NeuroImage 172 (2018) 689–702

Contents lists available at ScienceDirect

NeuroImage

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

Task-dependent enhancement of facial expression and identity representations in human cortex

Katharina Dobs [a,b,c,*], Johannes Schultz [a,d], Isabelle Bülthoff [a], Justin L. Gardner [b,e]

[a] Department Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Max-Planck-Ring 8, 72076 Tübingen, Germany
[b] Laboratory for Human Systems Neuroscience, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
[c] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar Street, Cambridge, MA 02139, USA
[d] Division of Medical Psychology and Department of Psychiatry, University of Bonn, Sigmund Freud Str. 25, 53105 Bonn, Germany
[e] Department of Psychology, Stanford University, 450 Serra Mall, Stanford, CA 94305, USA

ARTICLE INFO

Keywords:
Attentional enhancement
Face perception
Functional magnetic resonance imaging
Multivoxel pattern analysis
Dynamic faces

ABSTRACT

What cortical mechanisms allow humans to easily discern the expression or identity of a face? Subjects detected changes in expression or identity of a stream of dynamic faces while we measured BOLD responses from topographically and functionally defined areas throughout the visual hierarchy. Responses in dorsal areas increased during the expression task, whereas responses in ventral areas increased during the identity task, consistent with previous studies. Similar to ventral areas, early visual areas showed increased activity during the identity task. If visual responses are weighted by perceptual mechanisms according to their magnitude, these increased responses would lead to improved attentional selection of the task-appropriate facial aspect. Alternatively, increased responses could be a signature of a sensitivity enhancement mechanism that improves representations of the attended facial aspect. Consistent with the latter sensitivity enhancement mechanism, attending to expression led to enhanced decoding of exemplars of expression both in early visual and dorsal areas relative to attending identity. Similarly, decoding identity exemplars when attending to identity was improved in dorsal and ventral areas. We conclude that attending to expression or identity of dynamic faces is associated with increased selectivity in representations consistent with sensitivity enhancement.

Introduction

Humans effortlessly discern two distinct and, indisputably, ecologically important visual aspects of faces: identity and expression. Identity does not change with expression, and, while idiosyncrasies of expression can give clues to identity (O'Toole et al., 2002; Xiao et al., 2014; Lander and Butcher, 2015; Dobs et al., 2016, 2017), is largely associated with static features such as the shape of the eyes and mouth. Conversely, expression is eminently changeable and associated with movement of facial features. What neural mechanisms allow selective extraction of facial form features, such as identity, and facial motion features, such as dynamic expression, from this complex interplay of features? Much work has focused on how attention operates on simpler visual features (Maunsell and Treue, 2006; Bisley, 2011; Carrasco, 2011; Maunsell, 2015) or distinct high-level object categories (e.g., Peelen et al., 2009; Çukur et al., 2013; Peelen and Kastner, 2014). Considerably less is known about mechanisms for selecting different complex aspects of the same stimuli like facial identity and expression.

Attention to identity or expression is known to modulate activity in the diverse set of cortical areas associated with face processing. In humans, face-selective areas have been found along the fusiform (e.g., FFA), the inferior occipital cortex (OFA), and the posterior superior temporal sulcus (STS) (Kanwisher et al., 1997; Haxby et al., 2000), with similar organization in macaques (Tsao et al., 2008; Fisher and Freiwald, 2015; Weiner and Grill-Spector, 2015). While anatomical segregation of processing for expression and identity has been reported (Sergent et al., 1994; Haxby et al., 2000; O'Toole et al., 2002; Andrews and Ewbank, 2004), other reports suggest overlap (Calder and Young, 2005; Bernstein and Yovel, 2015; Lander and Butcher, 2015; Fisher et al., 2016). When subjects selectively attend to expression or eye gaze, greater average BOLD response in STS has been observed, while ventral areas such as FFA showed larger average responses when attending to identity (Hoffman and Haxby, 2000; Narumoto et al., 2001; but see also Ganel et al., 2005).

Attention thus modulates the average BOLD response in face

* Corresponding author. Department of Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany.
E-mail address: katharina.dobs@gmail.com (K. Dobs).

https://doi.org/10.1016/j.neuroimage.2018.02.013
Received 18 July 2017; Received in revised form 2 February 2018; Accepted 6 February 2018
Available online 10 February 2018
1053-8119/

processing areas; however, this existing evidence is equivocal as to how attentional selection for identity and expression functions. Average modulations of BOLD responses in cortical areas could be interpreted in at least two different ways. First, increased activity in late stage areas could be a signature of processes that improve sensory representations. This sensitivity enhancement would result in enhanced representations of the attended facial aspect, and might be similar to effects reported in early visual areas (Motter, 1994; McAdams and Maunsell, 2000; Reynolds et al., 2000; Martinez-Trujillo and Treue, 2004; Cohen and Maunsell, 2011; for review see Maunsell, 2015) or to effects reported in later face-selective areas when attention is directed to visual features which distinguish individual faces (Gratton et al., 2013). Second, aside from the fact that stronger responses may have higher signal-to-noise ratio, the magnitude of response and the fidelity of response patterns could be unrelated. This would be evident in an increase of overall activity but no increased distinctiveness of the neural representations of the attended facial aspect. Indeed, for representations of the image contrast of visual stimuli, BOLD responses in early visual cortex are increased with spatial attention (Buracas and Boynton, 2007; Li et al., 2008; Murray, 2008; Pestilli et al., 2011), but do not change the slope of their relationship with contrast as would be expected by a higher fidelity representation to differences in contrast. These increased responses can, nonetheless, account for behavioral performance enhancement by an efficient selection model (Pestilli et al., 2011; Hara and Gardner, 2014) in which responses with larger magnitude carry a larger effect on perceptual decisions (Pelli, 1985; Lee et al., 1999; Verghese, 2001; Eckstein et al., 2009; Mante et al., 2013). Thus, at least these two, non mutually-exclusive possibilities of sensitivity enhancement and selection exist for attentional selection for expression and identity of faces.

To test whether a selection mechanism or sensitivity enhancement provides a better account of the neural mechanisms underlying attentional selection of identity or expression, we measured cortical responses while subjects were instructed to attend to the identity or the expression of dynamic face stimuli. Specifically, subjects detected changes in either expression or identity of naturalistically animated avatar faces (Fig. 1). Importantly, these stimuli allow fine-tuning of changes along these dimensions (Dobs et al., 2014), which in turn allows matching task difficulty online during the experiment. We assessed fidelity of representation by our ability to decode (Kamitani and Tong, 2005) individual exemplars of identity and expression in functionally (Kanwisher et al., 1997; Kanwisher and Yovel, 2006; Kanwisher, 2010) and topographically (Wandell, 1999; Wandell et al., 2007) localized cortical areas across the visual hierarchy. We hypothesized that a pure selection account would result in no improvement in classification accuracy, but simply an increase in overall activity. In contrast, sensitivity enhancement would be evident in improved classification accuracy when each facial aspect is attended relative to when that aspect is not attended.

Material and methods

Subjects

Six observers (two female; mean age: 32 years) from the RIKEN Brain Science Institute volunteered as subjects. All observers were right-handed and had normal or corrected-to-normal vision and provided informed written consent prior to the experiment. All procedures for psychophysical and neuroimaging experiments were approved in advance by the RIKEN Functional MRI Safety and Ethics Committee.

Stimuli and display

Each of the four basic stimuli used in this experiment consisted of a short video displaying a female avatar face animated by a facial expression (from neutral to the peak facial expression). We used two female avatar faces and two dynamic facial expressions (Fig. 1A) to create these four clips. Each clip differed from the others either by the face or by the expression displayed. Two additional avatar faces and two additional facial expressions served as change stimuli for the change detection task. Briefly, the procedures used to make these basic and change animations were as follows (for details see Dobs et al., 2014). First, four facial expressions (angry, happy, disgust, surprise) were motion-recorded from one non-professional female actor following a previously validated and published procedure (Dobs et al., 2014). All facial expressions were 2 s long and started from a neutral expression that proceeded to the peak expression. Second, four avatar faces (ID1−ID4) were designed in Poser 8 (SmithMicro, Inc., Watsonville, CA, USA). We selected two identities (ID1 and ID2) and two facial expressions (angry and happy) to create the basic stimuli. The remaining two identities and two facial expressions were used to create change stimuli that served as targets (see Design and procedure). Change stimuli were created by linearly morphing a basic face into a target face (ID1 to ID3, ID2 to ID4; 10% morph steps) or by linearly morphing one basic facial expression into another target facial expression (angry to disgust, happy to surprise; 5% morph steps). Third, the two basic faces and their facial morphs were animated by the motion-captured facial expressions and their motion morphs (for details about the motion-retargeting procedure see Curio et al., 2006). Importantly, each basic clip could be modified parametrically either in terms of its identity information or its dynamic expression information, allowing matching and continuous control of task difficulty in both tasks (see Design and procedure below). Finally, the animations were rendered as Quicktime movies of 2 s duration (450 × 600 pixels, 30 frames at 60 Hz) in 3ds Max 2012 (Autodesk, Inc., San Rafael, CA, USA).

Stimuli were presented using MGL (http://justingardner.net/mgl) and Matlab (version R2010a; The MathWorks, Inc., Natick, MA, USA). Images were back-projected on a projection screen (Stewart Filmscreen, Torrance, CA, USA; 28.3 × 20.0 cm size, 800 × 600 pixel resolution, 60 Hz refresh rate) located inside the scanner at 50 cm viewing distance from the subject. Animation stimuli were scaled to a size of approximately 9° × 12° and positioned such that the tip of the avatar's nose was located at the center of the screen.

Design and procedure

Subjects performed tasks in which they detected changes in either expression (expression task) or identity (identity task) on a stream of one of the four basic moving face stimuli (Fig. 1A). The design of the task was adapted from a previously published paradigm used to investigate feature-based attention in low-level stimuli (Liu et al., 2011). Each 20 s trial (Fig. 1B) began with a 0.4 s period in which a centrally presented letter instructed subjects to either detect changes in the expression (letter ‘E’), identity (letter ‘I’), or to perform no task (letter ‘0’). Throughout the rest of the trial, subjects were required to fixate a cross presented at the center of the screen. Following a 0.2 s gray screen period, eight animated 2 s long face stimuli (Fig. 1A) were shown, each followed by a short gray screen of 0.3 s between stimuli (inter-stimulus interval, ISI). Trials were either no-change or change trials. In no-change trials, the stream of eight face stimuli consisted of eight repetitions of one of the four basic stimuli, whereas change trials consisted of repetitions of one basic stimulus interspersed with change stimuli containing small changes in expression or identity. Subjects had to press a button within the presentation of a stimulus (2 s) or up to 0.5 s afterwards to report a change in the identity (identity task) or expression (expression task) of the stimulus. Reaction time was measured from the beginning of stimulus presentation until the response occurred, or from the beginning of the previous stimulus for responses faster than 0.3 s. The last stimulus was followed by an extra 0.3 s gray screen (total of 0.6 s) to allow for a 2.6 s response window for this final stimulus. Each trial was followed by a 0.5 s inter-trial interval (ITI).

Changes in expression or identity of stimuli were controlled by staircase procedures to maintain performance across tasks, stimuli and subjects at a similar threshold level. In each trial, one to four of the eight presentations of the basic stimuli were randomly replaced by a change