NeuroImage 59 (2012) 625–634

Contents lists available at ScienceDirect

NeuroImage

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

Form and motion make independent contributions to the response to biological motion in occipitotemporal cortex

James C. Thompson *, Wendy Baccus

Department of Psychology, George Mason University, USA

ARTICLE INFO

Article history:
Received 10 April 2011
Revised 29 June 2011
Accepted 18 July 2011
Available online 2 August 2011

Keywords:
Biological motion
Motion
Form
Human MT
Human MST
Extrastriate body area

ABSTRACT

Psychophysical and computational studies have provided evidence that both form and motion cues are used in the perception of biological motion. However, neuroimaging and neurophysiological studies have suggested that the neural processing of actions in temporal cortex might rely on form cues alone. Here we examined the contribution of form and motion to the spatial pattern of response to biological motion in ventral and lateral occipitotemporal cortex, using functional magnetic resonance imaging (fMRI) and multivoxel pattern analysis (MVPA). We found that selectivity to intact versus scrambled biological motion in lateral occipitotemporal cortex was correlated with selectivity for bodies and not for motion. However, this appeared to be due to the fact that subtracting scrambled from intact biological motion removes any contribution of local motion cues. Instead, we found that form and motion made independent contributions to the spatial pattern of responses to biological motion in lateral occipitotemporal regions MT, MST, and the extrastriate body area. The motion contribution was position-dependent, and consistent with the representation of contra- and ipsilateral visual fields in MT and MST. In contrast, only form contributed to the response to biological motion in the fusiform body area, with a bias towards central versus peripheral presentation. These results indicate that the pattern of response to biological motion in ventral and lateral occipitotemporal cortex reflects the linear combination of responses to form and motion.

Introduction

Humans have a remarkable ability to perceive the movements of other people, even from stimuli consisting only of light points at each joint (Johansson, 1973). Psychophysical studies have indicated that the perception of point-light biological motion involves contributions of both body shape/form (Beintema and Lappe, 2002; Garcia and Grossman, 2008) and motion (Neri et al., 1998; Mather et al., 1992; Thurman and Grossman, 2008) cues. Computational models that can successfully discriminate features such as walking direction from biological motion using only a form-based template or "snapshot neurons" (Giese and Poggio, 2003; Lange and Lappe, 2006) or motion (Giese and Poggio, 2003) suggest that both of these cues might be important for action recognition. How each of these cues contributes to the neural processing of biological motion is, however, poorly understood.

Human observers appear to vary their dependence on form or motion depending on stimulus and task demands, suggesting a flexible cue integration process (Thirkettle et al., 2009; Thurman et al., 2010). Consistent with a role of both form and motion pathways in the processing of biological motion, selectivity for intact versus scrambled biological motion has been reported in human extrastriate body form-selective regions such as the extrastriate body area (EBA) and the fusiform body area (FBA), as well as the motion-selective human MT+ complex (Grossman and Blake, 2002; Jastorff and Orban, 2009; Peelen et al., 2006). The EBA and MT+ are at least partly overlapping, however, and the biological motion selectivity observed might reflect the activity of a single population selective for one cue only. For example, a study by Peelen et al. (2006) that used multivoxel pattern analysis (MVPA) to show that biological motion selectivity in EBA and MT+ was correlated with selectivity for static bodies (versus other objects) but not with selectivity for moving versus static gratings. This evidence suggests that selectivity to biological motion in lateral occipitotemporal cortex reflects the activity of a population selective for form cues alone. A recent monkey neurophysiology study also challenged the proposal that the motion pathway is involved in processing observed actions. Singer and Sheinberg (2010) found that cells in the monkey anterior superior temporal sulcus (STS) and inferotemporal (IT) cortex that showed discriminative responses to specific actions did so mostly via the encoding of sequences of poses rather than using continuous motion information. Others, however, have reported evidence for both static pose and moving action responsive cells in a similar region of monkey anterior STS and IT (Vangeneugden et al., 2011). The action cells showed greater discrimination between forward and backward walking than the static and action cells, suggesting that this region does receive motion

* Corresponding author at: Department of Psychology, George Mason University, 4400 University Drive MS3F5, Fairfax, VA, 22030, USA. Fax: +1 703 993 1359.
E-mail address: jthompsz@gmu.edu (J.C. Thompson).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.07.051

inputs relevant to action recognition (Oram and Perrett, 1996). It is therefore important to better understand the processing of both motion and form cues in the processing of biological motion in extrastriate cortex.

One important consideration is that neural populations in extrastriate cortex that process motion information from biological motion might not necessarily show selectivity for intact versus scrambled stimuli. These populations might still be important for coding biological motion features, such as movement dynamics, in a manner that is not dependent on configuration. For example, cells in MT process local speed and direction of motion within small receptive fields, and thus might be expected to respond similarly to intact and scrambled biological motion (Desimone and Ungerleider, 1986; Lagae et al., 1994). In other subregions in MT+, such as MSTd, cells show selectivity for particular motion patterns, although many cells respond to multiple patterns (Duffy and Wurtz, 1991; Geesaman and Andersen, 1996; Lagae et al., 1994). Here we sought to determine the extent to which motion responses combine with form responses in occipitotemporal regions to contribute to the response to biological motion.

To determine the contribution of form and motion to the response to biological motion, we used fMRI and multi-voxel pattern analysis (MVPA) in ventral and lateral occipitotemporal cortex regions of interest (ROIs) that were defined by their form- or motion-selectivity. One of the difficulties in manipulating the contribution of these two cues is that changing the form of the stimulus, such as through spatial scrambling, also changes the global pattern of motion. Similarly, adding motion noise can alter the temporally integrated form representation. In order to investigate the contribution of form and motion to the biological motion response, the present study instead used differences in the position-dependence of form and motion populations. Motion-selective cells in area MT are highly position-dependent and have receptive fields that do not include ipsilateral space, whereas MST neurons have receptive fields that cover ipsi- and contralateral space (Allman and Kass, 1971; Huk et al., 2002). In contrast, there is evidence of foveal and/or contralateral biases in form-selective regions that also respond to ipsilateral stimuli (Sayres and Grill-Spector, 2008; Schwarzlose et al., 2008; Weiner and Grill-Spector, 2011). We hypothesized that by manipulating spatial position we could vary the contribution of form and motion populations to the response to biological motion, and that the effects of spatial position would vary depending on the region.

Material and methods

Participants

Ten healthy individuals (4 males; age range = 22–37; mean = 26.2, S.E. = 1.6 years) participated in either two or three fMRI scanning sessions. All participants were right handed with normal or corrected-to-normal vision. Participants were compensated $15 per hour and provided written informed consent in accordance with the Human Subjects Review Board at George Mason University. Two of the participants were the authors.

fMRI data acquisition and preprocessing

fMRI data were collected using a Siemens Allegra 3 T scanner and a single channel birdcage coil at the Krasnow Institute for Advanced Study at George Mason University. Visual stimuli were displayed on a rear projection screen and viewed by participants on a coil mounted angled mirror. We acquired gradient-echo, echoplanar imaging scans (33 axial slices; 4 mm slice thickness and 1 mm gap; TR/TE = 2000/30 ms; flip angle = 90; 64×64 matrix with 3.75×3.75 mm in-plane resolution, field of view = 24 cm). In addition, between two and four T1 whole-head anatomical structural scans were collected using a three-dimensional magnetization-prepared rapid-acquisition gradient echo (MPRAGE) pulse sequence (160 1 mm thick slices; 256×256 matrix; field of view = 260 mm; 0.94 mm voxels, TR/TE = 2300/3 ms). Cortical surfaces were reconstructed from the MPRAGE scans using Freesurfer software (http://surfer.nmr.mgh.harvard.edu/). This automated processing involves motion correction, averaging of the images, removal of non-brain tissue, intensity normalization, and segmentation to create a representation of the pial surface. The pial surface model was also inflated to support visualization of activation occurring within cortical sulci.

Preprocessing of fMRI data included removal of the first three volumes from each run to compensate for the time it took to reach equilibrium magnetization. The FEAT (fMRI Expert Analysis Tool) tool of the FSL (fMRI of the Brain Software Library) toolbox (http://www.fmrib.ox.ac.uk/fsl/) was used for fMRI analysis. The fMRI time series were high-pass filtered at 128 s and motion corrected. No spatial smoothing was applied at any stage of analysis. For each run, a general linear model that included gamma function regressors (sd = 3, lag = 6) for motion correction and the onset and duration of each block was used to estimate the response to each category of stimuli. Prewhitening was also used to remove temporal autocorrelation of the fMRI time series. The results of the FEAT analyses were then projected onto the Freesurfer generated surface of each individual.

Identification of regions of interest

We used independent localizers to identify human body form-, motion-, and biological motion-selective regions in lateral and ventral occipitotemporal cortex. Stimuli for the localizer and the main experiment were presented during neuroimaging data acquisition using Presentation software (Neurobehavioral Systems, La Jolla, California, USA). Regions of interest (ROIs) were identified on the cortical surface of each participant using a combination of statistical and anatomical criteria. To identify the extrastriate body area (EBA) and fusiform body area (FBA), participants viewed blocks of headless bodies or blocks of chairs and performed a 1-back task (Downing et al., 2001). Stimuli covered 13.6° vertical by 8.3° horizontal visual angle. The EBA (Fig. 2) was identified as the cluster showing a greater response to bodies versus chairs located in or adjacent to the ascending portion of the posterior inferior temporal sulcus (pITS) (Downing et al., 2001; Peelen et al., 2006) (see Results). The FBA was defined at the cluster with a greater response to bodies versus chairs located in the middle fusiform gyrus (Peelen and Downing, 2005).

We sought to identify human motion areas MT and MST with techniques based on previously described methods (Beauchamp et al., 2007; Dukelow et al., 2001; Huk et al., 2002; Smith et al., 2006). We first identified the motion-selective hMT+ complex as the cluster located in or adjacent to the ascending portion of the pITS that responded to motion, by presenting participants with blocks of movies of expanding and contracting radially moving dots (100% coherence) or blocks of static images of dots. One hundred black dots were presented on a 50% gray background within a circle of 11.4° diameter, with the central 0.5° left blank. Dots moved at a speed of 7°/s, and had a lifespan of 500 ms. During the motion blocks, the motion alternated between expansion and contraction at a rate of 0.5 Hz. To identify MT and MST, we then had participants maintain central fixation as they were presented with a separate localizer consisting of blocks of expanding or contracting radial motion (5.5° diameter) centered 8.1° to the left or right of fixation. Human MT was then defined as the posterior or ventral cluster of at least 20 vertices within the MT+ cluster that showed a response to contralateral motion ($p<0.01$ uncorrected) but no response to ipsilateral motion ($p>0.05$) (Dukelow et al., 2001; Huk et al., 2002). MST was defined as the more anterior or superior cluster of at least 20 vertices within the MT+ cluster that showed a response to contralateral motion ($p<0.01$ uncorrected) and ipsilateral motion ($p<0.01$ uncorrected) (Dukelow et al., 2001; Huk et al., 2002). While these definitions are consistent