ELSEVIER

Neuropsychologia 45 (2007) 3234–3241

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Facial expression and gaze-direction in human superior temporal sulcus

Andrew D. Engell *, James V. Haxby

Department of Psychology, Center for the Study of Brain, Mind, and Behavior, Princeton University, United States

Received 2 January 2007; received in revised form 14 May 2007; accepted 4 June 2007
Available online 5 July 2007

Abstract

The perception of facial expression and gaze-direction are important aspects of non-verbal communication. Expressions communicate the internal emotional state of others while gaze-direction offers clues to their attentional focus and future intentions. Cortical regions in the superior temporal sulcus (STS) play a central role in the perception of expression and gaze, but the extent to which the neural representations of these facial gestures are overlapping is unknown. In the current study 12 subjects observed neutral faces with direct-gaze, neutral faces with averted-gaze, or emotionally expressive faces with direct-gaze while we scanned their brains with functional magnetic resonance imaging (fMRI), allowing a comparison of the hemodynamic responses evoked by perception of expression and averted-gaze. The inferior occipital gyri, fusiform gyri, STS and inferior frontal gyrus were more strongly activated when subjects saw facial expressions than when they saw neutral faces. The right STS was more strongly activated by the perception of averted-gaze than direct-gaze faces. A comparison of the responses within right STS revealed that expression and averted-gaze activated distinct, though overlapping, regions of cortex. We propose that gaze-direction and expression are represented by dissociable overlapping neural systems.

Keywords: fMRI; Gaze; Expression; Face; Perception; Non-verbal communication

Humans are highly sensitive to non-verbal cues that provide important information about the emotional state, focus of attention, and future behavior of others (Simpson & Crandall, 1972). The accurate perception and analysis of these cues is an important component of 'social cognition' and facilitates appropriate behavior in complex social groups. Two particularly salient and powerful cues are facial expressions and gaze-direction. An individual's facial expression is a reliable and accessible indication of their internal emotional state (Ekman & Rosenberg, 2005). The importance of expressions as social signals is highlighted by evidence that their frequency and intensity increase in social situations (Fridlund, 1994; Jancke & Kaufmann, 1994). The gaze-direction of others communicates their current focus of attention and provides clues to their future intentions. The importance of this information is underscored by evidence that humans as young as 10 weeks old (Hood, Willen, & Driver, 1998), non-human great apes (Tomasello, Call, & Hare, 1998), and monkeys (Emery, Lorincz, Perrett, Oram, & Baker, 1997) automatically orient their attention to the target of another's gaze (Friesen, Chris Kingstone, & Alan, 1998).

The brain regions responsible for the perception of facial identity and facial gesture have been shown to be largely dissociable both in the macaque (Hasselmo, Rolls, & Bayliss, 1989) and human (Andrews, & Ewbank, 2004; Engell, Gobbini, & Haxby, 2006; Hoffman & Haxby, 2000) cortex. Haxby, Hoffman, and Gobbini (2000) have, therefore, proposed a distributed human neural system for face perception that posits a primary role of the superior temporal sulcus (STS) in perception of dynamic facial features and a primary role of the ventral temporal cortex, including the fusiform gyrus, in the perception of facial features that are invariant over dynamic changes (i.e. identity).

Single-cell recording in monkeys (Hasselmo, Rolls, & Baylis, 1989) and humans (Ojemann, Ojemann, & Lettich, 1992), as well as human neuropsychological (Rapcsak, Kaszniak, & Rubens, 1989) and neuroimaging (Gur, Skolnick, & Gur, 1994; Narumoto, Okada, Sadato, Fukui, & Yonekura, 2001) studies, have all implicated the STS in expression processing. Similarly, the STS has been implicated in the perception of gaze-direction by single-cell recording in monkeys (Perrett et al., 1985), human

* Corresponding author at: Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08544, United States. Tel.: +1 609 258 1063; fax: +1 609 258 1113.
E-mail address: aengell@princeton.edu (A.D. Engell).

0028-3932/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2007.06.022

neuroimaging (Hoffman & Haxby, 2000; Puce, Allison, Bentin, Gore, & McCarthy, 1998; Wicker, Michel, Henaff, & Decety, 1998), and human electrophysiological responses as measured with scalp (N170; Puce, Smith, & Allison, 2000) and intracranial (N200; Allison, Puce, Spencer, & McCarthy, 1999; McCarthy, Puce, Belger, & Allison, 1999) electrodes.

Despite this similarity, it is unknown if a single system supports the perception of both types of stimuli. Many recent studies have investigated the behavioral effect of gaze-direction on expression recognition (e.g. Adams & Kleck, 2003, 2005; Ganel, Goshen-Gottstein, & Goodale, 2005) and of facial expression on the reflexive orienting effects of gaze-direction (e.g. Holmes, Richards, & Green, 2006; Hori et al., 2005; Mathews, Fox, Yeind, & Calder, 2003; Putman, Hermans, & van Honk, 2006). These studies suggest a behavioral interdependence that is further supported by evidence that the amplitude of event-related potentials recorded from the scalp of 4-month olds is larger during perception of angry faces with direct-gaze than angry faces with averted-gaze (Striano, Kopp, Grossmann, & Reid, 2006; but see Pourtois et al., 2004). However, to date, there have been no direct comparisons of the neural response evoked within the STS by both expression and gaze-direction perception.

The objective of our study was to explore whether perception of facial expression and averted-gaze resulted in dissociable patterns of activity within the STS region of the lateral temporal cortex. Subjects viewed blocks of either neutral faces with direct-gaze (“control” condition), emotional faces with direct-gaze (“expression” condition), or neutral faces with averted-gaze (“averted-gaze” condition) while we scanned their brains with functional magnetic resonance imaging (fMRI). We examined the signal changes evoked by the perception of emotional faces and averted-gaze relative to the control condition to determine whether they shared a single neural substrate.

Both facial expression and gaze-direction represent facially communicated social information. However, we hypothesized that the observation of these two gestures would evoke dissociable neural representations due to the distinctiveness of each gesture’s informational content (i.e. emotional state and target of attention, respectively). Further, we predicted that we would find this dissociation in the STS region given its putative role in processing facial gesture.

1. Materials and methods

1.1. Subjects

Thirteen participants (5 females and 7 males, age 22–33) were recruited from the community in and around Princeton University (Princeton, NJ). One subject was excluded from analysis due to excessive head motion. The participants were all right-handed and had normal or corrected vision. All participants gave informed consent prior to the experiment and were fully debriefed at its completion in accordance with the policies of Princeton University’s Institutional Review Panel.

1.2. Stimuli

Pictures were taken from the Pictures of Facial Affect (Ekman & Friesen, 1976) and modified to display different gaze directions. The full stimulus set included 11 individuals posing in nine conditions. The nine conditions included four emotional expressions (anger, disgust, fear, surprise) displaying a direct-gaze, four averted-gaze poses (full left, partial left, full right, partial right) displaying a neutral expression, and a neutral expression displaying a direct-gaze (Fig. 1a). The four basic expressions were selected in order to best mirror the confusability of the gaze stimuli. That is, just as full and partial gaze toward a given direction is presumably harder to distinguish than gaze to an opposite direction, the expressions of disgust and anger (and surprise and fear) are more likely to be confused for each other than for other expressions.

To create the averted-gaze images we used Adobe PhotoShop CS (Adobe, CA) on an Apple iMac G4 (Apple Computer, CA) to modify the original images from the Pictures of Facial Affect.

Visual stimuli were projected onto a screen at the rear of the bore of the magnet. Subjects viewed these images via an angled mirror attached to the RF coil and placed above their eyes.

1.3. Task

Each of the eight runs contained nine blocks of 14 face images. There was one block for each of the nine conditions in every run. Seven face images were presented twice in random order in each block. The faces selected for each block (7 of a possible 11) and the order of the nine blocks were randomized across runs.

All runs began with a 32-s presentation of a fixation cross. Faces were presented for 1 s. At the conclusion of each 14-s block, a black fixation cross would appear on a white screen for 1 s immediately followed for 1 s by another face (the “test face”) with the same expression and gaze-direction as the images in preceding block (Fig. 1b). The participant’s task was to report whether the identity of the test face was the same as any of the faces in that block. The experimenter monitored responses during the acquisition periods to ensure participants were engaged in the task but were not recorded, as they were orthogonal to the perception of expression or gaze. Each of these blocks was separated by a 16-s rest period in order to allow hemodynamic activity to return to baseline levels.

1.4. Image acquisition

The blood oxygenation level-dependent (BOLD) signal was used as a measure of neural activation (Kwong et al., 1992; Ogawa, Lee, Kay, & Tank, 1990). Echo planar images (EPI) were acquired with a Siemens 3.0 Tesla Allegra head-dedicated scanner (Siemens, Erlangen, Germany) with a standard “bird-cage” head coil (TR, 2000 ms, TE, 30 ms, flip angle, 90°, matrix size, 64×64). Near whole brain coverage was achieved with 33 interleaved 3-mm axial slices. At the beginning of each scan session a high resolution anatomical image (T1-MPRAGE, TR, 2500 ms, TE, 4.3 ms, flip angle, 8°, matrix size 256×256) was acquired for use in registering activity to each subject’s anatomy and for spatially normalizing data across subjects.

1.5. Preprocessing

Data was analyzed with Analysis of Functional NeuroImages (AFNI; Cox, 1996) using standard preprocessing procedures. These procedures included 6 parameter 3D motion-correction, undistortion (Cusack, Brett, & Osswald, 2003), spatial smoothing with a 6-mm full-width-at-half-minimum Gaussian kernel and signal normalization to percent signal change from the mean.

1.6. Image analysis

Each of the eight time-series was convolved with a hemodynamic response function to create a regressor for each of the nine facial configuration conditions. Regressors of non-interest were included in the multiple regression model to factor out variance associated with mean, linear, and quadratic trends in each run as well as subject head motion. The regression model yielded coefficients that represented the signal change from the mean for each condition within each voxel.

We tested three linear contrasts. In the expression minus control contrast, all four of the expression conditions (anger, disgust, fear, and surprise) were contrasted with the control condition to find regions that were more strongly