NeuroImage 142 (2016) 67–78

Contents lists available at ScienceDirect

NeuroImage

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

The neural representation of objects formed through the spatiotemporal integration of visual transients

Gennady Erlikhman [a,*], Gennadiy Gurariy [a], Ryan E.B. Mruczek [a,b], Gideon P. Caplovitz [a]

[a] Department of Psychology, University of Nevada, Reno, USA
[b] Department of Psychology, Worcester State University, USA

ARTICLE INFO

Article history:
Received 16 October 2015
Accepted 17 March 2016
Available online 24 March 2016

Keywords:
Spatiotemporal objects
MVPA
V3A/B
Dorsal stream
Shape perception

ABSTRACT

Oftentimes, objects are only partially and transiently visible as parts of them become occluded during observer or object motion. The visual system can integrate such object fragments across space and time into perceptual wholes or spatiotemporal objects. This integrative and dynamic process may involve both ventral and dorsal visual processing pathways, along which shape and spatial representations are thought to arise. We measured fMRI BOLD response to spatiotemporal objects and used multi-voxel pattern analysis (MVPA) to decode shape information across 20 topographic regions of visual cortex. Object identity could be decoded throughout visual cortex, including intermediate (V3A, V3B, hV4, LO1–2,) and dorsal (TO1–2, and IPS0–1) visual areas. Shape-specific information, therefore, may not be limited to early and ventral visual areas, particularly when it is dynamic and must be integrated. Contrary to the classic view that the representation of objects is the purview of the ventral stream, intermediate and dorsal areas may play a distinct and critical role in the construction of object representations across space and time.

Introduction

A fundamental function of the visual system is to parse the environment into surfaces and objects. As difficult as this problem is for the seemingly simple case of static, unoccluded objects, self-motion and object motion complicate the task by creating complex, dynamic patterns of visual stimulation that must be segmented and grouped over space and time. The relative motion of objects and observers can make previously visible object parts occluded as nearer objects pass in front of farther ones, while other, once invisible parts become gradually revealed over time. Under certain circumstances, the visual system is able to overcome the problem of dynamic occlusion and represent what we will refer to as spatiotemporal objects. Spatiotemporal objects arise though the spatial and temporal integration of piecemeal information from object surfaces and the interpolation of missing, never-visible regions (Palmer et al., 2006). Here, we applied functional magnetic resonance imaging (fMRI) to identify neural correlates of the processes that underlie the representation of spatiotemporal objects.

A number of studies have examined the behavioral aspects of spatiotemporal object perception. In general, humans are quite good at identifying moving objects that are seen through a narrow slit (anorthoscopic viewing) or through many small apertures (Plateau, 1836; Zöllner, 1862; von Helmholtz, 1867/1962; Parks, 1965; Hochberg and Haber, 1968; Mateeff et al., 1993; Palmer et al., 2006). How does the visual system know whether two object fragments are aligned and can be integrated if both are in motion and only one is visible at a time? Palmer et al. (2006) proposed that the position, orientation, and velocity of object fragments are encoded and stored when visible and then updated during occlusion to maintain correspondences with visible fragments. Intermediate, never-visible regions are interpolated between visible and occluded regions. These integration, updating, and interpolation processes operate together to unify object parts separated across space and time in order to construct representations of spatiotemporal objects. Although there is an extensive literature examining the neural representation of occluded objects that are static (Edelman et al., 1998; Grill-Spector et al., 1998b; Kourtzi and Kanwisher, 2000, 2001), little work has been done to identify the neural correlates of spatiotemporal object perception (Yin et al., 2002; Ban et al., 2013).

Static object representations have traditionally been localized to the "what" visual processing stream which includes regions along the posterior and ventral temporal lobes (Tanaka, 1996; Ishai et al. 1999; Haxby et al., 2001; Pietrini et al., 2004; for reviews, see DiCarlo et al., 2012; Kravitz et al., 2013; Grill-Spector and Weiner, 2014). The ventral pathway is thought to be hierarchically organized: information from striate cortex (V1) is sequentially processed by subsequent areas leading to more complex and abstract representations (Van Essen and Maunsell, 1983; Riesenhuber and Poggio, 1999; Serre et al., 2007). Evidence for such a representational hierarchy comes from increasing receptive field sizes, increased response latencies, and increasing complexity of the preferred stimuli as one advances through the cortical

* Corresponding author at: Department of Psychology, University of Nevada, Reno, NV 89557, USA.
E-mail address: gerlikhman@unr.edu (G. Erlikhman).

http://dx.doi.org/10.1016/j.neuroimage.2016.03.044
1053-8119/

areas that make up the hierarchy, eventually arriving at representations that can be used for object identification and categorization (Rousselet et al., 2004; Hegdé and Van Essen, 2007; Kravitz et al., 2013).

Information from striate cortex is also thought to be passed, in parallel, along a dorsal, “where” pathway that extends from the occipitoparietal cortex to the intraparietal sulcus (IPS) and the parietal cortex (Mishkin et al., 1983; Goodale and Milner, 1992; Goodale et al., 1994; but see de Haan and Cowey, 2011). This pathway has been associated with eye movements (Sereno et al., 2001), the allocation of attention to objects (Ikkai and Curtis, 2011), spatial attention (Silver et al., 2005), and object manipulation and planning (Goodale et al., 1994). Recently, it has been suggested that the dorsal pathway trifurcates, with each branch responsible for distinct functions: from posterior parietal cortex (PPC), one pathway leads to prefrontal areas including the dorsolateral prefrontal cortex (DLPFC) and is involved in spatial working memory; the second pathway leads to premotor areas and is involved in motor planning and spatial action; the third pathway leads to the medial temporal lobe and is involved in spatial navigation (Kravitz et al., 2013).

Moving, dynamically occluded objects are not easily captured by the functions of either the dorsal or ventral pathways. Interaction with objects and navigation through the environment require accurate representations of relative spatial relationships for coordinated movement and grasping. These relationships include the precise representation of an object's 3D shape, orientation, position and motion relative to the observer. Object recognition and categorization, in contrast, ignore such spatially and temporally specific information, and instead require stable, position-, size-, and rotation-invariant representations that allow for recognition across a variety of spatial configurations and viewpoints. This invariant representational scheme poses a problem for how the visual system might recover and represent the structure of objects that only become visible gradually over time. Consider two different parts of an object that are seen successively, one at a time. In order to understand how they relate to each other spatially, that they form a single perceptual unit, and what the global form of that object is, non-invariant information about position and velocity is needed in order to accurately group these object fragments over time. On the other hand, such information is abstracted away by successive shape processing areas as it is passed along the ventral stream. It is our central hypothesis that the neural correlates of *spatiotemporal objects* – objects whose shape gradually emerges over space and time – may therefore span both ventral and dorsal pathways and representational schemes.

Here, we examine spatiotemporal objects that are produced by spatiotemporal boundary formation (SBF). SBF is the perception of illusory boundaries, global form, and global motion from spatiotemporally sparse element transformations (Shipley and Kellman, 1993, 1994, 1997). A familiar example is the gradual accretion and deletion of texture elements as when one surface passes in front of another, similarly textured surface (Gibson et al., 1969). Texture accretion and deletion are, however, just one of a wide variety of element transformations that can give rise to the percept of illusory boundaries and global form. In a typical SBF display, an invisible or virtual object moves in a field of undifferentiated elements. Whenever an element enters or exits the boundary of the object, the element changes in some property such as color, shape, orientation, or position. The sequence of element transformations results in the perception of an illusory contour corresponding to the virtual object's boundary despite the fact that no information about the object is present in any single frame. Illusory figures can be seen even in sparse displays when only a single element transforms per frame (Shipley and Kellman, 1994). SBF is therefore a *spatiotemporal* process in that information about the object's shape arrives gradually over time and is incomplete, with many regions of the boundary missing and requiring interpolation. SBF is also a robust phenomenon – shapes can be seen even though their properties may change in between element transformations such as changes in orientation, velocity, size, and even curvature (i.e., non-rigid deformations; see Erlikhman et al., 2014). An example of an SBF stimulus and the shapes used in the study can be seen in Fig. 1 and Movie 1. In the experiment reported here, SBF shapes were generated by either the rotation or displacement of Gabor elements in an array of randomly placed and oriented Gabors.

Spatiotemporal boundary formation provides a uniquely suitable test-bed for examining spatiotemporal object perception. First, in static views of these displays, no form information is present; all perceived forms are the result of dynamic integration and interpolation processes that define spatiotemporal objects. Second, a wide variety of element transformations can be used to produce the same global figures, making it possible to extract spatiotemporal object representations that are independent of local stimulus properties. Finally, it is simple to create displays that contain element transformations, but which do not form a global percept, producing a natural control comparison.

Using SBF stimuli, we were able to disentangle the contributions of global motion, which was present in all displays, from spatiotemporal form perception, which only occurred for a subset of stimuli, all while

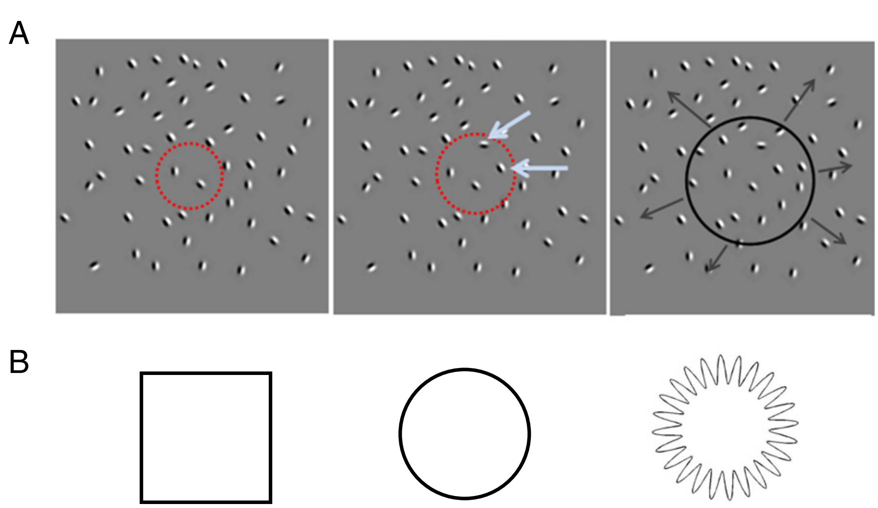

Fig. 1. Stimulus displays exemplifying spatiotemporal boundary formation (SBF) as used in the current experiments. A. An invisible object (red dotted circle, Frame 1) expanded and contracted. Elements entering the boundary of the object (blue arrows, Frame 2) rotated or were displaced by a small amount. The resulting percept (Frame 3) was of expanding and contracting illusory contours. B. The three shapes used in the experiment. The boundaries of the third shape could not be recovered because of the rapid modulation of the contour relative to the density of the background elements. The resulting percept was of flickering elements in a ring-like configuration, but without a clearly-defined form as for squares and circles. This served as the control, no-shape condition.