period. Compared with the standard change detection task, in which only a single chance exists to detect a change, this specific manipulation may improve subject performance. This factor may be responsible for apparent discrepancies in stationary conditions between the studies by Saiki (2003a, 2003b) and Wheeler and Treisman (2002). Saiki (2003a, 2003b) reported accurate performance under stationary conditions, whereas Wheeler and Treisman (2002) found significant impairment in binding conditions. The present study compared performance in the MOPT task with two switches ("switch-back condition") and with one switch ("no-switch-back condition").

Multidimensional MOPT with the type identification task replicated the basic findings of previous MOPT experiments (Saiki, 2003a,b), in that memory for feature bindings appears severely limited. Overall, patterns of results were consistent with the view that a single switch is insufficient to use memory for binding. Lack of object motion was insufficient for the maintenance of feature bindings, as the no-switch-back condition showed significant impairment even under stationary conditions. In addition to correct type identification rates, analyses of response pattern address an important theoretical issue. Model-based analyses revealed that the number of switches affects not only accuracy, but also contingency of stimulus and response types. Error analyses suggest that people rely more on partial conjunction information (i.e., shape-location, or color-location), at the time of first switch, but at the time of second switch, they rely on triple conjunction information. Combined with the accuracy data, the results of model fitting support the interpretation that shape-color-location binding is available only when a second switch is present.

The results of Experiment 1 suggest that triple conjunction representation becomes available mainly at the time of second switch. There are at least two factors which can produce this result. First, memory representations change their format from partial-conjunction to triple-conjunction between the first and second switches. Second, the first switch functions as a cue to selectively attend to a switching object, which makes the triple-conjunction representation more accessible. To examine the effects of these two factors, Experiment 3 in Saiki and Miyatsuji (2007) introduced two manipulations. First, to investigate the transition from partial- to triple-conjunctions, we used mixed switch trials. Unlike previous experiments, where the event type of the first and second switches was the same, mixed switch trials had two different switch types, allowing us to evaluate which switch leads to observers' type identification response. Second, to investigate effects of selective attention, we introduced a condition where the first and second switches occurred with different object pairs.

When two switches occur with the same pair of stationary objects, there was a strong bias toward reporting the second switch, which is consistent with the hypothesis that triple-conjunction representation becomes available before the second switch by selectively attending to switching objects. In contrast, the significant bias toward the first switch in the different-pair stationary condition is also consistent with attentional cueing hypothesis, because if the attention is focused on the pair of first switch objects, correct type identification of the second switch is less likely than the first switch, when attention was evenly distributed among four objects. Finally, Experiment 2 in Saiki and Miyatsuji (2007) investigated whether the use of occluder significantly impaired performance in MOPT, and showed that the occluder did not have any negative effects on performance.

These results are inconsistent with the popular claim that visual working memory can hold about four objects simultaneously (Cowan, 2001; Irwin, 1992; Kahneman, et al., 1992; Luck & Vogel, 1997) even when objects are stationary. If previous works with change detection tasks reflect the use of explicit memory for feature binding, similarly accurate performances would be expected in type identification tasks. One exception in previous studies using a change detection task is Wheeler and Treisman (2002), which showed significant impairment in the change detection of feature bindings. The findings of the present study appear consistent with their data, but the mechanisms underlying performance impairment may differ. As Wheeler and Treisman used a change detection task, the saliency-based detection strategy is available. Impairment as described by Wheeler and Treisman may thus reflect a reduction to salience change in the binding condition. In MOPT with type identification, on the other hand, saliency-based identification is almost impossible, and impairment in the no-switch-back stationary condition likely reflects the limit in feature binding. This issue is discussed in the next section.

Both accuracy data and event-response contingency analyses revealed that properties of binding memory are qualitatively different between conditions involving only one switch and those presenting a second chance. One interpretation is that visual memory, similar to visual perception (Treisman & Schmidt, 1982), is structured in a feature-based fashion when multiple objects require simultaneous storage. When attention is directed to an object, feature representations are integrated to form a coherent object representation (Treisman, 1988). In other words, availability of selective attention to a particular object could result in significant changes to performance. In the no-switch-back condition with a single switch, subjects must divide their attention between all four objects, since the subject does not know which object will change. In that state, the results suggest that only partial feature binding information is available. When the first switch occurred, if the objects were stationary, subjects were likely to detect a change to one or two objects, but were unable to identify the type. Subjects then direct attention to a suspected object, and if a second switch occurred, they could identify the switch type based on selective attention. An extreme view of this account is that we can hold feature-bound object only one at a time. One remaining issue is whether selective attention affects transition from feature-based memory to object-based memory, or modulates the availability of prestored object representation.

To evaluate visual working memory for feature binding, Saiki and Miyatsuji (2007) devised a type identification paradigm, and applied it to multiple object permanence tracking task (MOPT). Compared with previous results with simple change detection, task performance was greatly reduced, suggesting that previous data reflects memory for something other than feature binding, such as stimulus salience. The number of switches facilitates performance only when objects were stationary, and the model-based analyses and mixed design experiment showed that this improvement reflects the effects of selective attention on forming or strengthening feature-bound memory representation. In contrast, when objects were moving, the effect of second switch was quite small, suggesting that either detection of the first switch, or maintenance of feature binding across occlusion is disrupted by object motion. Type identification method is a powerful tool to investigate various aspects of feature binding memory in combination with model-based analyses and various experimental procedures.

3.3 Source of performance impairment: maintenance or retrieval?

A series of experiments using MOPT (Saiki, 2002, 2003a, 2003b; Saiki & Miyatsuji, 2007) so far revealed that task performance was severely impaired even when objects are stationary. One critical problem is to evaluate whether the performance impairment reflects memory retrieval or maintenance. Saiki and Miyatsuji (in press) addressed this issue.

A deficit in a memory task may be caused by a limit in storage capacity, or by a bottleneck in memory retrieval and/or comparison between memory and perceptual representations. Some studies have suggested that low estimated capacity for feature binding memory relative to feature memory may reflect differences in memory retrieval. Wheeler and Treisman (2002) compared the single-probe paradigm, where only one object was presented in the probe display to be judged for the presence of change, with the multiple-probe paradigm, where the whole probe display needed to be compared with the initial display. They showed that the single-probe condition significantly improved performance in the binding condition compared to the multiple-probe condition. This improvement in task performance can be interpreted as a reduction of interference and/or a facilitation of memory retrieval by the single probe.

The single probe advantage in the binding condition leaves some questions open regarding the nature of representation and processing in VSTM. First, the findings of Wheeler and Treisman (2002) do not necessarily imply that memory for object files in general suffer from a retrieval bottleneck in the multiple probe condition. Wheeler and Treisman investigated simple feature conjunctions such as color-location and shape-color conjunction, so whether a single probe advantage is observed with more complex representations (triple conjunction) remains unknown. If the previous findings reflect the nature of object files in general, the single probe advantage should be observed with triple conjunction representation. Conversely, if the previous findings hold true only in certain special situations, the single probe advantage may be limited to simple conjunction representations. In Saiki and Miyatsuji (in press), retrieval cueing and memory task manipulation were combined to achieve a better understanding of the nature of binding memory.

Experiment 1 combined the type identification task and retrieval cuing using the MOPT paradigm. A cue indicating the changing object was 100% valid, and was presented either just before (precue) or after (postcue) a change occurred. If a cue is effective, the precue condition is expected to show significantly better task performance compared with the no-cue control. The critical condition was the postcue condition. Estimated capacity from behavioral data is determined by two factors: maintenance capacity and costs in memory retrieval. Because the postcue condition substantially reduces retrieval costs, estimated capacity in the postcue condition is closer to the genuine maintenance capacity than in the no-cue condition. The performance facilitation by the postcue condition thus suggests that estimated capacity in the no-cue condition suffers from retrieval costs, while the lack thereof suggests that the estimated capacity in the no-cue condition reflects genuine maintenance capacity. Also, using the moving condition of MOPT, we compared effects of retrieval cueing between spatiotemporal updating and simple maintenance of complete object files. Experiment 1 failed to obtain facilitation by the retrieval cue, suggesting that the retrieval cue benefit does not occur for triple conjunctions.

One alternative account, however, is that the complexity of the type-identification paradigm eliminated any postcue benefit. In Experiment 2, a relevant-feature switch detection task (see section 2.5) was used. This task had only two response alternatives, as in a simple change-detection task, but required distinction between color- and shape-switch events. A relevant-feature switch detection task failed to show any effect of postcue, suggesting that the results in Experiment 1 were not simply due to the complexity of response mapping. Next, to eliminate a possibility that postcue manipulation is simply not effective in MOPT task, Experiment 3 used a simple change detection task. A simple change-detection task revealed significant facilitation in the stationary condition. A retrieval cue facilitates judgment of whether any kind of change is present, but does not help identify the type of switch. The postcue paradigm can thus reveal a facilitation effect similar to that found with the single-probe paradigm of Wheeler and Treisman (2002), suggesting that postcues used in this study can effectively function as a retrieval cue. Another interesting result was the lack of postcue effects in the moving condition, suggesting that the postcue is ineffective for moving objects. This may reflect that memory retrieval and matching operation are location-based, not object-based. These results are replicated in Experiment 4 where a simple switch detection and relevant-feature switch detection tasks were directly compared with a within-subject design. Finally, Experiment 5 revealed that these findings are not reflecting overwriting effects.

Taken together, the interaction between postcue benefit and task (significant benefit with the simple change-detection and no benefit with tasks requiring triple conjunctions) suggest that retrieval cue benefit occurs only for simple feature conjunctions, and that limits in triple conjunctions primarily reflect memory maintenance. Maintenance capacity for triple conjunctions is close to the estimated capacity, that is, one or two objects, whereas, that for simple conjunctions may be larger.

The present results argue against the view that memory of feature binding is a system composed of general object file representations. General object file representations include complex representations such as triple conjunctions, and should lead to postcue benefits in all different tasks used in this work, a possibility was unsupported by the data. Unlike a previous claim by Luck and Vogel (1997) that the content of object memory, object files, is complete, regardless of the number of features, the present study suggests that the content of object files are partial by default. The present study suggests that functional properties of object files differ depending on complexity, which is related to a recent argument regarding whether complexity of objects affects the capacity of VSTM (Alvarez & Cavanagh, 2004; Awh et al., 2007).

Alvarez and Cavanagh (2004) reported that capacity estimate using a simple change-detection task is a linear function of the complexity of the object measured by the slope in a visual search task, suggesting that the complexity of objects affects the capacity of VSTM. Recently, however, Awh et al. (2007) showed some evidence that these results could be explained by difficulty of matching between memory and percept, suggesting that the capacity of VSTM is fixed regardless of object complexity, but resolution of object representations becomes degraded with increasing complexity. As far as the simple change-detection task is concerned, the results of the present study appear consistent with the argument by Awh et al. as the significant postcue benefit with simple change detection suggests that performance impairment primarily reflects memory retrieval or matching of memory and percept, and not capacity per se. In contrast, the results with tasks requiring triple conjunctions seem consistent with the argument of Alvarez and Cavanagh (2004), suggesting that impairment primarily reflects maintenance capacity. When the task requires use of triple conjunctions, the capacity of object file representation is substantially reduced.

Taken together, the idea of fixed capacity with varied resolution may hold only in the context of simple change detection, and in general, the complexity of objects may reduce the maintenance capacity of memory representation.

The effects of retrieval cue on visual short-term memory depend on task requirement. Whereas a simple change-detection task shows a facilitatory effect as seen in previous studies, tasks requiring discrimination of different feature combinations failed to show facilitation, even when task difficulty was similar to the change detection. These results suggest that retrieval cue benefit occurs in memory for simple feature conjunctions, but not for more complex representations. Limits in memory for complex object files primarily reflect maintenance capacity, whereas maintenance capacity for simple conjunctions is underestimated by a simple change-detection task due to retrieval bottleneck.

3.4 Comparison between arbitrary and knowledge-based binding

So far, all experiments investigating feature binding in visual working memory using MOPT used arbitrary color-shape combinations. However, binding of various features of natural objects are usually not arbitrary. For example, banana has often a particular shape and a yellow color. Binding in natural objects is structural in the sense that a particular combination of component features is associated with a higher level description of objects, whereas binding discussed in visual working memory lacks such a higher level unit. In other words, a problem with stimuli used in visual working memory may be the lack of such structural relations. To address this issue, an experiment was conducted to compare the effect of higher level nodes on maintenance of feature binding in visual working memory. Two specific questions are addressed:

(1) Does pre-stored knowledge about shape-color correspondence facilitate memory for feature bindings, and if so, how?

(2) Does constant mappings of shape-color correspondence within an experimental session facilitate memory for feature bindings, and if so, how?

If manipulations of (1) or (2) facilitate performance, the limited capacity for feature bindings in previous works is likely to reflect the arbitrary and independent nature of feature conjunctions used in the experiments. In contrast, if the factors above do not facilitate performance, then the capacity limit is likely to be more general.

Using the multidimensional MOPT paradigm with the type identification procedure, the roles of prestored memory representations of color-shape conjunctions in maintaining object information in visual working memory were evaluated. An experiment was conducted to investigate (1) whether known color-shape conjunctions facilitate maintenance of multiple object representations in visual working memory, (2) whether fixed color-shape conjunction facilitates maintenance of multiple object representations, and (3) whether patterns of errors demonstrate the roles of prestored conjunctions in visual working memory.

Saiki (2007) investigated whether this limitation is specific to the use of arbitrary combinations of color-shape. Two main independent variables were object type and motion type. The object types were natural when natural objects were used, geometric-constant when geometric figures were used as in previous studies, while the shape-color correspondences were fixed, and geometric-varied, which is identical to previous studies. The motion types were object motion and occluder motion. Shapes used for objects in the geometric conditions were circle, square, hexagon and triangle. Objects used in the natural condition were lobster, frog, banana, and violin, which had clear associated colors, based on a preliminary survey. Colors were those typically associated colors: red, green, yellow, and brown, for both natural and geometric conditions. A total of four events were possible: object-switch with simultaneous switch of color and shape; color-switch alone; shape-switch alone; and no switch. Participants were asked to identify event types without feedback as to which was correct.

The results showed only a weak tendency toward performance improvement in the natural and geometric-constant conditions, and these conditions showed severe performance impairment under the moving condition. The natural and geometric-constant conditions were virtually the same in accuracy, suggesting that prestored color-shape conjunctions had limited effect on percent correct data.

However, analyses of error types demonstrated strong effects of prestored conjunction on task performance. Compared with geometric conditions, the natural condition showed significantly more errors confusing between color-switch and shape-switch, suggesting that observers were quite sensitive to detect a change in object identity, but not able to accurately identify the switch type. In the natural condition, color and shape form a unit of object identity, but to identify the switch type, its component (either color or shape) and location needs to be bound. Observers can detect the occurrence of color or shape switch when they see a green lobster, but they are not good at telling whether a red lobster changed to a green lobster (i.e., color switch), or a green frog changed to a green lobster (i.e., shape switch). In fact, they had a strong bias to judge any switch involving identity change as a color switch.

In contrast, although the error rates were about the same under the geometric-constant condition, the pattern of errors is quite different. Color and shape behave more independently, even when the conjunctions are completely fixed. Unlike the case of lobster, when a predefined red-square combination changed to a red-circle (i.e., shape-switch), errors were more likely to be an indication of no switch (i.e., overlooking the shape-switch), and in the case of feature confusion, errors occurred evenly in both directions.

Results for the natural condition support a view that visual features are first bound together to form a type representation, before further binding to a spatiotemporal location to form a token (Kanwisher, 1991). Moreover, this view holds only when type information is prestored in LTM, and without prestored types, shape-color conjunctions played no significant role.

More importantly, the availability of type information did not facilitate task performance in MOPT. Binding of type representations to their spatiotemporal location appears to be quite difficult. This raises a possibility that even the feature binding in structural descriptions may have a similar limitation. As Hummel and Biederman (1992) described, structural description is not simply a co-activation of a set of geons, but also a binding of parts with relations. Given part representation is a set of its components, it is similar to the type representation discussed here. Thus, structural description needs binding of parts (types) with spatial information, which corresponds to the binding of types with their locations in MOPT task. Thus, the formal structure has a certain level of similarity between multiple objects in the MOPT task and an object's structural description.

However, there are important differences as well. For example, parts are tightly grouped by connectedness and other grouping factors (Saiki & Hummel, 1998), but objects are completely separated in MOPT. Binding in structural description formation is limited to shape information, but shape and color (and other object features) are used in MOPT. Clearly, how these factors affect binding performance is an issue for further studies, but

Saiki (2007) shows that limits in feature binding in visual working memory are not simply an artifact of arbitrary feature combinations, and these limits may have a broader common ground including binding in object recognition.

3.5 Neural correlate of feature binding in VSTM

Although the MOPT task clarified the cognitive aspects of object representation in visual working memory, the underlying neural mechanisms remain unclear. Several neuroimaging studies have addressed either dynamic updating (Culham et al., 1998; Culham et al., 2001; Jovicich et al., 2001) or feature binding (Prabhakaran et al., 2000; Mitchell et al., 2000; Shafritz et al., 2002), but none have addressed both simultaneously. In this respect, the MOPT task provides a unique means of investigating the neural basis behind the interactions of feature binding and dynamic updating, which are a crucial part of our visual object representation.

Previous studies on dynamic updating have reported quite consistent results showing activation of the dorsal frontoparietal network involving the frontal and the parietal areas (Culham et al., 1998; Culham et al., 2001; Jovicich et al., 2001). In contrast, results from feature binding studies are less consistent, showing participation of the anterior prefrontal (Prabhakaran et al., 2000; Mitchell et al., 2000) and parietal areas (Shafritz et al., 2002; Corbetta et al., 1995; Friedman-Hill et al., 1995; Ashbridge et al., 1997), and the hippocampus (Mitchell et al., 2000). One possible reason for such discrepancies is the diversity of experimental paradigms. Importantly, most of these paradigms do not reflect feature binding in a strict sense. To investigate feature binding, standard and test stimuli should contain an identical set of features, only differing in the combination, and the task should require the use of combination information. Thus far, no clear demonstrations of neural correlates for feature bindings in memory have been presented, and exploring these with the MOPT paradigm is important.

We performed a functional magnetic resonance imaging (fMRI) experiment during the MOPT task. Sixteen observers performed two kinds of MOPT task that enabled us to compare brain activities during dynamic (object-moving) and static (object-stationary) situations under matched circumstances.

We found that the MOPT task induced not only dorsal frontoparietal activation, but also right anterior and bilateral ventral parts of frontal activation. The spatial pattern of this activation did not vary, irrespective of whether objects were in motion or stationary. This result was clearly inconsistent with a hypothesis that the MOPT task would induce only dorsal frontoparietal activation. Thus, the dorsal frontoparietal network alone cannot maintain object representations. In the ROI analyses, patterns of anterior and inferior parts of frontal activation differed from those of the dorsal frontoparietal activations. This discrepancy between two activation groups suggests that these two activation groups reflect two distinct cognitive processes. Object representation in visual working memory thus appears to be maintained by active interactions between the dorsal frontoparietal network and other frontal regions.

The present study revealed that the MOPT task induces both dorsal frontoparietal and the anterior and ventral activation of the frontal cortices. This result suggests that object representations are not contained within a single neural system such as the frontal area or dorsal frontoparietal network, but instead are represented by cooperation of distributed neural systems involving the coherency control system in the anterior frontal area and the dynamic updating system in the dorsal frontoparietal network.

However, the epoch-related design in Imaruoka et al. (2005) prevents us from further analysis, given that activities in the maintenance and change-detection periods were confounded. Recently, Takahama et al. (2005) conducted a follow-up to the Imaruoka study, using an event-related design and modifying the experimental paradigm in several points. First, with extensive practice trials before the fMRI sessions, the accuracy of behaviour data was set to be quite high, and there was no substantial difference in task difficulty across conditions. Second, visual stimuli in the maintenance period of control conditions now were exactly the same as those in the experimental conditions, so that differences in brain activity could be said to reflect top-down control of memory maintenance. Third, activities in the maintenance and change-detection periods were now decomposed by event-related design. Although still preliminary, the results were largely consistent with those of Imaruoka et al., but with some new findings. Regarding activity in posterior areas, maintenance activity showed the pattern similar to that of Imaruoka et al. (2005). By contrast, the event-related design revealed further qualifications about anterior prefrontal activity. The effect of load (moving vs. stationary) was observed during the maintenance period such that the moving condition showed stronger activation in both control and test conditions, without task effect. The effect of task (binding vs. control conditions) was instead observed during the change-detection. These data suggest that manipulation of memory representation during the maintenance period increases anterior prefrontal activity, whereas binding of color and location affects the memory retrieval and matching process. Manipulation-related activity in the anterior prefrontal area is consistent with Mohr et al. (2006), and binding-related activity at the time of change-detection appears to imply that the anterior PFC is not the storage place of feature binding, but, rather, involved in carrying out judgments based on a change in feature binding. Because all the reports of binding-related activity in the anterior PFC used epoch-related design (Imaruoka et al., 2005; Mitchell et al., 2000; Prabhakaran et al. 2000), this interpretation is consistent with those previous studies.

Taken together, studies using the MOPT paradigm support the view that maintenance and updating of feature-bound object representations cannot be carried out autonomously within the frontoparietal network. Updating of color-location binding requires activity of the anterior PFC, suggesting that the conventional MOT task is unlikely to be actually investigating the tracking of feature-bound object representations. Although memory for color-location binding cannot fully function within the frontoparietal network, there are some alternative explanations regarding the functional architecture of binding memory. First, as suggested by Wheeler and Treisman (2002), visual working memory is inherently feature-based, and memory judgment on feature conjunction, such as switch detection, is carried out by combining states of two feature-based memory systems. Alternatively, memory representations in the inferior IPS may be feature-bound, but the frontoparietal network cannot detect a change in feature combination autonomously. In other words, representations in the inferior IPS may be implicitly feature-bound, but explicit detection of change requires prefrontal activity. Recently, some studies reported that intraparietal sulcus (IPS) revealed brain activation proportional to memory load (Todd & Marois, 2004; Vogel & Machizawa, 2004; Song & Jiang, 2006; Xu & Chun, 2006). Earlier studies use simple color-location tasks (Todd & Marois, 2004; Vogel & Machizawa, 2004) and more recent ones manipulated object complexity (Xu & Chun, 2006) and the number of objects and features (Song & Jiang, 2006), showing that IPS activation is modulated by both object complexity (in particular superior IPS) and the number of features. However, whether explicit representation of complex feature conjunctions resides in IPS remains unclear. Further investigations are necessary to resolve this issue.

4. Summary and future directions

I have reviewed a series of studies using MOPT task in this chapter. In this section, I summarize the major findings and relate them in a broader context of object representations in visual cognition.

The findings from MOPT experiments can be summarized in the following way in terms of transformation, maintenance and retrieval of object representations.

(1) Transformation of object representations: Even when objects are moving slowly and in completely predictable fashion, updating object features (colors and shapes) as objects' movement is extremely costly. Extreme difficulty in feature switch detection in the moving condition reflects failure in updating feature binding, not in tracking or feature encoding alone.

(2) Maintenance of object representations: The results from the type identification and relevant-feature switch detection tasks revealed that maintenance of multiple object representations without any spatiotemporal transformation is also quite difficult. Unlike often reported capacity estimate of 3-5 objects using simple change detection tasks, the estimated capacity for complex feature conjunction representations is somewhere between 1 and 2 objects.

(3) Retrieval of object representations: Whether impairment in task performance reflects failure in memory retrieval depends on tasks. Memory evaluated by a simple change detection task suffers significant effects of retrieval bottleneck, whereas that evaluated by tasks measuring triple conjunction representations does not. This suggests that capacity estimates of 1-2 objects for complex object representations primarily reflect maintenance capacity, whereas estimates of 3-5 objects for simple conjunctions are underestimated by retrieval failures.

(4) Object representations with previous experiences: MOPT with natural objects revealed that difficulty in integrating feature and location information is not specific to arbitrary combination of features. However, the structure of memory representation is different: arbitrary objects form a set of feature-location bindings (color-location and shape-location), whereas natural objects form a type (shape-color conjunction) bound to location.

(5) Neural correlate of transformation of object representation: Although still preliminary, fMRI experiments suggest that a large network of brain areas, in particular frontoparietal network and anterior prefrontal cortex, is necessary to maintain and transform multiple object representations.

These findings suggest that although subjectively we feel that we can see multiple complex objects simultaneously, and can maintain their memory for a short period of time, it may not be the case. A strict test of object memory using MOPT paradigm shows that our capacity of visual working memory for complex multi-feature objects is surprisingly small. These findings suggest that previous findings regarding capacity of 3-5 objects may reflect partial representations of complex objects, such as just color-location or shape-location. When detailed representation of complex object is necessary, one may need to direct selective attention to the object. The reason why we do not have any serious problems in everyday life may be that selective attention mechanism is quite efficient to direct attention to an object or region which needs detailed processing just in time.

If our visual cognition system operates with minimum number of complex object representations, this becomes an important constraint in human-interface design. A situation like MOPT, color switch in the middle of object motion, rarely occurs in natural environments, but it becomes certainly possible in artificial environments, particularly computer-generated virtual environments with much less physical constraints. Thus, in dynamic control of complex systems such as driving vehicles and air-traffic control, careless design of interface may cause an accident or other serious consequences.

Are findings with MOPT paradigm specific to this paradigm? The severe limitation with complex objects is also reported studies with typical change detection tasks (for example, Xu & Chun, 2006), which is consistent with the MOPT studies. Recently, using a different experimental paradigm called spatiotemporal search, I found the results consistent with these studies (Saiki, in press).

One issue remaining unclear is the role of implicit mechanisms. All experiments with MOPT so far used an explicit task, thus even if we cannot maintain many complex object representations explicitly, they may be maintained in some implicit fashion. Indeed, object file preview effects (Kahneman et al., 1992) suggest that it may be the case. Implicit representation of feature binding and its role in visual cognition is an important future direction, which needs new ways of investigating the issue. Modification of MOPT may contribute to this line of research as well.

5. Conclusion

Multiple object permanence tracking (MOPT) task revealed that our ability of maintaining and transforming multiple representations of complex feature-bound objects is limited to handle only 1-2 objects. Often reported capacity of 3-5 objects likely reflects memory for partial representations of objects and simple cases such as just color and their locations. Also, performance in multiple object tracking (MOT) task is likely mediated by spatiotemporal indices, not by feature-bound object representations. MOPT paradigm is quite useful in investigating maintenance, retrieval and transformation of dynamic object representations with properly controlled experimental setting.

6. Acknowledgement

This work was partially supported from Grants-in-Aid (#13610084, #14019053, and #19500226) from JMEXT, the Global COE Program "Revitalizing Education for Dynamic Hearts and Minds," from JMEXT, and PRESTO from JST. I would like to think collaborators of the MOPT project, Hirofumi Miyatsuji, Toshihide Imaruoka, and Sachiko Takahama.

7. References

Alvarez, G. A. & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15, 106-111, ISSN: 0956-7976.

Ashbridge, E.; Walsh, V. & Cowey, A. (1997). Temporal aspects of visual search studied by transcranial magnetic stimulation. *Neuropsychologia*, 35, 1121-1131, ISSN: 0028-3932.

Awh, E.; Barton, B. & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, 18, 622-628, ISSN: 0956-7976.

Ballard, D. H.; Hayhoe, M. M.; Pook, P. K. & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences,* 20, 723-767, ISSN: 0140-525X.

Behrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition*, 10, 949-963, ISSN: 1350-6285.

Chun, M. M. & Cavanagh, P. (1997). Seeing two as one: linking apparent motion and repetition blindness. *Psychological Science*, 8, 74-79, ISSN: 0956-7976.

Corbetta, M.; Shulman, G. L.; Miezin, F. M. & Petersen, S. E. (1995). Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science,* 270, 802-805, ISSN 0036-8075.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences,* 24, 87-185, ISSN: 0140-525X.

Culham, J. C.; Brandt, S. A.; Cavanagh, P.; Kanwisher, N. G.; Dale, A. M. & Tootell, R. B. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology,* 80, 2657-70, ISSN: 1522-1598.

Culham, J. C.; Cavanagh, P. & Kanwisher, N. G. (2001). Attention response functions: characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron,* 32, 737-745, ISSN: 0896-6273.

Friedman-Hill, S. R.; Robertson, L. C. & Treisman, A. (1995). Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. *Science,* 269, 853-855, ISSN 0036-8075.

Horowitz, T. S. & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, 394, 575-577, ISSN: 0028-0836.

Hummel, J. E. & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review,* 99, 480-517, ISSN: 0033-295X.

Imaruoka T.; Saiki J. & Miyauchi S. (2005). Maintaining coherence of dynamic objects requires coordination of neural systems extended from anterior frontal to posterior parietal brain cortices. *NeuroImage* 26, 277-284, ISSN: 1053-8119.

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, 23, 420-456, ISSN: 0010-0285.

Itti, L. & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489-1506, ISSN: 0042-6989.

Jovicich, J.; Peters, R. J.; Koch, C.; Braun, J.; Chang, L. & Ernst, T. (2001). Brain areas specific for attentional load in a motion-tracking task. *Journal of Cognitive Neuroscience,* 13, 1048-1058, ISSN: 0898-929X.

Kahneman, D.; Treisman, A. & Gibbs, B. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology, 24*, 175-219, ISSN: 0010-0285.

Kanwisher, N. G. (1991). Repetition blindness and illusory conjunction: Errors in binding visual types with visual tokens. *Journal of Experimental Psychology: Human Perception and Performance,* 17, 404-421, ISSN: 0096-1523.

Luck, S. J. & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281, ISSN: 0028-0836.

Mitchell, K. J.; Johnson, M. K.; Raye, C. L. & D'Esposito, M. (2000). fMRI evidence of age-related hippocampal dysfunction in feature binding in working memory. *Cognitive Brain Research,* 10, 197-206, ISSN: 0926-6410.

Mohr H. M.; Goebel R. & Linden D. E. J. (2006). Content- and task-specific dissociations of frontal activity during maintenance and manipulation in visual working memory. *Journal of Neuroscience*, 26, 4465-4471, ISSN: 1529-2401.

Olson, I. R. & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception & Psychophysics*, 64, 1055-1067, ISSN: 0031-5117.

Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics,* 44, 369-378, ISSN: 0031-5117.

Prabhakaran, V.; Narayanan, K.; Zhao, Z. & Gabrieli, J. D. (2000). Integration of diverse information in working memory within the frontal lobe. *Nature Neuroscience,* 3, 85-90, ISSN: 1097-6256.

Pylyshyn, Z. Q. & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision,* 3, 179-197, ISSN: 0169-1015.

Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision Research,* 40, 1469-1487, ISSN: 0042-6989.

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, ISSN: 0956-7976.

Saiki, J. (2002). Multiple-object permanence tracking: limitation in maintenance and transformation of perceptual objects. In J. Hyona, D. P. Munoz, W. Heide and R. Radach (Eds.), *The Brain's eye: neurobiological and clinical aspects of oculomotor research (Progress in Brain Research Vol. 140) (pp.133-148)*. Elsevier Science, ISBN: 0-444-51097-4, Amsterdam.

Saiki, J. (2003a). Feature binding in object-file representations of multiple moving items. *Journal of Vision,* 3, 6-21, ISSN: 1534-7362.

Saiki, J. (2003b). Spatiotemporal characteristics of dynamic feature binding in visual working memory. *Vision Research,* 43, 2107-2123, ISSN: 0042-6989.

Saiki, J. (2007). Feature binding in visual working memory. In N. Osaka, I. Rentschler, & I. Biederman (Eds.) *Object Recognition, Attention & Action (pp. 173-185).* Springer - Verlag, ISBN 978-4-431-73018-7, Tokyo.

Saiki, J. & Hummel, J. E. (1998). Connectedness and the part-relation integration in shape perception. *Journal of Experimental Psychology: Human Perception and Performance,* 24, 227-251, ISSN: 0096-1523.

Saiki, J. & Miyatsuji, H. (2007). Feature binding in visual working memory evaluated by type identification paradigm. *Cognition. 102,* 49-83, ISSN: 0010-0277.

Saiki, J. & Miyatsuji, H. (in press). Estimated capacity of object files in visual short-term memory is not improved by retrieval cueing. *Journal of Vision,* ISSN: 1534-7362.

Saiki, J. (in press). Functional roles of memory for feature-location binding in event perception: Investigation with spatiotemporal visual search. *Visual Cognition,* ISSN: 1350-6285.

Scholl, B. J. & Pylyshyn, Z. W., (1998). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, 38, 259-290, ISSN: 0010-0285.

Shafritz, K. M.; Gore, J. C. & Marois, R. (2002). The role of the parietal cortex in visual feature binding. *Proceedings of the National Academy of Sciences, U.S.A., 99*, 10917-10922, ISSN: 1091-6490.

Song, J. H. & Jiang, Y. (2006). Visual working memory for simple and complex features: an fMRI study. *NeuroImage*, 30, 963-972, ISSN: 1053-8119.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs,* 74(11) [whole No. 498]), 29.

Takahama, S.; Saiki, J.; Misaki, M. & Miyauchi, S. (2005). The necessity of feature-location binding activates specific brain regions in visual working memory task: an event-related fMRI study. *Society for Neuroscience Annual Meeting Abstract.*

Tipper, S. P.; Weaver, B.; Jerreat, L. M. & Burak, A. L. (1994). Object-based and environment-based inhibition of return of visual selection. *Journal of Experimental Psychology: Human Perception and Performance,* 20, 478-499, ISSN: 0096-1523.

Todd, J. J. & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature,* 428, 751-754, ISSN: 0028-0836.

Towse, J. N.; Hitch, G. J. & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory and Cognition,* 28, 341-348, ISSN: 0090-502X.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology,* 40A, 201-237, ISSN: 1747-0218.

Treisman, A. (1999). Solutions to the binding problem: Progress through controversy and convergence. *Neuron,* 24, 105-110, ISSN: 0896-6273.

Treisman, A. & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology,* 14, 107-141, ISSN: 0010-0285.

Valdes-Sosa, M.; Cobo, A. & Pinilla, T. (1998). Transparent motion and object-based attention. *Cognition*, 66, B13-B23, ISSN: 0010-0277.

Verstraten, F. A. J.; Cavanagh, P. & Labianca, A. T. (2000). Limits of attentive tracking reveal temporal properties of attention. *Vision Research*, 40, 3651-3664, ISSN: 0042-6989.

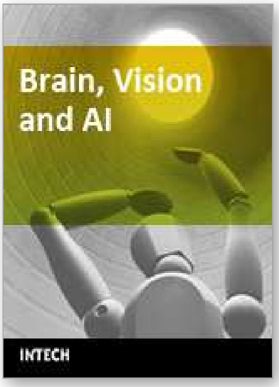
Vogel, E. K. & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428, 748-751, ISSN: 0028-0836.

Vogel, E. K.; Woodman, G. F. & Luck, S. J. (2001). Storage of features, conjunctions and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 92-114, ISSN: 0096-1523.

Wheeler, M. E. & Treisman, A.M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131*, 48-64, ISSN: 0096-3445.

Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 458-468, ISSN: 0096-1523.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440*, 91-95, ISSN: 0028-0836.



Brain, Vision and AI
Edited by Cesare Rossi

ISBN 978-953-7619-04-6
Hard cover, 284 pages
Publisher InTech
Published online 01, August, 2008
Published in print edition August, 2008