



Research article

Neural correlates of visual short-term memory for objects with material categories

Sachio Otsuka^{a,*}, Jun Saiki^b^a Faculty of Culture and Information Science, Doshisha University, Japan^b Graduate School of Human and Environmental Studies, Kyoto University, Japan

ARTICLE INFO

Keywords:

Neuroscience
Cognitive neuroscience
Cognition
Cognitive psychology
Learning and memory
Fusiform face area
Tactile properties
Visual short-term memory
Material information
Precentral/postcentral gyrus

ABSTRACT

Behavioral and neuroscience studies have shown that we can easily identify material categories, such as metal and fabric. Not only the early visual areas but also higher-order visual areas including the fusiform gyrus are known to be engaged in material perception. However, the brain mechanisms underlying visual short-term memory (VSTM) for material categories are unknown. To address this issue, we examined the neural correlates of VSTM for objects with material categories using a change detection task. In each trial, participants viewed a sample display containing two, four, or six objects having six material categories and were required to remember the locations and types of objects. After a brief delay, participants were asked to detect an object change based on the images or material categories in the test display (image-based and material-based conditions). Neuronal activity in the brain was assessed using functional magnetic resonance imaging (MRI). Behavioral results showed that the number of objects encoded did not increase as a function of set size in either image-based or material-based conditions. By contrast, MRI data showed a difference between the image-based and material-based conditions in percent signal change observed in a priori region of interest, the fusiform face area (FFA). Thus, we failed to achieve our research aim. However, the brain activation in the FFA correlated with the activation in the precentral/postcentral gyrus, which is related to haptic processing. Our findings indicate that the FFA may be involved in VSTM for objects with material categories in terms of the difference between images and material categories and that this memory may be mediated by the tactile properties of objects.

1. Introduction

The visual perception of materials is based on the estimation and categorization of a wide range of properties inherent to an object (Fleming, 2014). For example, it is important for us to estimate an object's state and make decisions based on material perception (Komatsu and Goda, 2018; Fleming, 2014; e.g., decide whether or not a strawberry is fresh and can be eaten). Previous studies of material perception have focused on glossiness, translucency, and surface roughness (e.g., Nishida and Shinya, 1998; Fleming et al., 2003). It has been demonstrated that both the lower and higher visual areas of humans and nonhuman primates are engaged in material perception, such as the V4, posterior inferior cortex, and fusiform gyrus (e.g., Cant and Goodale, 2011; Hiramatsu et al., 2011; Goda et al., 2014). However, it is unclear how we store information related to material categories in the visual memory. Here, we examined the neuronal mechanisms associated with the visual short-term memory (VSTM) for objects with material categories.

Several previous studies of material properties have focused on the visual estimation of specific properties such as glossiness and surface roughness (see Fleming, 2014). For example, Nishida and Shinya (1998) reported that participants could judge the specular reflectance of a glossy surface simulated by a computer. Motoyoshi and Matoba (2012) indicated that the statistical characteristics of the illumination influenced the perceived glossiness. Additionally, Ho et al. (2006) reported that the judgment of surface roughness is systematically biased by the illumination. By contrast, material categorization refers to the ability to judge which class a material belongs to, and this helps us to guess its properties (Fleming et al., 2013). Thus, material categorization results in a simple label, such as metal, glass, and fabric, and this may not require a precise representation of the material properties. Psychological research has shown that we can rapidly perceive and categorize the material for objects made of plastic and fabric (Sharan et al., 2009). For example, in the study by Fleming et al. (2013), participants viewed images of different material categories and rated subjective qualities, such as hardness,

* Corresponding author.

E-mail address: sotsuka@mail.doshisha.ac.jp (S. Otsuka).

glossiness, and prettiness. Even when they did not know that the samples belonged to different classes, the subjective ratings of the samples were systematically classified into material categories. These results suggest that we can group materials according to the visual appearance of the material properties. Additionally, Cant and Goodale (2007), Cant et al. (2009), and Cavina-Pratesi et al. (2010a, b) have reported that brain areas around the collateral sulcus and the fusiform gyrus are involved in the processing of individual features of surface properties (e.g., color or texture). Cavina-Pratesi et al. (2010b) have indicated that the medial regions of the occipitotemporal cortex are engaged in the estimation of surface materials.

How do we identify and categorize the material of an object from these properties? Recently, using functional magnetic resonance imaging (fMRI), Hiramatsu et al. (2011) examined how information about materials (e.g., low-level image features such as spatial frequency) was encoded and transformed into perceptual representations in the brain when participants viewed the images of nine materials such as metal, ceramic, and glass. As a result, the neural representations gradually shifted from image-based representation in early visual areas (V1/V2) to perceptual-category representations in higher-order visual areas around the fusiform gyrus. Similar results have also been observed in the visual cortex of macaque monkeys (Goda et al., 2014). Additionally, Suzuki (2015) reported that a patient with damage to the left ventromedial occipitotemporal cortex, which included the fusiform gyrus and lingual gyrus, experienced difficulty matching and naming the textures of real materials. These results indicate that material categorization requires more than the low-level processing of material images, and the fusiform gyrus and its surrounding regions play an important role in the perception of material categories.

This study aimed to clarify the neuronal mechanisms in the brain associated with VSTM for material information using event-related functional MRI. As described above, previous studies have identified the brain areas engaged in perceiving the low-level features of material properties and forming material categories. However, it is still unclear whether there are neural representations of VSTM for materials and, if so, which region is involved in the maintenance of material information. To address this issue, we used a change detection task (e.g., Todd and Marois, 2004; Xu and Chun, 2006), in which participants are required to remember colors, shapes, or relatively complex features from the same set of objects. They first view one to four or six (or eight) objects in a sample display. As items in the display increase, it is presumed that visual working memory load also increases, which requires more of the store's capacity (Ma et al., 2014). After a brief delay, participants are required to detect a shape-feature change in a test display. Previous studies have demonstrated that people can retain memories pertaining to three to four simple features of objects or colors (Hakim et al., 2019; Luck and Vogel, 1997; Todd and Marois, 2004; Xu and Chun, 2006). By contrast, the number of complex objects encoded does not increase with set size and reaches a plateau of about 1.5 at set size two (e.g., Alvarez and Cavanagh, 2004; Brady and Alvarez, 2015; see also Awh et al., 2007).

According to this slot model (e.g., Cowan, 2001; Luck and Vogel, 1997), individuals store each item in independent memory slots. Instead of the slot model, a resource model proposed that people allocate their resources to distinct items, and this item can be recalled when enough resources are allocated to it (e.g., Bays and Husain, 2008; Ma et al., 2014; Wilken and Ma, 2004). More recently, it has been shown that higher-order summary information and ensemble representations of items presented in a display, which are not separate in visual working memory, influence performance in a change detection task (Brady and Tenenbaum, 2013; Liesefeld and Müller, 2019; Liesefeld et al., 2019). Additionally, brain activation in the lateral occipital cortex and intraparietal sulcus (IPS), specifically the superior IPS, reflects the VSTM for colors or objects during encoding and maintenance (e.g., Todd and Marois, 2004; Xu and Chun, 2006).

Consistent with the previous neuroscience studies of VSTM (e.g., Todd and Marois, 2004; Xu and Chun, 2006) and material perception

(Hiramatsu et al., 2011), we assessed the neural responses to memory for material categories using a region of interest (ROI) approach and calculated the averaged signal changes from two functionally defined ROIs: the superior IPS, which is engaged in VSTM (e.g., Todd and Marois, 2004; Xu and Chun, 2006), and the fusiform face area (FFA), which is sensitive to the perception of material categories (Hiramatsu et al., 2011). In this study, we used two types of change detection tasks to measure VSTM for material categories: an image-based change detection task and a material-based change detection task. In the image-based change detection task, participants were required to decide whether the test display image itself had changed in relation to the sample display. By contrast, in the material-based change detection task, participants decided whether a material category had changed in the test display. Consequently, if the VSTM for material categories were represented in the superior IPS and the FFA, a difference between the signal changes in image (i.e., object shape)-based and material-based change detection tasks would emerge in these regions. Therefore, we also examined the association between the brain activities in the superior IPS and the FFA using a correlation analysis.

2. Material and methods

2.1. Participants

Twenty-six participants from the participant pool at Kyoto University took part in this experiment (13 males and 13 females, mean age: 21.4 years; age range: 20–27 years).¹ They participated in this experiment in exchange for a book coupon (5,000 yen). Data from four participants were excluded from the following analyses because there were significantly several no-responses in the main scan (more than 50%) or because of anatomical findings (atrophy in the left temporal region). All participants were right-handed based on the H. N. Handedness Inventory (Hatta, 1975), had self-reported normal or corrected-to-normal visual acuity with non-magnetic glasses, and had normal color vision based on the Ishihara test for color blindness (Ishihara, 1968). All experimental protocols were approved by the Institutional Review Board of Kyoto University. All participants provided informed consent for inclusion in the study.

2.2. Stimuli and procedure

Forty-eight images rendered on eight nonsense shapes were selected from those in Hiramatsu et al. (2011), and these belonged to six material categories (metal, glass, stone, bark, leather, and fur; Figure 1). We created monochromatic images from each original image, by transforming the photograph from RGB to the $x^*y^*Y^*$ color mode and subsequently discarding the chromatic components x^* and y^* of the colored images, leaving only Y^* (luminance information).² These images were presented on a gray background.

Visual stimuli were displayed using a digital projector with a refresh rate of 60 Hz (U2-X2000, Plus, Japan). Image resolution was $1,024 \times 768$ pixels. Participants viewed visual stimuli with a mirror, which was positioned above the eyes, at a viewing distance of approximately 20 cm.

¹ We did not perform a priori power analysis. However, our sample size is comparable to that of some previous studies of VSTM (Hakim et al., 2019; Todd and Marois, 2004; Xu and Chun, 2006).

² As pointed out by Fleming (2014), there are a considerable number of variations in the possible appearance of a broad material class such as plastics. Hence, polyethylene bags, straws, and containers have widely diverging shapes and colors. A given exemplar, such as a container, can have many different shapes. Therefore, it should be noted that plastic may have any shape or color and that these low-level features were rarely diagnostic for plastic objects. We corrected all material and color stimuli in advance to make screen gamma values to 2.2 using Mcalibrator2 software (Ban and Yamamoto, 2013). However, we used monochromatic images of material objects.



Figure 1. Material images used in this study, from Hiramatsu et al. (2011). The labels presented at the bottom of the images arranged in tandem represent the name of the materials (starting from the left, metal, glass, stone, bark, leather, and fur).

The experiment was controlled by MATLAB (The MathWorks, Inc., Natick, MA) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Behavioral responses were collected using an MRI-compatible fiber-optic button box.

First, participants took part in one main scan consisting of four runs: two runs of image-based and two runs of material-based change detection tasks. The order of two change detection tasks was counterbalanced across participants. In other words, half of the participants first performed the image-based change detection task (two runs) and subsequently performed the material-based change detection task (two runs). The remaining participants did the two tasks in the opposite order: participants were informed whether the task was an image-based or a material-based change detection task in advance. In each trial, a sample display with two, four, or six objects on the circumference of a circle was presented for 500 ms, and participants were required to remember what images or materials were presented at each location. Each object subtended a visual angle of approximately 9.7° vertically and 6.3° horizontally, and the eccentricity between the fixation point and each object was approximately 11.4° . No two or more objects with the same categories were ever presented in a sample display. Following a delay screen for 1,500 ms, a test display was presented for 2,000 ms. Participants were required to detect a change in an object's image (Figure 2) or material (Figure 3) (50% of the trials were no change trials; the remaining 50% of the trials were change trials). The participants were to press with their index finger for a change response and with their middle finger for a no-change response. The finger configuration for making a response was counterbalanced across participants. Thus, half of the participants used their index finger for a change response and their middle finger for a no-change response, while the inverse of this configuration was used by the remaining participants. In the material-based change detection task, participants were required to decide whether the object in the test display changed at the material level. Hence, the test display in the

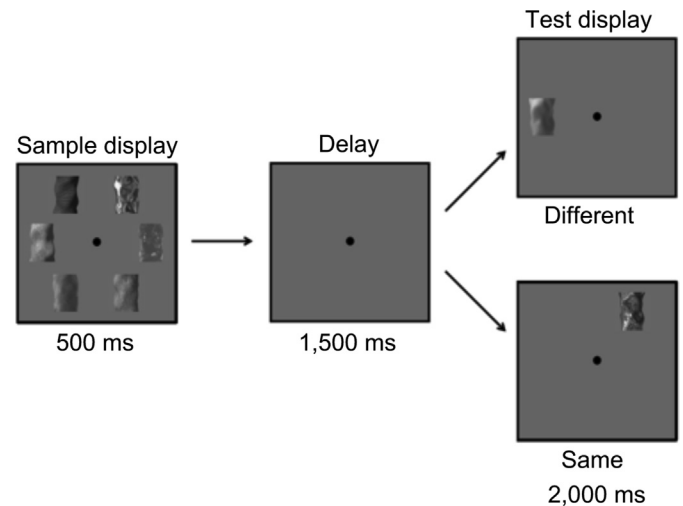


Figure 3. Examples from the material-based change detection task in the main scan. In the material-based change detection task, participants were required to decide whether the object in the test display changed at the material level. Hence, the test display in the material-based detection task always contained a stimulus of a shape different from the sample display, both in the same and in the different conditions. That is, in the different condition of the material-based task, both shape and material category changed in the test display (e.g., from stone to fur), whereas in the same condition, only the shape changed in the test display and the material category remained the same (e.g., metal).

material-based detection task always contained a stimulus of a shape different from the sample display, both in the same and in the different conditions. We did not include objects for which there was no change in either shape or material in the material-based condition to exclude the possibility that participants would perform the material-based change detection task using shape information. Note that in the image-based change detection task, there were two types of *different* conditions. In the material-change condition (*different condition 1*), both shape and material category changed in the test display, and participants had to judge it as different. In the shape-change condition (*different condition 2*), only the object shape changed in the test display, while the material category remained the same (e.g., bark shown in both the sample and test displays). In this case, these two objects were different at the image level; therefore, participants had to judge them as different. In the image-based task, neither the shape nor the category changed in the *same* condition. While at first, the image-based change detection task contained both shape-change and material-change detection, participants were only instructed to decide whether the object in the test display changed at the image level. The rest period was 2,000, 4,000, or 6,000 ms. Participants were required to press the button within 4,000 ms after the test display. Furthermore, there were several blank trials, in which only the fixation dot was presented for 6,000 ms. The presentation order of the set size (2, 4, and 6) and blank trials was random in each run. Each participant was tested with four runs (two runs for each task), each containing 12 trials per set size and 12 blank trials, resulting in 96 trials for each task (the image-based change detection task included 18 trials in the *same* condition, nine trials in *different condition 1*, and nine trials in *different condition 2* in each run; the material-based change detection task included 18 trials each in the same and different conditions, respectively, in each run).

To define the superior IPS, we conducted a color-based change detection scan with a design identical to that of the main scan, consisting of two runs (e.g., Todd and Marois, 2004; Xu and Chun, 2006). This procedure was the same as that of the main scan, except that we used six colors (red, green, blue, cyan, yellow, and magenta), each color square subtended a visual angle of approximately 6.3° vertically and horizontally, and for each trial, participants decided if a color change had occurred during the test display (Figure 4).

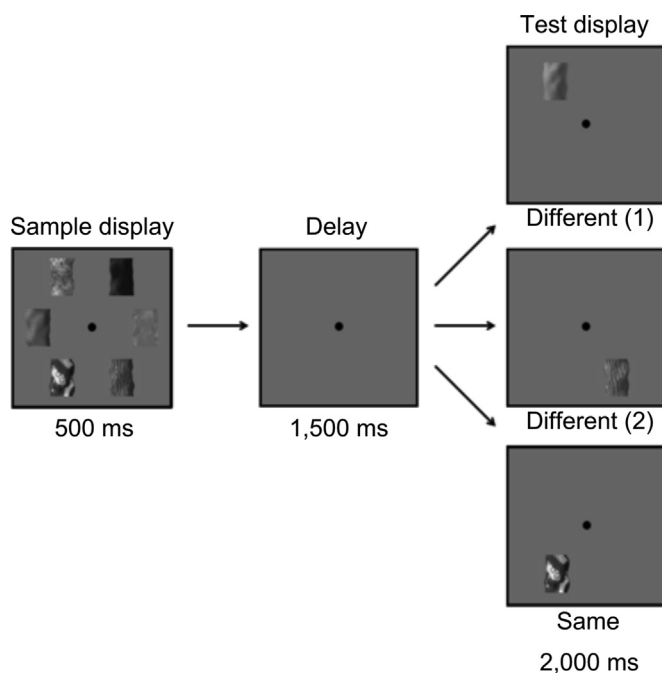


Figure 2. Examples from the image-based change detection task in the main scan. Participants performed the change detection task by deciding whether an image itself had changed or not, irrespective of whether the material categories were the same or different. In the image-based change detection task, there were two types of different conditions. In different condition (1), both shape and material category changed in the test display (e.g., from stone to fur). In different condition (2), only the shape changed in the test display and the material category remained the same (e.g., bark). In the image-based task, the shape and category did not change in the same condition (e.g., metal).

Finally, to define the FFA, participants took part in one run of an additional localizer scan. Each block consisted of 12 faces (six male faces and six female faces) from those in [Matsumoto and Ekman \(1988\)](#) or 12 scenes (six indoor scenes and six outdoor scenes) from the SUN database ([Xiao et al., 2010](#)). [Figure 5](#) depicts the face and scene stimuli used in the localizer scan. The two categories alternated every 30 s. The order of blocks was counterbalanced across participants. Participants judged whether the faces were male or female or whether the scenes were indoors or outdoors by pressing one of the two buttons. For each block, a fixation cross was displayed for 6 s with an instruction about finger/key constraints (“press with your index finger for a male or indoor scene” and “press with your middle finger for a female or outdoor scene”). Each image was subsequently presented for 500 ms with a 1,500-ms inter-stimulus interval, resulting in 30-s blocks. There were eight blocks of each type, resulting in 16 blocks.

Before the experiment, participants performed practice trials for each run outside the scanner. Additionally, they classified which objects belonged to each of the six material categories ([Figure 1](#)).

2.3. fMRI data acquisition

Data were collected on a 3T Siemens scanner (3.0T MAGNETOM Verio) at Kokoro Research Center, Kyoto University. Functional data were acquired with a T2*-weighted gradient-echo, echo-planar imaging (EPI) sequence (echo time [TE] = 25 ms; repetition time [TR] = 2,000 ms; flip angle = 75°; matrix = 64 × 64; field of view = 224 mm; 3.5 × 3.5 × 3.5 mm voxel size) with 34 axial slices. We acquired 180 volumes for each test run, 180 volumes for each color-localizer run, and 240 volumes for the additional localizer run. Structural images were acquired using a T1-weighted anatomical sequence (three-dimensional [3-D] magnetization-prepared rapid acquisition with gradient echo; TE = 3.51 ms; TR = 2250 ms; flip angle = 9°; matrix = 256 × 256; 1.0 × 1.0 × 1.0 mm voxel size).

2.4. Preprocessing

Preprocessing and statistical analyses were performed using the Statistical Parametric Mapping software (SPM) (SPM 8; <http://www.fil.ion.ucl.ac.uk>) in MATLAB. Images were corrected for slice acquisition time, motion-corrected with realignment to the first volume, spatially

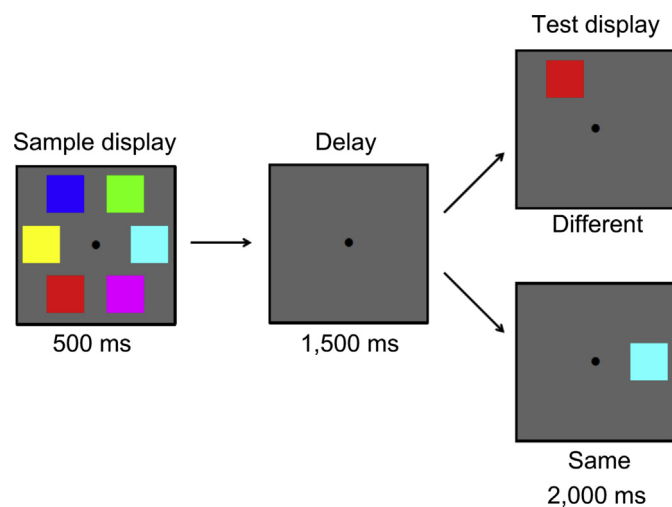


Figure 4. Examples from the color-based change detection task in the localizer scan. In the color-based change detection task, participants performed the change detection task by determining whether a color changed or not in the test display. In the different condition of the color-based task, a color changed in the test display (e.g., from blue to red), whereas in the same condition, the color remained the same (e.g., cyan).

normalized using the Montreal Neurological Institute EPI template, and spatially smoothed using an 8-mm Gaussian kernel.

2.5. Data analyses

To examine how neural responses differed with respect to the image-based and material-based change detection tasks within two a priori ROIs from the localizer scans (bilateral FFA and superior IPS), we ran three general linear model (GLM) analyses using SPM. First, we ran one GLM analysis on the data from the test scan, with four event types (set size 2, set size 4, set size 6, and blank trials) and task conditions (image-based vs. material-based) as separate regressors. To localize the superior IPS for each participant, we ran the second GLM analysis on the data from the color-localizer scan, with four event types as separate regressors and the behavioral performance as a covariate. Finally, to localize the FFA for each participant, we ran the third GLM analysis on the data from the additional localizer scan, with two block types (facial images and natural images) and fixation durations as separate regressors. As covariates of no interest, we also included six regressors for each dimension of head motion in each GLM analysis. These models estimated the contribution of each condition to the blood-oxygen-level dependent response using a boxcar function that was convolved with a canonical hemodynamic response function.

Next, to localize the superior IPS for each participant, we conducted a voxel-wise contrast for all set size conditions versus the blank trials on the data from the color-localizer scan. Facial images were also contrasted with images of natural scenes to localize the FFA for each participant. In each region, the voxel showing the greatest t value was selected as the center of a 4-mm spherical ROI, only if the t value reached a cluster level of significance threshold of $p < 0.05$ corrected (the voxel level threshold $p < 0.001$ uncorrected). Additionally, we set the cluster size threshold as at least 20 k , consistent with that of the previous studies ([Lieberman and Cunningham, 2009](#); [Woo et al., 2014](#)). We used the MarsBaR ROI toolbox to calculate percent signal changes in the test scan ([Brett et al., 2002](#)). Percent signal changes were calculated separately in the left and right regions, and these data were collapsed.

Additionally, we examined the association between the brain activities in the superior IPS and the FFA.

3. Results

3.1. Behavioral results

First, we analyzed the behavioral data from the main scan. Accuracy in the change detection task decreased as a function of set size in both the image-based (set size 2, 79.8%; set size 4, 68.0%; set size 6, 60.8%) and material-based blocks (set size 2, 64.8%; set size 4, 61.6%; set size 6, 56.7%). Consistent with the previous studies ([Todd and Marois, 2004](#); [Xu and Chun, 2006](#)), we used Cowan's K formula ([Cowan, 2001](#)) to estimate the number of objects encoded: $K = (\text{hit rate} + \text{correct rejection rate} - 1)N$ for each set size, where K is the number of objects encoded and N is the number of objects presented. If the number of objects encoded increases with an increase in set size, the trend in [Figure 6](#) will show a linear increase up to three or four. However, different from the previous studies using the change detection task with color or object stimuli (e.g., [Hakim et al., 2019](#); [Luck and Vogel, 1997](#); [Todd and Marois, 2004](#); [Xu and Chun, 2006](#)), the number of objects encoded for each set size did not increase in either the image-based or the material-based change detection tasks (see [Figure 6](#)). We performed a two-way within-subjects analysis of variance (ANOVA) on K values in the main scan, with the task (image-based vs. material-based) and set size conditions (2, 4, vs. 6). Although we observed a significant effect of task (1.29 in the image-based condition vs. 0.77 in the material-based condition), $F(1, 21) = 10.49$, $p = 0.004$,

Female faces



Male faces



Indoor scenes



Outdoor scenes



Figure 5. Face and scene stimuli used in the localizer scan (Matsumoto and Ekman, 1988; Xiao et al., 2010).

neither an effect of set size, $F(2, 42) = 1.39$, $p = 0.262$, nor interaction between task and set size conditions was observed, $F < 1$.³

Next, we compared correct responses between the image-based and material-based conditions using only the *same* trials. We performed a two-way within-subjects ANOVA with the task (image-based vs. material-based) and set size (2, 4, vs. 6) and observed a significant effect of task (image-based task, 53.0%; material-based task, 36.0%), $F(1, 21) = 37.13$, $p < 0.001$. This result suggests that *same* trials in the material-

³ We performed the same ANOVA on K values with all 26 participants. Although we observed a significant effect of task (1.23 in the image-based condition vs. 0.81 in the material-based condition), $F(1, 25) = 5.68$, $p = 0.025$, neither an effect of set size, $F(2, 50) = 1.01$, $p = 0.370$, nor interaction between task and set size conditions was observed, $F < 1$. As described in the “ROI analyses” section (3.2.1.), we observed the effect of set size only in the color-based change detection task used in the localizer scan (see Figure 6).

based change detection task are more difficult than those in the image-based change detection task.

3.2. Results of fMRI

3.2.1. ROI analysis

It should be noted that images from trials with errors were not included in the first-level analysis of the fMRI data. We examined VSTM for objects with material categories within two a priori ROIs: the superior IPS and the FFA. First, according to Xu and Chun (2006), the effect on VSTM was examined by comparing responses for the task and set size in our superior-IPS ROIs, identified in each participant from the functional localizer of a color-based change detection experiment. ROIs were found in 18 participants. Although we performed a two-way within-subjects ANOVA on the signal changes with the task (image-based vs.

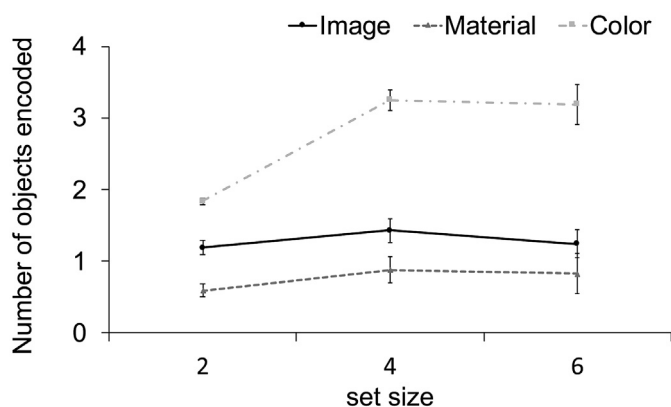


Figure 6. The number of objects encoded during the change detection task for the main and color-localizer scans as a function of the type of task (image-based vs. material-based) and set size (2, 4, and 6). Error bars represent the standard errors of the means.

material-based) and set size (2, 4, vs. 6), we did not observe any significant difference in signal change as a function of each condition, $F_s < 1$.

Next, we performed the same ANOVA on the signal changes in our bilateral FFA ROIs. ROIs were found in 18 participants; however, the data for one participant were excluded from the ANOVA as an outlier (the data of this participant deviated by at least 2 standard deviations from the mean signal changes). Figure 7 shows percent signal changes as a function of each condition. We observed a significant main effect of task, indicating lower activation under the material-based than the image-based conditions (0.06 vs. 0.11), $F(1, 16) = 4.83$, $p = 0.043$. However, neither the effect of set size nor interaction was significant, $F(2, 32) = 1.75$, $p = 0.190$, and $F < 1$, respectively.

We did not observe any reliable difference in signal changes between the same and different conditions (i.e., we did not observe either an effect of same/different trials or interaction including this factor) in either the superior-IPS or the FFA ROIs, which suggests that our results would not be explained by adaptation (repetition) effects.

One important issue for behavioral results is that we did not observe an effect of set size in either the image-based or material-based change detection task. One possible explanation might relate to the visual objects with material categories that were used in this study. In other words, it might have been more difficult for participants to memorize material objects than colors. In fact, we performed a two-way within-subjects ANOVA including the behavioral data from the color-localizer scan and observed a significant interaction between the task (image-based, material-based, vs. color-based) and set size (2, 4, vs. 6), $F(4, 84) = 5.19$, $p = 0.001$, which indicated that the effect of set size was significant in the color-based change detection task, $F(2, 20) = 56.41$, $p < 0.001$, but not in

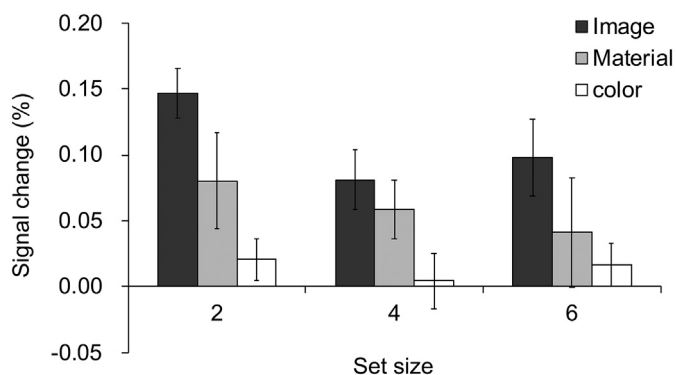


Figure 7. Percent signal change in the fusiform face area as a function of the type of task and set size (including both the same and different conditions). Error bars represent the standard errors of the means.

the image-based and material-based tasks, $F(2, 20) = 1.04$, $p = 0.371$ and $F(2, 20) = 1.21$, $p = 0.321$, respectively. Multiple comparisons (Bonferroni-corrected) showed that the VSTM capacity for set size 4 and 6 (3.3 and 3.2, respectively) was higher than that for set size 2 (1.8), $p_s < 0.001$ (see Figure 6). According to the behavioral data, the neural responses in the superior IPS increased as a function of set size. A two-way within-subjects ANOVA on the signal changes in the superior IPS revealed a marginally significant interaction between the task (image-based, material-based, vs. color-based) and set size (2, 4, vs. 6), $F(4, 68) = 2.33$, $p = .065$, which indicated that the effect of set size was significant in the color-based change detection task, $F(2, 16) = 5.17$, $p = .019$, but not in the image-based and material-based tasks, $F(2, 16) = 1.93$, $p = .178$ and $F(2, 16) = 0.15$, $p = .862$, respectively. Multiple comparisons (Bonferroni-corrected) showed that the signal changes for set size 4 (0.40) and set size 6 (0.38) were higher than that for set size 2 (0.31), with significance only for set size 4 ($p = .013$; set size 6, $p = .321$), in the color-based change detection. The pattern of these results is consistent with that of some previous studies that showed that people can retain an average of three to four simple features of objects or colors (Hakim et al., 2019; Luck and Vogel, 1997; Todd and Marois, 2004; Xu and Chun, 2006). These findings suggest that our manipulation of set size in the change detection task was adequate, and no effect of set size in either the image-based or material-based tasks attributes to the stimuli. It is possible that the low performance in the main scans is due to the removal of color information from material objects.

3.2.2. Correlation analyses

To examine the association between the brain activities in the superior IPS and the FFA, we performed a correlation analysis on the parameter estimates of the activity in the superior IPS and the FFA during the image-based versus material-based conditions (both the superior IPS and the FFA were localized in 16 participants). Because we did not observe a significant effect of set size in the behavioral and ROI analyses, we collapsed this variable. The correlation analysis was not significant, $r(16) = .033$, $p = 0.904$.

Next, we used the difference in parameter estimates between the image-based and material-based change detection tasks in the bilateral FFA or superior IPS as a covariate and performed a whole-brain analysis using the data from the main scan to examine the association between neural activation in the FFA or the superior IPS and other brain areas. Voxels were judged to show a significant difference for image-based versus material-based change detection tasks only if the t value reached a cluster level of significance threshold of $p < 0.05$ corrected (the voxel level threshold $p < 0.001$ uncorrected). This analysis revealed that activation was greater in the bilateral region of the precentral/post-central gyrus (right, 60, -4, 12; left, -64, -22, 20; $p_s < 0.05$ corrected; extent = 637/533; peak $t_s = 5.34/6.94$) for image-based versus material-based change detection tasks, including all three set size conditions, dependent on the FFA (Figure 8A and B).⁴ By contrast, we did not observe any significant regions dependent on the superior IPS.

4. Discussion

This study examined the brain mechanisms underlying VSTM for objects with material categories. To address this issue, we used a change detection task for visual shapes with six material categories. First, we defined two ROIs based on the localizer scans, and performed ROI

⁴ We labeled each region using the Talairach Client (Lancaster et al., 1997, 2000) after converting from the Montreal Neurological Institute to Talairach coordinates (<http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach>). It is not surprising that we also observed greater activation in the regions of the parahippocampal place area/fusiform face area (right, 32, -38, -12; left, -30, -48, -14; $p_s < 0.05$ corrected; extent = 376/518; peak $t_s = 8.66/6.83$) for image-based versus material-based change detection tasks.

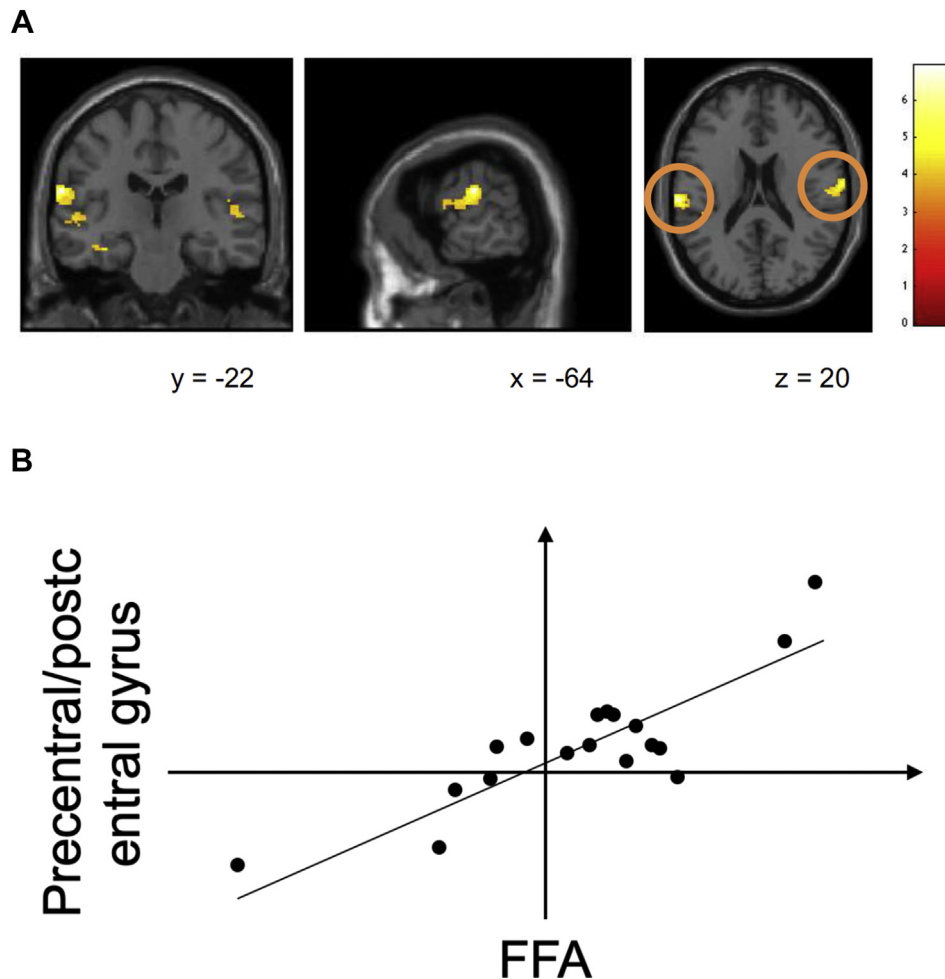


Figure 8. Whole-brain analysis from the main scan contrasting between the image-based and material-based change detection tasks with the parameter estimate in the bilateral fusiform face area (FFA) as a covariate. (A) The bilateral precentral/postcentral gyrus was activated in the image-based versus material-based change detection tasks. (B) Scatter plot of the parameter estimate between the bilateral FFA and bilateral precentral/postcentral gyrus.

analyses with the FFA and superior IPS as our ROIs. Although we did not observe any significant difference in percent signal changes in the superior IPS between the two conditions, we observed a significant difference in percent signal changes in the FFA between the image-based and material-based change detection tasks, which might be comparable to the previous studies on the perception of material categories (Hiramatsu et al., 2011; Suzuki, 2015). Our results indicate that the FFA may be involved in VSTM for objects in material categories in terms of the difference between images and material categories, which is related to the perceptual processing of multiple visual features such as color and texture, as shown by Cant and Goodale (2011).

It is worth noting that signal changes in the FFA were lower in the material-based change detection task than in the image-based task. One possible explanation is that the material-based change detection task might have contained more decision processes than the image-based task. For example, when one metal shape was presented in the upper left of a sample display and another metal shape was presented at the same position on a probe display, “change” detection may have been based on the visual features of the two metal shapes. However, participants might subsequently have had to correct their decisions to “no change” based on the material information of the two metal shapes. This additional step in the change detection task might have decreased both the capacity K value and signal changes in the FFA in the material-based condition. Indeed, our results indicate that no-change trials in the material-based change detection task were more difficult than those in

the image-based task. Alternatively, it is possible that our results rely on the difference in the conditions between the image-based and material-based change detection tasks. As mentioned in the “Stimuli and procedure” section, the image-based change detection task contained two types of different conditions in which there was a change in either shape or shape and material category, whereas the material-based task contained one different condition in which there was a change in both the shape and material category. Thus, participants could have just attended to the shape only in the image-based change detection task and monitored a shape change without the need to encode its material property. In other words, the difference between these two conditions might not be whether material was relevant to change detection but whether shape was relevant. Consequently, differential activity in the FFA might have reflected the processing of shape information in the image-based change detection rather than that of material information in the material-based change detection. Indeed, Cant and Goodale (2011) have shown that compared with perception of the surface texture, that of surface shape elicited stronger signal changes in the right FFA. However, Cant and Goodale have also reported stronger signal changes in the same regions when perceiving material properties rather than surface shape in their Experiment 2.

Additionally, differential activity between the image-based and material-based change detection tasks might depend on whether participants needed to ignore irrelevant changes rather than to store relevant information for change detection. Indeed, the shape of objects in the test

display changed during each trial, but participants had to ignore this type of change in the material-based change detection task. In other words, participants were required to respond no change when only the shape of objects changed. Alternatively, there were no irrelevant changes in the image-based change detection task. More recently, [Liesefeld et al. \(2017\)](#) reported that redundant changes (color and orientation) for a change detection task shortened reaction times in change detection, similar to pop-out in a visual search task. In our experiment, irrelevant changes of shapes might enforce different strategies on participants between the two conditions. Although it has been shown that visual search does not require accumulation of information about objects ([Horowitz and Wolfe, 1998](#)), it may be that VSTM for irrelevant information influences performance in a change detection task. Some of these factors might have contributed to the counterintuitive findings that the FFA activity was lower in the material-based condition than in the image-based condition.

More interestingly, however, we observed a significant positive association between the activations in the FFA and bilateral precentral/postcentral gyrus. In other words, we observed less brain activity in the left and right precentral/postcentral gyri in the material-based versus image-based change detection tasks, and these results were dependent on brain activity in the FFA. It has been demonstrated that these areas in particular are engaged in haptic perception (e.g., [Bodegård et al., 2001](#); [Reed et al., 2004](#)). This association suggests that even when we encode and maintain visual representations of material categories, haptic processing of visual information, such as roughness and glossiness, may be related to the VSTM for objects with material categories. Although some previous studies have reported that brain regions such as the ventrolateral occipitotemporal cortex and early visual cortex are involved in cross-modal interactions between visual and haptic processing (e.g., [Masson et al., 2016](#); [Snow et al., 2014](#)), this study provides a novel hypothesis that both the higher visual areas and somatosensory cortex may play important roles in the VSTM for objects with material properties. More recently, [Sun et al. \(2016\)](#) used fMRI to examine whether visual-tactile cross-modal activation occurs when people only observe material surfaces. In their study, participants observed 3-D objects made of four types of materials: Glossy, Glossy Control (Matte), Rough, and Rough Control (Textured). Participants observed a stream of these objects and were required to perform a 1-back matching task. Using multi-voxel pattern analysis to calculate classification accuracies for the above materials, [Sun et al. \(2016\)](#) showed that differences between Glossy and Rough conditions were decoded not only by the visual areas (e.g., V1, V2, V3, V4, and the lateral occipital region) but also by the secondary somatosensory cortex. In contrast, they did not observe a reliable difference between the Glossy Control and Rough Control conditions in the somatosensory cortex.⁵ Additionally, they showed that decoding performance was higher for the Gloss vs. Rough comparison than for the Glossy Control vs. Rough Control comparison only in the secondary somatosensory cortex. [Sun et al. \(2016\)](#) suggested that surface properties retrieved from visual stimuli activate a visual-tactile cross-modal network, which facilitates the multisensory processing of texture and roughness. Although we examined neural correlates of VSTM for objects with material properties comparing brain activities between the material-based and image-based conditions, it is possible that the image-based change detection is not adequate as a control task, different from the control conditions in [Sun et al.](#), because of the contamination accompanied by manipulation (i.e., the image-based change detection task in our control condition required participants to detect changes in

⁵ For example, in both the Glossy Control and Rough Control conditions, the objects looked matte because participants could not utilize the information for each perception ([Sun et al., 2016](#)). Brain encoding predicts brain response patterns from the sensory stimuli and experimental conditions, and brain decoding predicts behavioral responses from brain responses with computation of classification accuracies for stimuli and conditions (c.f. [Kriegeskorte and Douglas, 2019](#)).

either shape or in shape and material), making it difficult to interpret differential brain activities. In future studies, we will need to examine brain regions responsible for VSTM for material properties with a control task that is identical in all except for VSTM for material properties.

However, note that the above hypothesis is the reverse inference from the involvement of the somatosensory cortex with VSTM for objects with material properties. Thus, we need to take care of interpretation of our results, and conduct further studies to examine this issue in more detail, for example, with a combination of behavioral and fMRI data and the usage of neuroimaging databases (see [Poldrack, 2006](#)). Furthermore, it is unclear whether our results were actually reflective of VSTM, specifically the maintenance of visual representations of material categories rather than the perceptual processing of material categories. To dissociate whether brain activations observed in the IPS were reflective of VSTM encoding, maintenance, or retrieval, [Todd and Marois \(2004\)](#) conducted a change detection task with the retention interval extended to 9,200 ms (see also Experiment 3 by [Xu and Chun, 2006](#)). This additional experiment showed that the IPS was more activated at larger set sizes than smaller set sizes (set size 3 > set size 1) during both encoding and maintenance, but not during retrieval. They discussed that the IPS is sensitive to working-memory load during encoding and maintenance. However, we did not conduct this type of additional fMRI experiment to shorten the experiment time in this study.

As described in the “Results” section, we did not observe an effect of set size in either the image-based or material-based but in the color-based change detection tasks. Regarding the no effect of set size, we failed to achieve our research aim. One possible explanation is that participants might have had difficulty in discriminating the monochromatic visual objects with material-category changes. For this reason, participants might have used a different strategy in both the image-based and material-based change detection tasks from that used in the color-change detection task (e.g., memory of luminance or contrast of the visual objects presented in the display).⁶ It is possible that this unexpected strategy might have disguised the effect of set size. Alternatively, it might have been more difficult for participants to memorize the visual objects with material categories than colors. Previous studies have revealed that the VSTM capacity for facial expressions is slightly lower than four ([Švegar, 2011](#)). Because [Xu and Chun \(2006\)](#) observed lower memory capacity for more complex objects than simple objects, they examined whether behavioral performance was due to perceptual processing or memory limitations. As a result, longer encoding times for sample displays did not improve the VSTM capacity, and they concluded that the reduced VSTM capacity for more complex objects was due to memory rather than perceptual processing limitations. Thus, an upper storage limit of approximately four items can be observed for simple objects. As the amount of visual information per item increases, the storage limit decreases to lower levels ([Alvarez and Cavanagh, 2004](#)). Instead of the slot model, as used in our study (e.g., [Cowan, 2001](#); [Luck and Vogel, 1997](#)), the performances on the change detection tasks may be explained by flexible resource models (e.g., [Bays and Husain, 2008](#); [Ma et al., 2014](#); [Wilken and Ma, 2004](#)). According to resource models, resources are allocated to all of the items in a sample display. In conditions with smaller set size, the resources are sufficient for allocation to all of the items in the display, thus resulting in a higher precision of recall and a better decision in the change detection. In contrast, in conditions with larger set size, with sustained visual working memory capacities, the memory resources can be unevenly distributed among the items, thus resulting in prioritization of the items that are stored with enhanced precision compared to those that are not prioritized. Additionally, [Machizawa et al. \(2012\)](#) have shown that the amplitudes of the neural

⁶ We examined the mean luminance (metal: 83.06; glass: 90.95; stone: 86.71; bark: 83.83; leather: 82.69; fur: 85.87) and root mean square contrast of visual objects with material-category changes (metal: 39.99; glass: 20.00; stone: 28.29; bark: 27.43; leather: 26.64; fur: 26.21).

signals that are sensitive to the number of items in working memory (the contralateral delay activity) correlate with precision. More importantly, these resource models account for the variability of the visual working memory capacities in the stimulus category (Ma et al., 2014). Considering the results of these previous studies, our results may not necessarily suggest memory capacities of less than one for material categories. It is possible that the maintenance of the memories for material objects requires more memory resources than those for color stimuli, thus resulting in a lack of effect of set size. Additionally, in the material-based change detection task, the participants needed to maintain memories for the visual features and the categorical information of the materials. In other words, the material-based change detection task required the participants to use more memory resources than the image-based task did, which might have resulted in less precision in the material-based task. Thus, we need to do further research to determine whether the precision in the material categories depends on set size in the memory displays using delayed estimation tasks (Wilken and Ma, 2004), instead of typical change detection tasks with Cowan's K.

5. Conclusion

Although the present study was conducted with the aim of elucidating the neuronal mechanisms associated with VSTM for objects with material categories, we failed to achieve this research aim. However, our findings suggest that the FFA may be involved in VSTM for objects belonging to material categories, in terms of the difference between images and material categories and that this memory is mediated by the tactile properties of objects. It will guide further work to clarify the mechanisms underlying material perception, cognition, and memory.

Declarations

Author contribution statement

Sachio Otsuka: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Jun Saiki: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

This work was supported by JSPS KAKENHI [Grant Number 25135719, 16H01672, 18H05006].

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

This study was conducted using the MRI scanner and related facilities of Kokoro Research Center, Kyoto University.

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. *Psychol. Sci.* 15, 106–111.
 Awh, E., Barton, B., Vogel, E.K., 2007. Visual working memory represents a fixed number of items regardless of complexity. *Psychol. Sci.* 18, 622–628.
 Ban, H., Yamamoto, H., 2013. A non-device-specific approach to display characterization based on linear, nonlinear, and hybrid search algorithms. *J. Vis.* 13, 1–26.
 Bays, P.M., Husain, M., 2008. Dynamic shifts of limited working memory resources in human vision. *Science* 321, 851–854.

Bodegård, A., Geyer, S., Grefkes, C., Zilles, K., Roland, P.E., 2001. Hierarchical processing of tactile shape in the human brain. *Neuron* 31, 317–328.
 Brady, T.F., Alvarez, G.A., 2015. No evidence for a fixed object limit in working memory: spatial ensemble representations inflate estimates of working memory capacity for complex objects. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 921–929.
 Brady, T.F., Tenenbaum, J.B., 2013. A probabilistic model of visual working memory: incorporating higher order regularities into working memory capacity estimates. *Psychol. Rev.* 120, 85–109.
 Brainard, D.H., 1997. The psychophysics toolbox. *Spat. Vis.* 10, 433–436.
 Brett, M., Anton, J.L., Valabregue, R., Poline, J.B., 2002. Region of interest analysis using an SPM toolbox. *Neuroimage* 16 (suppl. 1), 1140–1141. Retrieved from <https://www.sciencedirect.com/journal/neuroimage/vol/16/issue/2/suppl/S1>.
 Cant, J.S., Arnott, S.R., Goodale, M.A., 2009. fMR-adaptation reveals separate processing regions for the perception of form and texture in the human ventral stream. *Exp. Brain Res.* 192, 391–405.
 Cant, J.S., Goodale, M.A., 2007. Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cerebr. Cortex* 17, 713–731.
 Cant, J.S., Goodale, M.A., 2011. Scratching beneath the surface: new insights into the functional properties of the lateral occipital area and parahippocampal place area. *J. Neurosci.* 31, 8248–8258.
 Cavina-Pratesi, C., Kentridge, R.W., Heywood, C.A., Milner, A.D., 2010a. Separate processing of texture and form in the ventral stream: evidence from fMRI and visual agnosia. *Cerebr. Cortex* 20, 433–446.
 Cavina-Pratesi, C., Kentridge, R.W., Heywood, C.A., Milner, A.D., 2010b. Separate channels for processing form, texture, and color: evidence from FMRI adaptation and visual object agnosia. *Cerebr. Cortex* 20, 2319–2332.
 Cowan, N., 2001. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114.
 Fleming, R.W., 2014. Visual perception of materials and their properties. *Vis. Res.* 94, 62–75.
 Fleming, R.W., Dror, R.O., Adelson, E.H., 2003. Real-world illumination and the perception of surface reflectance properties. *J. Vis.* 3, 347–368.
 Fleming, R.W., Wiebel, C., Gegenfurtner, K., 2013. Perceptual qualities and material classes. *J. Vis.* 13, 1–20.
 Goda, N., Tachibana, A., Okazawa, G., Komatsu, H., 2014. Representation of the material properties of objects in the visual cortex of nonhuman primates. *J. Neurosci.* 34, 2660–2673.
 Hakim, N., Adam, K.C.S., Günseli, E., Awh, E., Vogel, E.K., 2019. Dissecting the neural focus of attention reveals distinct processes for spatial attention and object-based storage in visual working memory. *Psychol. Sci.* 30, 526–540.
 Hatta, T., 1975. A study of handedness: handedness and manual activity. *Tekisei-kenkyu* 9, 1–13.
 Hiramatsu, C., Goda, N., Komatsu, H., 2011. Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. *Neuroimage* 57, 482–494.
 Ho, Y.X., Landy, M.S., Maloney, L.T., 2006. How direction of illumination affects visually perceived surface roughness. *J. Vis.* 6, 634–648.
 Horowitz, T.S., Wolfe, J.M., 1998. Visual search has no memory. *Nature* 394, 575–577.
 Ishihara, S., 1968. Ishihara's Tests for Colour-Blindness. Handaya, Tokyo.
 Komatsu, H., Goda, N., 2018. Neural mechanisms of material perception: quest on shitsukan. *Neuroscience* 392, 329–347.
 Kriegeskorte, N., Douglas, P.K., 2019. Interpreting encoding and decoding models. *Curr. Opin. Neurobiol.* 55, 167–179.
 Lancaster, J.L., Rainey, L.H., Summerlin, J.L., Freitas, C.S., Fox, P.T., Evans, A.C., Toga, A.W., Mazziotta, J.C., 1997. Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. *Hum. Brain Mapp.* 5, 238–242.
 Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., Fox, P.T., 2000. Automated talairach atlas labels for functional brain mapping. *Hum. Brain Mapp.* 10, 120–131.
 Lieberman, M.D., Cunningham, W.A., 2009. Type I and type II error concerns in fMRI research: Re-balancing the scale. *Soc. Cogn. Affect. Neurosci.* 4, 423–428.
 Liesefeld, H.R., Liesefeld, A.M., Müller, H.J., 2019. Two good reasons to say 'change!' - ensemble representations as well as item representations impact standard measures of VWM capacity. *Br. J. Psychol.* 110, 328–356.
 Liesefeld, H.R., Liesefeld, A.M., Müller, H.J., Rangelov, D., 2017. Saliency maps for finding changes in visual scenes? *Atten. Percept. Psychophys.* 79, 2190–2201.
 Liesefeld, H.R., Müller, H.J., 2019. Current directions in visual working memory research: an introduction and emerging insights. *Br. J. Psychol.* 110, 193–206.
 Luck, S.J., Vogel, E.K., 1997. The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281.
 Ma, W.J., Husain, M., Bays, P.M., 2014. Changing concepts of working memory. *Nat. Neurosci.* 17, 347–356.
 Machizawa, M.G., Goh, C.C.W., Driver, J., 2012. Human visual short-term memory precision can be varied at will when the number of retained items is low. *Psychol. Sci.* 23, 554–559.
 Masson, H.L., Bulthé, J., Op de Beeck, H.P., Wallraven, C., 2016. Visual and haptic shape processing in the human brain: unisensory processing, multisensory convergence, and top-down influences. *Cerebr. Cortex* 26, 3402–3412.
 Matsumoto, D., Ekman, P., 1988. Japanese and Caucasian Facial Expressions of Emotion (JACFEE) and Neutral Faces (JACNeuF). San Francisco State University, San Francisco, CA.
 Motoyoshi, I., Matoba, H., 2012. Variability in constancy of the perceived surface reflectance across different illumination statistics. *Vis. Res.* 53, 30–39.
 Nishida, S., Shinya, M., 1998. Use of image-based information in judgments of surface-reflectance properties. *J. Opt. Soc. Am. A: Opt. Image Sci. Vis.* 15, 2951–2965.

- Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442.
- Poldrack, R.A., 2006. Can cognitive processes be inferred from neuroimaging data? *Trends Cogn. Sci.* 10, 59–63.
- Reed, C.L., Shoham, S., Halgren, E., 2004. Neural substrates of tactile object recognition: an fMRI study. *Hum. Brain Mapp.* 21, 236–246.
- Sharan, L., Rosenholtz, R., Adelson, E.H., 2009. Material perception: what can you see in a brief glance? *J. Vis.* 9 (8), 784.
- Snow, J.C., Strother, L., Humphreys, G.W., 2014. Haptic shape processing in visual cortex. *J. Cogn. Neurosci.* 26, 1154–1167.
- Sun, H.C., Welchman, A.E., Chang, H.F., Luca, M.D., 2016. Look but don't touch: visual cues to surface structure drive somatosensory cortex. *Neuroimage* 128, 353–361.
- Suzuki, K., 2015. Visual texture agnosia in humans. *Brain Nerve* 67, 701–709.
- Švegar, D., 2011. Visual working memory capacity for emotional facial expressions. *Psychol. Top.* 20, 489–502. Retrieved from. <https://www.ffri.hr/psihologija/en/articles-2011-20-3>.
- Todd, J.J., Marois, R., 2004. Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428, 751–754.
- Wilken, P., Ma, W.J., 2004. A detection theory account of change detection. *J. Vis.* 4 (11), 1120–1135.
- Woo, C.-W., Krishnan, A., Wager, T.D., 2014. Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. *Neuroimage* 91, 412–419.
- Xiao, J., Hays, J., Ehinger, K., Oliva, A., Torralba, A., 2010. SUN Database: large-scale scene recognition from abbey to zoo. *IEEE Conf. Comput. Vis. Pattern Recognit.*
- Xu, Y., Chun, M.M., 2006. Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440, 91–95.