Neuroscience 349 (2017) 48–63

HARD-WIRED FEED-FORWARD VISUAL MECHANISMS OF THE BRAIN COMPENSATE FOR AFFINE VARIATIONS IN OBJECT RECOGNITION

HAMID KARIMI-ROUZBAHANI, [a,b]
NASOUR BAGHERI [a] **AND REZA EBRAHIMPOUR** [c,d,b*]

[a] *Department of Electrical Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran*

[b] *School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran*

[c] *Cognitive Science Research Lab., Department of Computer Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran*

[d] *Institute for Advanced Technologies, Shahid Rajaee Teacher Training University, Tehran, Iran*

Abstract—Humans perform object recognition effortlessly and accurately. However, it is unknown how the visual system copes with variations in objects' appearance and the environmental conditions. Previous studies have suggested that affine variations such as size and position are compensated for in the feed-forward sweep of visual information processing while feedback signals are needed for precise recognition when encountering non-affine variations such as pose and lighting. Yet, no empirical data exist to support this suggestion. We systematically investigated the impact of the above-mentioned affine and non-affine variations on the categorization performance of the feed-forward mechanisms of the human brain. For that purpose, we designed a backward-masking behavioral categorization paradigm as well as a passive viewing EEG recording experiment. On a set of varying stimuli, we found that the feed-forward visual pathways contributed more dominantly to the compensation of variations in size and position compared to lighting and pose. This was reflected in both the amplitude and the latency of the category separability indices obtained from the EEG signals. Using a feed-forward computational model of the ventral visual stream, we also confirmed a more dominant role for the feed-forward visual mechanisms of the brain in the compensation of affine variations. Taken together, our experimental results support the theory that non-affine variations such as pose and lighting may need top-down feedback information from higher areas such as IT and PFC for precise object recognition.

Key words: invariant object recognition, feed-forward vision, psychophysics, EEG, computational model.

*Correspondence to: R. Ebrahimpour, Cognitive Science Research Lab., Department of Computer Engineering, Shahid Rajaee Teacher Training University, Post Office Box: 16785-163, Tehran, Iran.
E-mail address: rebrahimpour@srttu.edu (R. Ebrahimpour).
URL: http://ccvlab.ir/ (R. Ebrahimpour).

http://dx.doi.org/10.1016/j.neuroscience.2017.02.050
0306-4522/

INTRODUCTION

Primates can accurately perform object categorization in fractions of a second (Thorpe et al., 1996; Fabre-Thorpe et al., 1998), despite substantial variations in objects' size, position, pose and the environmental lighting conditions. It has been suggested that, rapid object categorization is likely to be feed-forward (Riesenhuber and Poggio, 1999; VanRullen, 2007; DiCarlo and Cox, 2007; Afraz et al., 2014), and that more complex stimulus processing is achieved by feedback projections from higher visual areas. The latter include situations in which the target objects are in clutter (Hupe et al., 1998; Lamme et al., 1998; Bullier, 2001), occluded and in low contrast (Wyatte et al., 2012). There are also studies suggesting that object representations which are robust to variations in size and position are mainly constructed in a hardwired feed-forward manner in the visual pathways (Serre et al., 2005; Serre et al., 2007b), whereas for non-affine variations such as pose and lighting the brain needs to activate its top-down information resources at higher areas such as IT and PFC to yield invariant object representations (Bullier, 2001; Serre et al., 2005). However, these hypotheses lack supporting empirical data, which provided the motivation for the current study.

A set of behavioral studies have addressed the impact of individual variations on categorization performance. These include the studies reporting that changes in objects' size (Jolicoeur, 1987; Peissig et al., 2006; Zoccolan et al., 2009), position (see (Kravitz et al., 2008) for a review), pose (Edelman, 1995; Troje and Bulthoff, 1996), and the lighting conditions (Braje et al., 1998) of the environment exert a considerable influence on categorization performance (i.e. on both accuracy and time). However, no conclusions could be made on the contributions from the feed-forward and/or feedback mechanisms in these studies, since those studies have put no constraints on the feed-forward sweep or the feedback processing of visual information. To separate the contribution of the feed-forward/feedback pathways of information, the backward masking strategy has been frequently used and revealed to be highly effective in blocking the influence of feedback resources on categorization (Pollen, 1999; Lamme and Roelfsema, 2000; Serre et al., 2007a). A recent study, which investigated the feed-forward categorization using backward masking, suggested that the impact of variations on categorization is relative to the level of the applied variation (Ghodrati et al., 2014). However, since the evaluated variations were combined in that work, the relative impact of individual variations on feed-forward performance remains unknown. In the current study, we aimed to find possible differences between the mentioned variations in feed-forward object categorization to understand whether there is a potential need for feedback information when encountering some specific variations rather than others. To do this, an image set was generated in which 3D object models underwent parametrically controlled variations in lighting, pose, size and position independently from one another. The image set was used in a behavioral object categorization set-up with a backward masking protocol. A short stimulus presentation time was chosen to reduce the probability of the integration of top-down with the ongoing bottom-up visual information, so that the observed results can be associated with the feed-forward visual mechanisms.

Although useful in the study of human categorization performance, the behavioral experiments may be influenced by undesirable effects such as decision-related cognitive processes and response-related motor actions inherent in such experiments. To avoid these effects, we also designed a passive EEG recording experiment to gain access to the brain correlates for our behavioral observations. We expected to see differences between variations, since it was previously reported that the activity levels of IT neurons in non-human primates were highly modulated by object variations (Desimone et al., 1984; Booth and Rolls, 1998; Ashbridge et al., 2000; Vogels and Biederman, 2002; Hung et al., 2005; Freiwald and Tsao, 2010), and that such modulations were simply decodable from whole-brain MEG/EEG data when objects underwent variations in position (Carlson et al., 2011; Karimi-Rouzbahani et al., 2017) and size (Isik et al., 2014; Karimi-Rouzbahani et al., 2017). Our goal was to reduce the intervention of the top-down signals from higher visual areas in categorization. Therefore, contrary to our behavioral experiment, the subjects performed a category-irrelevant color-matching task while their whole-brain EEG signals were recorded. We used a shorter stimulus presentation time compared to previous studies (Carlson et al., 2011; Isik et al., 2014), to avoid potential impacts of feedback information from higher visual areas and to confine the task to feed-forward visual processing. A new analysis method was proposed in this work, which used 'Dunn' clustering index (Dunn, 1973) to explore the representational space of the EEG signals. The method helped in explaining the behavioral observations in time and space and provided several key advantages to traditional decoding approaches (Carlson et al., 2011; Hung et al., 2005; Isik et al., 2014). Isik et al. (2014) compared the dynamics of the appearance of size- and position-invariant representations. Results showed that the processing of size preceded position in time. Here we argue that such comparisons could have been biased since no attempt was made to equalize the separability of size- and position-affected images in the pixel space, nor was the potential bias removed from the representational results in the brain space. To avoid such problems, here we defined a modulation index to provide an unbiased comparison between the four different variations in the representational space.

Finally, a hierarchically organized feed-forward computational model was used as a 'proof of existence' to provide support that a feed-forward structure seems to be enough to explain the behavioral as well as the EEG observations. The model was selected based on several recent studies supporting its brain plausibility from both the performance as well as the representational aspects (Khaligh-Razavi and Kriegskorte, 2014; Yamins et al., 2014; Cadieu et al., 2014).

EXPERIMENTAL PROCEDURES

Stimulus set

An object image set was generated in which 3D object models underwent variations in size, position, pose and lighting. The image set included sixteen distinct object exemplars (freely downloaded from http://tf3dm.com/), which were categorized into the groups of 'animals', 'cars', 'faces' and 'planes' (Fig. 1A). To apply the parametrically controlled variations, Blender software was used (https://www.blender.org/). The size, position and pose of the objects as well as the lighting conditions of the 3D space were altered in different conditions. In the size conditions, the objects were resized so as to cover approximately from 5 k to 250 k pixels in the pixel space in 9 linear steps. This ranged approximately from 2 to 13.5 degrees of visual angle when the images were presented on the screen in the psychophysical and EEG experiments (Fig. 1B, third row from top). In position conditions, objects were put at different circular radii from the image center to provide different levels of eccentricity from the fovea. This led to 9 steps of position conditions ranging from 0.8 to 7.7 degrees of visual angle into the periphery in the experiments (Fig. 1B, forth row). The variation in pose was applied by rotating the objects around their X, Y and Z Cartesian axes simultaneously in steps of 45 degrees. This led to a total of 8 conditions ranging from 0 to 360 degrees of orientation (Fig. 1B, second row). Size, position and pose conditions shared a default condition which is shown only once in Fig. 1B, highlighted by the orange box. However, this condition (i.e. which shows the objects in 0 pose orientation, 5.8 degrees of size and 0 degree of position eccentricity) is considered in the evaluation of pose, size and position conditions. A uniform light source was used in the three above-mentioned variations which had almost no influence on the objects as they underwent those variations. However, the uniform light source was replaced by a pointing light source in different lighting conditions and was directed to the objects at the same distance but from different angles to generate the nine lighting conditions: top left, top, top right, right, bottom right, bottom, bottom left, left and front (Fig. 1B, first row). A unique gray-background, 512-by-512 pixel image was generated from each object-exemplar-variation-condition making a total of 560 images in the image set (i.e. 16 exemplars in 35 conditions).