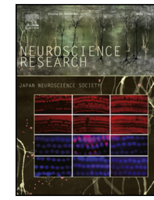
Neuroscience Research 89 (2014) 61–68

Contents lists available at ScienceDirect

Neuroscience Research

journal homepage: www.elsevier.com/locate/neures



Neuronal substrates characterizing two stages in visual object recognition

Tomoya Taminato[a], Naoki Miura[b,c], Motoaki Sugiura[c,d,*], Ryuta Kawashima[c]

[a] *Tohoku University School of Medicine, Sendai, Japan*
[b] *Faculty of Engineering, Tohoku Institute of Technology, Sendai, Japan*
[c] *IDAC, Tohoku University, Sendai, Japan*
[d] *IRIDeS, Tohoku University, Sendai, Japan*

ARTICLE INFO

Article history:
Received 27 May 2014
Received in revised form 25 August 2014
Accepted 1 September 2014
Available online 13 September 2014

Keywords:
Visual perception
Recognition
Semantic
Magnetic resonance imaging

ABSTRACT

Visual object recognition is classically believed to involve two stages: a perception stage in which perceptual information is integrated, and a memory stage in which perceptual information is matched with an object's representation. The transition from the perception to the memory stage can be slowed to allow for neuroanatomical segregation using a degraded visual stimuli (DVS) task in which images are first presented at low spatial resolution and then gradually sharpened. In this functional magnetic resonance imaging study, we characterized these two stages using a DVS task based on the classic model. To separate periods that are assumed to dominate the perception, memory, and post-recognition stages, subjects responded once when they could guess the identity of the object in the image and a second time when they were certain of the identity. Activation of the right medial occipitotemporal region and the posterior part of the rostral medial frontal cortex was found to be characteristic of the perception and memory stages, respectively. Although the known role of the former region in perceptual integration was consistent with the classic model, a likely role of the latter region in monitoring for confirmation of recognition suggests the advantage of recently proposed interactive models.

1. Introduction

Neuropsychological studies have indicated that visual object perception involves several processing stages. Most classic models distinguish between visual identification in the perception stage, which processes presented objects, and the memory stage, which verifies the resulting perceptual representations against representations stored in memory (Humphreys et al., 1999; Op de Beeck et al., 2000). The perception stage involves part-based analysis and analysis of global form; that is, feature extraction, segmentation, and shape analysis. During the memory stage, perceptual information is matched to each form stored in memory, which includes memory about the form of an object, its semantic properties, and its name (Humphreys and Riddoch, 1987; Humphreys et al., 1999). Although the concept of sequential processing has become outdated, the notion of temporal dynamics from primarily perceptual to more complicated processes is still valid and has provided the basis for recently developed interactive models.

Based on the classic two-stage model, neuroimaging studies using positron emission tomography (PET) or functional magnetic resonance imaging (fMRI) have addressed the candidate neural substrates involved in these stages and have added some aspects to the model (Gerlach et al., 1999, 2000, 2002; Op de Beeck et al., 2000; Pernet et al., 2004). Although the occipitotemporal regions are considered to play important roles in visual object processing, their detailed roles and the order of the processing sequence remain unclear. For example, involvement of the lateral occipitotemporal region in the perceptual stage has been suggested by the observations that this region responded to non-sense geometric objects (Shen et al., 1999) or without differentiating familiar and unfamiliar objects (Martin et al., 1996). On the other hand, modulation of activation in this region has been reported to depend on the degree of recognition success (Bar et al., 2001) or implicit semantic access (Pins et al., 2004), suggesting the involvement of this region in the semantic stage.

The validity of these arguments may, however, be questioned due to the limitations inherent in the experimental designs of these previous studies. Because visual identification is completed

* Corresponding author at: Department of Functional Brain Imaging, Institute of Development, Aging and Cancer (IDAC), Tohoku University Seiryo-machi 4-1, Aoba-ku, Sendai 980-8575, Japan. Tel.: +81 (0)22 717 7988; fax: +81 (0)22 717 7988.
E-mail address: motoaki@idac.tohoku.ac.jp (M. Sugiura).

http://dx.doi.org/10.1016/j.neures.2014.09.001

very quickly, ordinarily accomplishing many sub-processes in only 300 ms (Potter, 1976), previous fMRI and PET studies enhanced specific sub-processes by loading additional cognitive tasks targeting a specific operation. For example, to identify the process of accessing structural knowledge, subjects were required to judge whether presented objects were real objects or non-objects, and to isolate the semantic access process, subjects had to judge whether presented pictures were natural objects or artifacts (Gerlach et al., 2000). Although these experimental designs have undoubtedly provided important information about the functional anatomy of the visual recognition mechanism, they have left several points unanswered. First, these designs do not provide information about the order of the recruited sub-processes. Second, and more importantly, it is possible that the observed activation reflected some cognitive processes related solely to the given task per se; that is, processes unrelated to natural visual processing.

These issues of processing order and task-related activation can be addressed using a degraded visual stimuli (DVS) task. In a DVS task, subjects are first presented with low-spatial-resolution images to prevent immediate object identification, and the spatial resolution of the images is gradually improved to allow delayed identification. This special visual presentation technique enabled us to clarify the temporal characteristics of the sub-processes of visual recognition. Unlike the tasks in previous studies, which required additional cognitive operations, the DVS task requires only object recognition.

However, the DVS task has not yet been used to identify the neural substrates of the two stages in the classic models. Instead, it has been popularly used to demonstrate the temporally non-discrete nature of these stages and the existence of multiple sub-processes. A gradual transition between the two stages has been suggested (Bar et al., 2001, 2006; Ploran et al., 2007, 2011), consistent with the neural-network model assuming mutual facilitation of the two stages (Farah et al., 1993; Humphreys et al., 1999; Bar, 2003). Multiple sophisticated time-series models of expected signal change beyond the two-stage concept have been used, and distinct but overlapping sets of cortical regions were identified for each model (Carlson et al., 2006; Ploran et al., 2007, 2011).

In the current study, we used a DVS task to evaluate the classic two-stage model of visual object recognition. We characterized the early and late periods of the visual object-recognition process in terms of neural activation, with a focus on evaluation of the classic two-stage model; that is, we assigned perceptual and memory-related processes to the two periods. This evaluation method has not yet been implemented previously, and the data potentially obtained by this technique represent missing steps in the procession of this academic field from the classic two-stage model to recent interactive models. Although previous studies using the DVS task described the transition of the involvement of different areas through the recognition process, they did not test the validity of assigning each region to a specific stage or process. For example, Ploran et al. (2007) categorized the brain regions involved in object recognition according to temporal patterns of activation, applying an exploratory clustering approach. This procedure did not include statistical validation of the categorization of specific regions to one group versus another. Incorrect categorization was possible in these approaches with the technical consideration of a temporal correlation between models for different stages and the physiological consideration of the inhomogeneous hemodynamic function inherent across cortical regions.

In this study, subjects were asked to view low-spatial-resolution pictures (e.g., household items, animals, and fruit) that were gradually revealed. The following three characteristics were features of our version of the DVS task. First, to separate the periods in which each processing stage dominated, we asked subjects to respond once when they were able to guess what an object was and again when they were certain of their identification; the response times were used for the construction of the time-series model for analysis. Second, to render the time-series model for the two periods separable, we inserted “incomplete” trials in which a trial was interrupted in the middle of either the perception or memory stage. Finally, to remove the effects of the button press and image resolution on low-level visual processing, we included a control condition in which each subject performed the same task with a specific picture after having sufficiently practiced prior to the main fMRI experiment.

2. Materials and methods

2.1. Participants

In total, 47 healthy right-handed volunteers (41 males, 6 females, aged 19–31 years) participated. All subjects had normal vision, and none had a history of neurological or psychiatric illness. Handedness was evaluated using the Edinburgh Handedness Inventory (Oldfield, 1971). Written informed consent was obtained from all the subjects according to the guidelines of the ethics committee of Tohoku University and the declaration of Helsinki (1991).

Because data from 12 subjects were of insufficient quality (see Section 2.4), only the data of 35 subjects (31 males, 4 females) were analyzed.

2.2. Stimuli and tasks

Each stimulus set was a suite of 26 images that differed in their degree of degeneration (i.e., spatial resolution), produced from a single picture of an ordinary object with a white background by applying the Spatter filter of Photoshop CS (Adobe Systems, San Jose, CA). The degrees of degeneration were adjusted so that when the images were sequentially presented in order of resolution from the most degraded to wholly intact, the perceiver could guess the object in the middle of the sequence and could definitely recognize the object a few images before the end of the sequence (Fig. 1a). In total, 46 stimulus sets were prepared using pictures of different ordinary objects, including household items, animals, and fruit, which were selected from an original compilation of 56 stimulus sets used in a preliminary experiment (detailed later) and correctly recognized by all subjects. Among the 46 sets selected, 45 were used for the main condition, and one specific set, depicting an apple, was used for the control condition. During the fMRI experiment, each visual stimulus was back-projected onto a semi-translucent screen attached to the head coil of the MRI scanner, and subjects viewed the stimulus via a mirror. During the presentation of each stimulus set, the images were altered at a rate of 2.5 images/s, making the presentation duration of each image approximately 383 ms.

In each trial, the subject was required to respond twice: first, at the moment when he or she could guess what an object was and, second, at the moment when he or she was sure what the object was. For the second response, subjects were required to indicate whether the object was the same as the one they had guessed (or different). Subjects used their right middle fingers for the first response, and indicated “same” or “different” in the second response using their right index or ring finger, respectively. Subjects were encouraged to give their first response as soon as they could make a guess and to avoid careful decision making in an attempt to increase the number of times they could indicate “same” for the second response. We designated the period from the onset of the stimulus set to the first response as P, that from the first to the second response as M, and that from the second response to the end of the stimulus presentation as I. The perception and memory stages of the identification process were assumed to be maximally recruited in the P and M periods, respectively. The I period was