NeuroImage 61 (2012) 1287–1299

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Neural correlates of temporal integration in face recognition: An fMRI study

Yunjo Lee [a,*,1], David Anaki [b,1], Cheryl L. Grady [a,c], Morris Moscovitch [a,c]

[a] Rotman Research Institute, Baycrest Centre, Toronto, Canada
[b] Bar-Ilan University, Ramat-Gan, Israel
[c] University of Toronto, Toronto, Canada

ARTICLE INFO

Article history:
Accepted 25 February 2012
Available online 5 March 2012

Keywords:
Analytic processing
Configural processing
Face identification
Face part
Part-based processing

ABSTRACT

Integration of temporally separated visual inputs is crucial for perception of a unified representation. Here, we show that regions involved in configural processing of faces contribute to temporal integration occurring within a limited time-window using a multivariate analysis (partial least squares, PLS) exploring the relation between brain activity and recognition performance. During fMRI, top and bottom parts of a famous face were presented sequentially with a varying interval (0, 200, or 800 ms) or were misaligned. The 800 ms condition activated several regions implicated in face processing, attention and working memory, relative to the other conditions, suggesting more active maintenance of individual face parts. Analysis of brain-behavior correlations showed that better identification in the 0 and 200 conditions was associated with increased activity in areas considered to be part of a configural face processing network, including right fusiform, middle occipital, bilateral superior temporal areas, anterior/middle cingulate and frontal cortices. In contrast, successful recognition in the 800 and misaligned conditions, which involve analytic and strategic processing, was negatively associated with activation in these regions. Thus, configural processing may involve rapid temporal integration of facial features and their relations. Our finding that regions concerned with configural and analytic processes in the service of face identification opposed each other may explain why it is difficult to apply the two processes concurrently.

Introduction

The visual world surrounding us is replete with complex stimuli that cannot be apprehended instantaneously. Perception of a unified representation, therefore, depends crucially on temporal integration of correlated information both across views and within a view. Despite growing interest in understanding the interaction between temporal structure (synchronized changes in visual information) and spatial vision (e.g., see reviews by Melcher and Colby, 2008 and Blake and Lee, 2005), the neural and anatomical correlates of temporal integration are largely unknown. Moreover, the temporal integration processes involved in perceiving complex stimuli are yet to be determined. In face processing, for example, most research has focused on how facial features are spatially combined to form a unified representation and only a small number of studies have examined temporal aspects of integration (e.g., Anaki and Moscovitch, 2007; Anaki et al., 2007; Singer and Sheinberg, 2006). While a few studies examined perceptual awareness of an occluded face (Hulme and Zeki, 2007; Yi et al., 2008) or flashed face (Keysers et al., 2005), the focus of those studies was to measure consciousness in the absence of stimulus perception. Other studies examined the time course of whole-face recognition processes (Barbeau et al., 2008) or discrimination of face parts or spacing among them (Pitcher et al., 2007). Though related to some of the issues addressed in those studies, the present study differs from them in that we investigated neural correlates of temporal integration of static face parts when they are separated by varying time intervals that promote either configural or analytic processing of faces (Anaki and Moscovitch, 2007; Anaki et al., 2007). Although our study was concerned with temporal integration and faces, and in particular the distinction in this regard between configural (holistic) and non-configural processes, our findings and conclusions may not be specific to faces but may apply to other configural and non-configural processing in other domains.

Face perception is thought to entail particularly well-adapted perceptual processes, commonly referred to as *holistic* (Tanaka and Farah, 1993) or *configural*, which involve fine integration of facial features into a unitary representation. A typical marker of these processes is enhanced recognition of upright faces as compared to inverted faces (the face inversion effect, Maurer et al., 2002; Yin, 1969) or to misaligned faces whose top and bottom parts are spatially offset but shown simultaneously (Young et al., 1987). This impediment in processing misaligned or inverted faces is attributed to the difficulty in extracting holistic or configural information from them (Jacques and Rossion, 2010). Recently, we have shown that face parts, separated by blank intervals up to 400 ms, can be integrated and processed

* Corresponding author at: Rotman Research Institute, Baycrest Centre, University of Toronto, 3560 Bathurst Street, Toronto, ON, Canada M6A 2E1. Fax: +1 416 785 2862.
E-mail address: ylee@rotman-baycrest.on.ca (Y. Lee).
[1] These authors contributed equally to this research.

1053-8119/$ – see front matter
doi:10.1016/j.neuroimage.2012.02.073

configurally, yielding inversion effects comparable in magnitude to those of faces presented as a whole (Anaki et al., 2007). Beyond that interval, *analytic* or *part-based* processing appears to prevail, indicated by a marked reduction in the behavioral face-inversion effect. Presentation of a random pattern mask during the interval between the two face parts disrupted integration in the 200 ms interval condition, resulting in decreased recognition; the mask had no effect either on integration or perception of the separate parts in the 800 ms condition (Anaki and Moscovitch, 2007). These findings, along with others (see Anaki and Moscovitch, 2007), indicate that integration is achieved through a short-lasting, limited capacity buffer which temporarily maintains the visual input while integration occurs, and allows configural processing. We speculated that facial temporal integration would likely occur within iconic memory during *information persistence* (Coltheart, 1980) which lasts 150–300 ms after stimulus offset (see discussion in Anaki et al., 2007). Although we tested only faces, it is likely that similar processes occur in other domains (Ruff et al., 2007; Saneyoshi et al., 2011), though the content on which these processes operate, and hence some of the regions that are implicated, may be specific to each domain.

To date, there are no studies on the neural and anatomical manifestations of temporal integration of separate facial parts leading to face identification. Are temporally integrated faces handled by the same regions that integrate spatial information into a configural representation of faces? Prime candidates include "core" face-sensitive areas, such as the fusiform gyrus. The fusiform face area (FFA), functionally defined by contrasting response to faces vs. response to other categories of objects (Kanwisher et al., 1997), is engaged in face detection (Nestor et al., 2008; Rossion, 2008; Tong et al., 2000), representation of generic faces (Loffler et al., 2005), and identification of individual faces (Grill-Spector et al., 2004; reviewed in Kanwisher and Yovel, 2006). Although the FFA is involved in holistic representation of faces (Andrews et al., 2010; Rossion et al., 2000; Schiltz and Rossion, 2006), it is equally activated to facial features, both internal and external (shape), and to their configuration (Liu et al., 2009; Rotshtein et al., 2007; Yovel and Kanwisher, 2004; also see Andrews et al., 2010; Axelrod and Yovel, 2010). Crucially, in Mukamel et al. (2004) where a stimulus was successively flashed, striate and extrastriate areas show persistent neural activity even after stimulus termination resulting in signal increases not proportional to a stimulus presentation rate. This nonlinearity was stronger in higher visual areas such as the FFA even for non-preferred stimuli (i.e., faces as well as houses), and it might provide the short-term visual memory buffer needed for the temporal integration to occur (Mukamel et al., 2004).

Other distinct areas within the fusiform gyrus, in the vicinity of the FFA but not encompassed by it, appear to be sensitive only to changes in face configuration (Maurer et al., 2007; Ng et al., 2006; Schiltz and Rossion, 2006; Schiltz et al., 2010). Furthermore, configural face processing may involve a number of regions outside the occipitotemporal cortex (Maurer et al., 2007; Ng et al., 2006; Rotshtein et al., 2007). For example, Rotshtein et al. correlated discrimination of configural change in the face (measured outside the scanner) with blood-oxygen-level dependence (BOLD) responses when the participant monitored such stimulus changes. They found a positive correlation in several areas, such as the left middle cingulate gyrus, right insula, putamen and prefrontal regions, as well as the right fusiform and bilateral inferior occipital gyri. Although the anterior/middle cingulate cortex is not traditionally considered part of the face network (Haxby et al., 2000a), significant activity in this region was observed during configural face processing (Ng et al., 2006) and face encoding and subsequent recognition (Haxby et al., 1996). Prefrontal areas were also observed in configural processing of faces as opposed to featural processing (Maurer et al., 2007). If a temporally integrated face is represented in a configural manner, we would expect to find a positive correlation between activity of these regions and recognition performance in the short interval conditions. Taken together, we would expect the FFA proper to respond early in the short interval conditions providing a visual buffer facial parts to be used in temporal integration, but additional regions to contribute to configural representations of integrated faces.

Previous findings have shown that if face components were separated by a long enough interval such that the first one was not integrated with the second that arrived later in the visual buffer, configural representation would suffer and performance would depend on analytical processes based on individual face parts (Anaki and Moscovitch, 2007). In such cases, investigators have speculated that activity in configural processing regions may be detrimental to identification by component parts (de Gelder and Rouw, 2000; Macrae and Lewis, 2002; see Fig. 6 in Maurer et al., 2007), yet no-one has provided neural evidence for the incompatibility of analytic and configural processing of faces. If such incompatibility exists, activation in configural regions should be positively correlated with performance on tests sensitive to configural processing but negatively correlated with tests sensitive to analytic processing. The reverse should hold for regions that support identification based on piecemeal information derived separately from each of the component parts. Such regions, however, are not as clearly delineated as those associated with configural processing; for example, Rotshtein et al. (2007) found no regions showing brain-behavior correlations for featural changes.

To investigate the neural correlates of facial temporal integration, we used an event-related fMRI design, in which we presented famous faces whose top and bottom halves were separated by either 0, 200 or 800 ms inter-stimulus intervals (ISI 0, ISI 200, ISI 800, respectively) and measured the participant's recognition of the faces (yes/no responses) as an index of temporal integration. Although whole faces were not shown in any condition, the 0 ms condition had no blank screen, so that the top and bottom halves were presented sequentially, creating a whole-face percept. In a comparison condition, we used a misaligned face (MIS) in which both parts were presented simultaneously. As noted, identification of such faces has been shown to be based on analytic, rather than configural, processes. We chose to use misaligned rather than inverted faces to maintain a common orientation across our stimuli.

We first assessed the magnitude of BOLD signals in the FFA using a univariate, region of interest (ROI) analysis. Then, we used a multivariate method, Partial Least Squares (PLS; McIntosh et al., 1996, 2004), to assess a functional network of distributed neural regions whose activity co-varies with the stimulus conditions (task PLS) and with recognition performance in each condition (behavior PLS). Behavior PLS was used to identify a set of regions that contributed directly, either positively or negatively, to recognition performance, i.e., areas where activity was correlated with recognition (see Materials and methods section for justification for using PLS).

In the univariate ROI analysis of task-related effects in the FFA, a couple of potential results are possible. If FFA activity is related to temporal integration then it should be more active in the 0 and 200 conditions (i.e., ISI 0, ISI 200 > ISI 800, MIS). On the other hand, it is possible that the FFA would show some increase of activity in the short interval conditions, but a larger increase in the long intervals, compared to no interval or MIS conditions (i.e., ISI 800 > ISI 200 > ISI 0, MIS), simply because in the former case there is a double pulse produced by the sequential presentation of the two face halves. Activity might also be larger in the FFA during the long interval condition (ISI 800) because this condition would involve the greatest demand from maintenance of individual facial parts that are clearly segregated.

We also considered it important to examine the activity in the rest of the brain to see how that activity was related specifically to recognition. To do so, we chose to use PLS analysis because we believed it to be the best tool for the purpose (see justification in