ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

NeuroImage

NeuroImage 19 (2003) 861–869

www.elsevier.com/locate/ynimg

The human temporal lobe integrates facial form and motion: evidence from fMRI and ERP studies

Aina Puce,[a,b,1,*] Ari Syngeniotis,[b] James C. Thompson,[a,1] David F. Abbott,[b] Kylie J. Wheaton,[a] and Umberto Castiello[c,2]

[a] *Brain Sciences Institute, Swinburne University of Technology, PO Box 218, Hawthorn, Victoria 3122, Australia*
[b] *Brain Research Institute, Austin & Repatriation Medical Centre, Banksia Street, West Heidelberg, Victoria 3081, Australia*
[c] *Department of Psychology, University of Melbourne, Parkville Victoria 3052, Australia*

Received 11 October 2002; revised 30 December 2002; accepted 23 January 2003

Abstract

Physiological studies in humans and monkeys indicate that the posterior temporal cortex is active when viewing the movements of others. Here we tested the premise that this region integrates form and motion information by presenting both natural and line-drawn displays of moving faces and motion controls where motion was continuously presented in the same part of the visual field. The cortex in and near the STS and on the fusiform gyrus (FG) responded to both types of face stimuli, but not to the controls, in a functional magnetic resonance imaging study in 10 normal subjects. The response in the STS to both types of facial motion was equal in magnitude, whereas in the FG the natural image of the face produced a significantly greater response than that of the line-drawn face. In a subsequent recording session, the electrical activity of the brain was recorded in the same subjects to the same activation task. Significantly larger event-related potentials (ERPs) to both types of moving faces were observed over the posterior temporal scalp compared to the motion controls at around 200 ms postmotion onset. Taken together, these data suggest that regions of temporal cortex actively integrate form and motion information—a process largely independent of low-level visual processes such as changes in local luminance and contrast.

Introduction

Despite the popularity of mobile phones and e-mail to aid communication, face-to-face contact is still favored when unambiguity is necessary. What makes us prefer face-to-face interaction over methods of so-called "impoverished communication"? In a face-to-face interaction we can see if our message is being received from nonverbal signals sent by the other person. From these intended, and unintended, nonverbal messages we make informed judgments about another's emotional state and their future course of action.

Physiological studies in monkeys and humans suggest that the interpretation of the movements and actions of others recruit specialized neural pathways (Allison et al., 2000; Blakemore and Decety, 2001). Specifically, the perception of animate motion activates the cortex of the superior temporal sulcus (STS) in monkeys (Perrett et al., 1985; Oram and Perrett, 1994) and humans (Bonda et al., 1996; Puce et al., 1998), as well as the frontal cortex (Gallese et al., 1996; Rizzolatti et al., 1996a). In monkeys, it has been proposed that the STS response is a result of the integration of form and motion information by the cortex of the anterior superior temporal polysensory area (STPa) (Oram and Perrett, 1996). The frontal cortex (monkey F5) has been proposed to contain "mirror neurons" that respond to both observing and executing grasping movements of the hand. Human neuroimaging studies using point-light Johannson displays of human motion have shown activation in motion sensitive regions in posterior temporal/inferior parietal cortex (Bonda et al., 1996; Grossman et al., 2000). Addition-

* Corresponding author. Fax: +1-304-293-7142.
E-mail address: apuce@hsc.wvu.edu (A. Puce).
[1] Current Address: Center for Advanced Imaging, Department of Radiology, PO Box 9236, West Virginia University, Morgantown WV, 26506-9236, USA.
[2] Current Address: Department of Psychology, Royal Holloway-University of London, Egham, Surrey, TW20 0EX, United Kingdom.

1053-8119/03/$ – see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1053-8119(03)00189-7

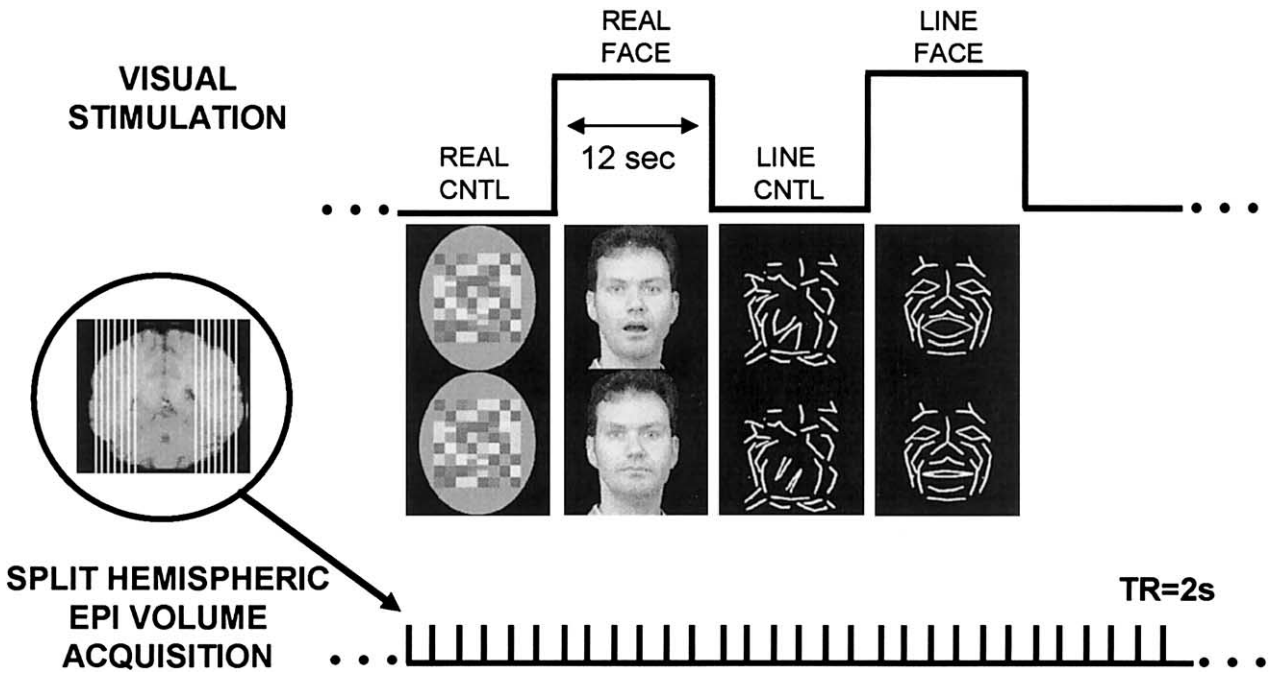

Fig. 1. Task timing and the four viewing conditions. Each viewing condition had a 12-s duration and consisted of the alternation of the two images depicted below the time line. The REAL CNTL condition consists of apparent motion of a checkerboard pattern to simulate a mouth opening and closing movement depicted in the REAL FACE condition. A similar correspondence exists for the LINE CNTL and LINE FACE conditions, where line-drawn images are alternated to display apparent motion in a control and face stimulus. In all conditions, nonlinear motion in the same part of the visual field is always depicted. The four viewing condition cycle was repeated a total of 10 times in each imaging run. Split hemispheric EPI volume acquisition at a TR of 2 was performed while the subject viewed the visual stimuli.

ally, these regions have been shown to respond to motion involving natural images of faces (Puce et al., 1998), as well as to natural images of implied motion (Kourtzi and Kanwisher et al., 2000).

What is not clear is whether the same regions in the human brain respond to both natural and Johannson-like (Johannson, 1973) displays of biological motion, that is, as does area STPa in the monkey. For this to be the case, these regions would have to integrate form and motion. If this were so, then these regions could be regarded as potential human STPa analogs. Previously, we have demonstrated that a discrete region of cortex centered on the posterior STS was active when human subjects viewed the motion of natural face images (Puce et al., 1998). Here, we investigated the functional significance of this activation by comparing brain activation to both natural and visually impoverished line drawn facial displays. We used two physiological assessment techniques. First, using high-field strength functional magnetic resonance imaging (fMRI)—a neuroimaging technique with superior spatial resolution (in millimeters)—we tested to what degree the STS was active in response to the motion types. Second, we recorded scalp event-related potentials (ERPs)—a technique with superior temporal resolution (in milliseconds)—so as to accurately determine *when* these processes occurred. We consider two possibilities that could produce the human STS activation, namely integration of visual form and motion information or activity due simply to luminance/contrast changes in the local visual field.

We studied 10 normal right-handed subjects using fMRI while they passively viewed visual displays alternating between natural images of facial motion (REAL FACE) and a respective motion control (REAL CNTL) and a line-drawn facial motion display (LINE FACE) with a respective motion control (LINE CNTL). Motion of the face consisted of mouth opening and closing movements, whereas the motion controls also had nonlinear motion components of comparable amplitudes (Fig. 1). All motion stimuli were presented in the same part of the visual field and subjects were asked to focus their gaze on the bridge of the nose in the facial stimuli and on a comparable spatial landmark in the control stimuli. All subjects participated in a second recording session in which scalp ERPs were recorded in response to exactly the same activation task, so as to characterize the time course of the neural activity elicited in response to viewing the facial motion and the motion controls.

Material and methods

Subjects

Ten neurologically normal subjects (ages: 21–46 years, mean 27.5) consented to participate as subjects in a study

approved by the Human Research Ethics Committee of Swinburne University of Technology. All subjects had either normal or corrected-to-normal vision and were strongly right-handed (LQ: mean = 82.5) as assessed by the Edinburgh Handedness Inventory (Oldfield, 1970) and were equally distributed in gender.

fMRI activation task

Subjects passively viewed a screen where stimuli subtended 4 × 4 degrees through a mirror mounted on the standard quadrature birdcage headcoil. The display was back-projected onto a screen in the scanner room while fMRI scans were performed. Four stimulus conditions were presented in an alternating block design experiment (Fig. 1) as follows

1. REAL FACE consists of a colored image of a natural face in which the mouth alternately opened and closed 7 times over a 12-s stimulus block. A single exemplar face was presented with two possible configurations, mouth open and mouth closed (depicted in Fig. 1), which were alternated in an apparent motion display (e.g., Puce et al., 1998).

2. REAL CNTL consists of a colored checkerboard pattern with similar overall color, luminance, and contrast to the natural face (Fig. 1). A group of alternating checks produced a motion stimulus in the same part of the visual field as the mouth in the REAL FACE, using identical timing for motion and block duration (e.g., Puce et al., 1998).

3. LINE FACE consists of a white line-drawn face on a black background. The mouth alternately opened and closed 7 times over a 12-s stimulus block as for the REAL FACE condition. A single exemplar face was presented with two possible configurations (mouth open and mouth closed, depicted in Fig. 1) that were alternated in an apparent motion display. The face was created from a multimarker recording of facial expressions using specialized biological motion creation software [Elite Motion Analysis System (BTS, Milan, Italy)].

4. LINE CNTL consists of a spatially "scrambled" version of the LINE FACE, with spatially rearranged features presenting with two possible configurations (depicted in Fig. 1) and identical spatial frequency and luminance/contrast as LINE FACE. The apparent motion stimulus was presented in the same part of the visual field as the mouth in LINE FACE.

Identical timing for motion and block duration was used for all conditions. Motion was presented in the same part of the visual field for all conditions. The activation task was stored on videotape. Subjects viewed the stimulus display and attempted to maintain fixation on the bridge of the nose on the face and on a similar region of space on the control.

The stimulus conditions occurred in a sequence of ABCDABCD . . . over 10 cycles totaling a viewing period of around 8 min ([ABCD] = 48 × 10 cycles = 480 s). There was also a "front-end" and "tail-end" to visual stimulation, allowing initially for steady-state magnetization to occur following commencement of scanning and subsequently for the hemodynamic response to tail-off following the completion of the imaging run. The whole experiment consisted of two imaging runs of 8 min each, with counterbalanced starting order (i.e., RUN1 = ABCD . . . and RUN2 = CDAB . . .).

Data acquisition

Functional MRI scans

Fourteen sagittal slices of fMRI data were acquired in two parasagittal blocks of 7 slices each (Fig. 1), designed to sample the STS along its entirety. This acquisition plane was chosen to minimize susceptibility artifacts in the superior temporal lobes. A series of 240 gradient echo echoplanar volumes were acquired over the 8 min stimulation period using the following parameters: TE = 40, TR = 2000, α = 40°, NEX = 1, FOV = 25 mm, matrix = 128×128 (in-plane resolution of 1.95 mm), slice thickness = 5 mm, gap = 1 mm.

Structural MRI scans

Additionally, three sets of structural images were acquired as follows:

1. A sagittal T1 series, consisting of the same 14 slices that were sampled in the fMRI sequence, were acquired to provide a structural template onto which activation images could be overlayed (TE = 14, TR = 500, NEX = 1, FOV = 25, matrix = 256 × 256, slice thickness = 5 mm, gap = 1 mm).

2. A sagittal MRA series of the same 14 slices that were sampled in the fMRI sequence were acquired so as to identify potential sources of false fMRI activation (TE = 6.9, TR = 16, NEX = 1, FOV = 25, matrix = 256 × 256, slice thickness = 5 mm, gap = 1 mm).

3. A high-resolution anatomical 3D (IR SPGR) sagittal series (TE = 1.9, TR = 9, TI = 500, NEX = 1, FOV = 25, matrix = 256 × 256 (in-plane resolution of 0.977 mm), slice thickness = 2 mm, no gap) consisting of a total of 80 images.

Data analysis

Motion detection

All fMRI data were screened for head movement (defined by a center of mass change >0.3 voxel) and other artifacts. There were no detectable head movements during the imaging runs in any of the subjects. In three subjects, spatial misregistration occurred between the two imaging runs. In these subjects, all volumes of the second imaging run were realigned to the first volume of the first imaging run using the automatic image registration (AIR) algorithm (Woods et al., 1998) running under MEDx 3.2 (Sensor systems), using the linear algorithm, rigid body (Six-parameter estimation), and trilinear interpolation approach.

False-positive "activated" voxel determination

An uncorrected statistical threshold of $P < 0.001$ was chosen based on an analysis of expected false-positive rates generated from randomly time "scrambled" EPI volumes across both imaging runs. The volumes within each hemisphere of two subjects (the two strongest activators) were scrambled in time and an unpaired *t* test was performed on the time-scrambled data. The *t* test maps were then thresholded at a number of different probability levels and the number of false-positive voxels in the brain were counted for both the positive and negative tails of the *t* test. This procedure (scrambling of volumes in time, performing unpaired *t* test, thresholding, counting false-positive voxels) was repeated a total of three times. The average false-positive voxel rate was 0.63 for $P < 0.001$. This low false-positive voxel rate was selected given the focal nature of the activations and relatively small number of voxels that made up the cluster of activation.

Statistical analysis

We sought to explicitly identify brain regions responding to both types of facial motion, that is (REAL FACE + LINE FACE) versus motion controls (REAL CNTL + LINE CNTL) by using unpaired *t* tests on the fMRI data of each hemisphere. Thresholded *t* maps ($P < 0.001$ uncorrected) showing activated voxels were overlayed on T1 anatomical images. Activated voxels were classified by anatomical location and counted using orthogonally viewed high-resolution anatomical T1 images.

Time courses of activated voxels

Time-course data were obtained from two regions of interest (ROIs) in the superior temporal sulcus (STS) and fusiform gyrus (FG). Previous studies (Puce et al., 1996; McCarthy et al., 1997) indicate that the presence of faces would activate the FG, hence this formed a control ROI in this study. An averaged single-cycle time course (over 20 cycles for the four stimulus conditions) was generated for the STS and FG in individual subjects. The time-course data were expressed as a percentage of signal change values. Group activation time courses were generated across subjects for each ROI.

Mean percentage signal change values for each condition were generated from each subject's activated voxels by calculating the mean signal change of the latter half of the viewing epoch for both the STS and FG ROIs. Repeated-measures ANOVAs were performed to determine whether there were significant differences in mean signal strength as a function of viewing condition for each of the ROIs. Next, a planned within-subjects contrast was performed using the general linear model to specifically test for differences in activation strength between the (1) REAL FACE and the LINE FACE, (2) the REAL FACE and the REAL CNTL, and (3) the LINE FACE and the LINE CNTL.

Finally, Talairach coordinates of activated voxels were calculated.

ERP study

In a subsequent recording session the 10 subjects viewed the same activation task while 64 channels of continuous EEG were recorded using a band pass of 0.1–100 Hz and a gain of 5000 with respect to the nose. An Electrocap with tin electrodes based at 10–20 system sites and at points equidistant from 10–20 sites was used. The vertical and horizontal electro-oculogram was also recorded from electrodes placed below and above the left eye and on the outer canthus of each eye. Event markers identifying motion onset for each movement had been recorded on the audio channel of the stimulus videotape. The event markers were digitized and stored with the continuous EEG file and then used to subsequently identify and extract the ERP epochs. Prior to averaging, the EEG file was screened for EOG and EMG artifacts (criterion: $\pm 50\ \mu V$). Trials with artifacts were excluded from subsequent analysis. ERPs were averaged according to stimulus type for each subject. The ERP to the first stimulus in each block was not included in the average, so as to minimize exogenous ERP components. ERP component amplitudes and latencies were measured for each subject at bilateral temporoparietal sites. A grand-average ERP was also generated for all subjects and topographical maps of voltage across the scalp were generated at time points corresponding to ERP peak latencies.

Statistical analysis

Differences in temporoparietal N170 peak amplitude and latency were evaluated using repeated-measures four-way ANOVAs for Condition (Face vs Control) × Type (Real vs Line) × Hemisphere (Left vs Right) × Electrode (T5/6, T5O1/T6O2) using SPSS for Windows Release 9.0.1 (SPSS Inc).

Results and discussion

fMRI study

Unpaired *t* tests were performed on the combined facial motion conditions (REAL FACE + LINE FACE) versus the motion controls (REAL CNTL + LINE CNTL). Activation was observed in both the posterior STS and the FG. Example of such activation in three individual subjects is illustrated in Fig. 2a. The MR signal increased during the conditions where the facial motion was presented and subsided during the corresponding control conditions in both the STS and the FG (Fig. 2b and 2c).

Overall 9 of the 10 subjects activated the STS. Of these 9 subjects, 7 activated the right STS only, 1 subject activated the left STS only, and 1 subject showed bilateral STS activation. The mean Talairach coordinates for the STS activation were (R) +44, −47, 0; (L) −30, −58, +7. Ten subjects showed FG activation, which was bilateral in 6 subjects and right-sided in 4 subjects. The mean Talairach

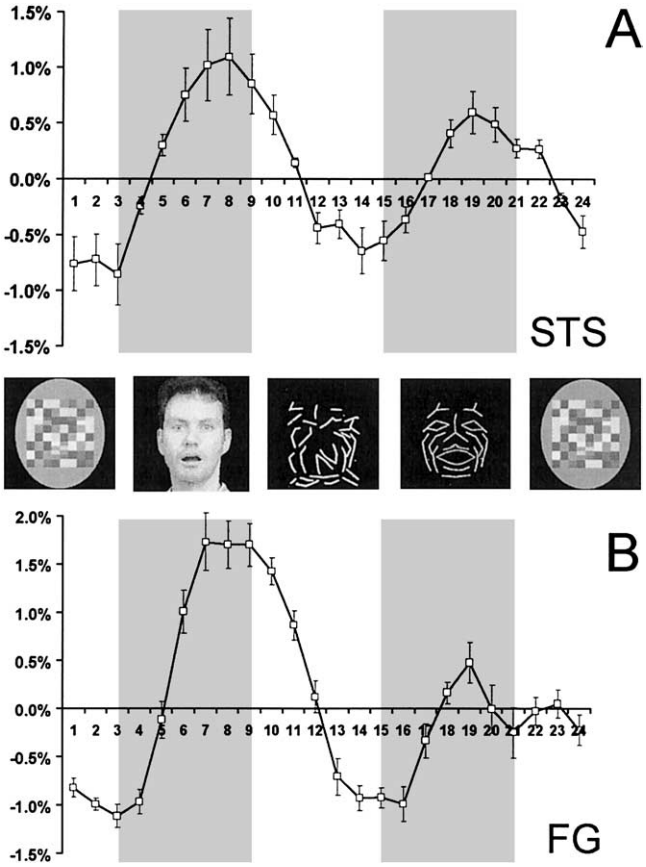

Fig. 3. Group activation time courses where voxels activating to both types of facial motion were identified. (A) STS; (B) FG. The error bars indicate SEM.

coordinates for the FG activation were (R) +33, −58, −26; (L) −33, −55, −24. The group activation time courses (Fig. 3) for the STS and the FG were generated across both hemispheres, as there were no systematic differences in response across hemispheres. The STS group activation time course shows robust activity to both facial motion conditions (Fig. 3a), whereas that in the FG appears to be strongest to the natural image of the face (Fig. 3b). These differences were verified statistically using a repeated-measures ANOVA for a main effect of Condition (REAL FACE, LINE FACE, REAL CNTL, LINE CNTL): STS [$F(3, 9) = 33.63$, $P < 0.001$] and the FG [$F(3, 12) = 57.25$, $P < 0.001$]. Importantly, when within subjects contrasts were performed across the conditions, in the STS there was no significant difference in MR signal between the two facial motion conditions [$F(1, 9) = 1.37$, $P > 0.1$]. This was in contrast to the FG, where the MR signal was significantly larger to the natural image of the face relative to the line drawn face [$F(1, 15) = 22.71$, $P < 0.001$]. Additionally, for both the STS and the FG significant differences were seen between each face condition relative to their respective controls [REAL FACE vs REAL CNTL: STS, $F(1, 9) = 41.51$, $P < 0.005$; FG, $F(1, 15) = 86.21$, $P < 0.001$; LINE FACE vs LINE CNTL: STS, $F(1, 9) = 14.66$, $P < 0.005$; FG, $F(1, 15) = 14.43$, $P < 0.005$].

ERP study

All stimulus categories elicited clear ERP activity within 250 ms of the motion transient in 8 of the 10 subjects. Technical difficulties prevented the analysis of ERP data in one subject of the two remaining subjects. In the other, clear ERPs could not be discerned in an otherwise technically adequate recording session. Hence, the data from 8 subjects were pooled and analyzed. The most prominent feature was a negative potential peaking at around 170 ms (N170; Fig. 4a) over the bilateral temporal scalp, which appeared to be larger in the right hemisphere (Fig. 4b). N170s to both types of faces were larger than those seen to control stimuli; however, control motion stimuli elicited N170s with a similar topography to that seen to facial motion. Group mean N170 amplitudes and latencies as a function of stimulus condition and electrode site are shown in Tables 1 and 2, respectively.

The repeated-measures ANOVA for N170 amplitude differences revealed a significant main effect for stimulus type for Condition (Face vs Control) [$F(1, 7) = 15.39$, $P < 0.01$]. Interestingly, there were no other significant main effects for N170 amplitude, that is, the apparent difference in N170 amplitude across hemisphere (Fig. 4b) was not significant [$F(1, 7) = 1.96$, $P > 0.05$], nor were there differences as a function of facial image type [$F(1, 7) = 1.32$, $P > 0.05$]. There was also no difference between N170 amplitude at the two electrode sites within each hemisphere [$F(1, 7) = 1.38$, $P > 0.05$]. No significant main effects were observed in an analogous repeated-measures ANOVA for N170 latency [Condition: $F(1, 7) = 0.17$, $P > 0.05$; Stimulus Type: $F(1, 7) = 0.06$, $P > 0.05$; Hemi-

Table 1
Group mean N170 peak amplitude values and their standard errors as a function of condition and electrode

STIM	T5		T5O1		T6		T6O2	
	Peak amplitude	SE	Peak amplitude	SE	Peak amplitude	SE	Peak amplitude	SE
REAL FACE	−2.09	0.54	−3.28	0.90	−3.15	0.48	−3.36	0.75
LINE FACE	−2.67	0.34	−3.02	0.68	−3.24	1.15	−3.29	0.53
REAL CNTL	−1.10	0.39	−0.99	0.40	−1.55	0.58	−1.63	0.65
LINE CNTL	−1.86	0.54	−2.42	0.61	−2.28	0.65	−2.41	0.50

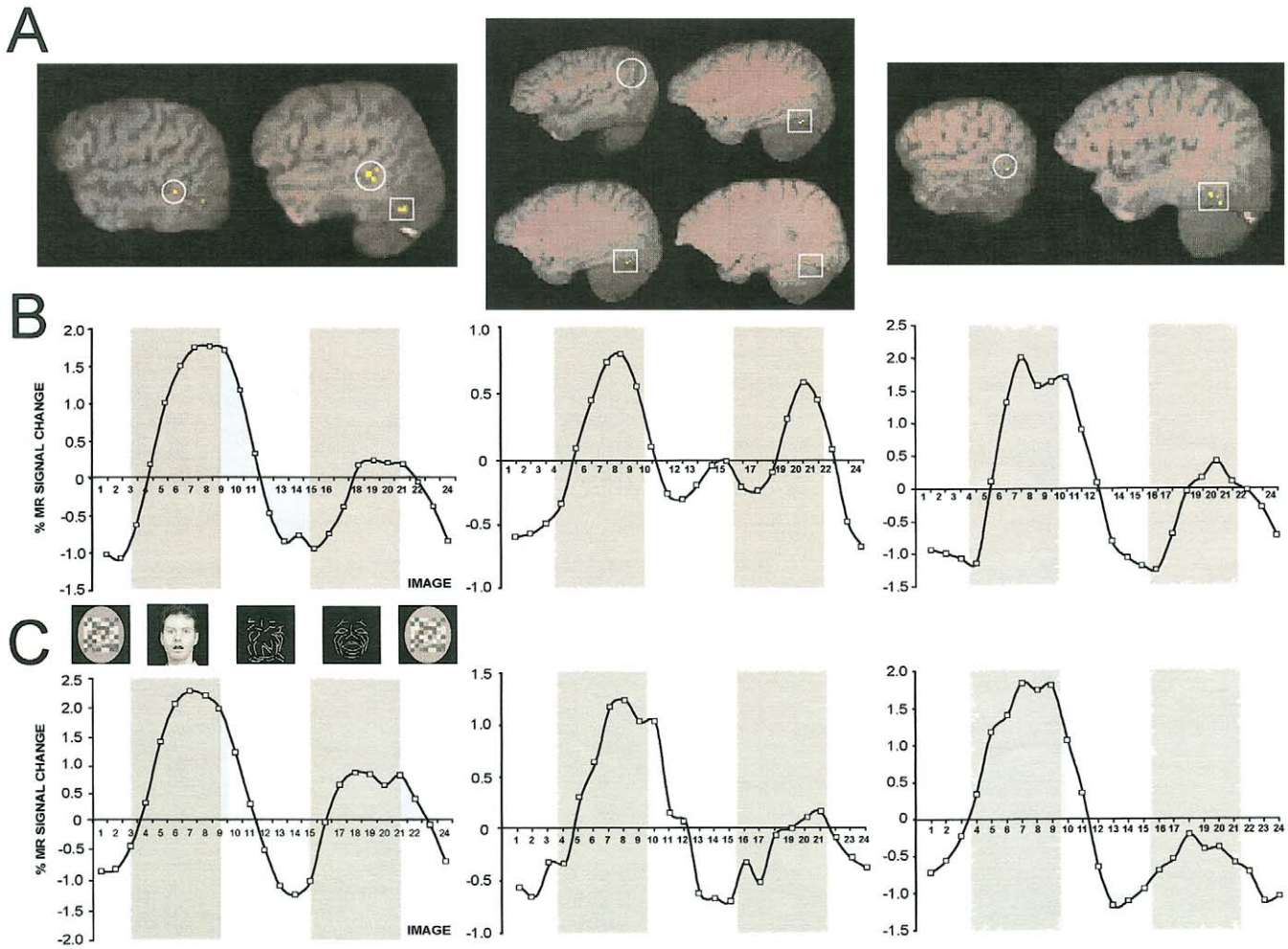

Fig. 2. Right hemisphere activation (data from three individual subjects depicted in three columns). (A) Sagittal slices showing the activation of both the right STS and the FG (yellow voxels enclosed by white circle and white square, respectively). (B) Averaged single-cycle time courses from activated STS voxels. (C) Average single-cycle time courses from activated FG voxels.

sphere: $F(1, 7) = 0.58$, $P > 0.05$; Electrode: $F(1, 7) = 1.54$, $P > 0.05$].

To summarize the major finding of this study was that human posterior temporal cortex responds to the movement from facial displays independent of whether they are natural facial images or impoverished representations devoid of standard color and shading cues. Our fMRI data indicate that the STS is the central locus for this neural activity. The robust STS response elicited to the line-drawn facial motion stimuli indicates that this region indeed integrates visual form and motion information. Color and luminance/contrast cues, however, do also contribute to the blood flow response observed in this region, as indicated by significantly larger fMRI percentage of signal changes to real face images relative to line drawn faces in the FG. Both the fMRI and ERP data suggest that bilateral temporal regions were involved; however, the extent of the neural activity was largest in the right hemisphere (Fig. 4b) and more subjects

Table 2
Group mean N170 peak latency values and their standard errors as a function of condition and electrode

STIM	T5		T5O1		T6		T6O2	
	Peak latency	SE	Peak latency	SE	Peak latency	SE	Peak latency	SE
REAL FACE	176	6	176	5	175	5	174	5
LINE FACE	167	7	168	8	179	7	174	6
REAL CNTL	171	8	171	8	168	7	167	6
LINE CNTL	182	7	181	7	167	3	168	3

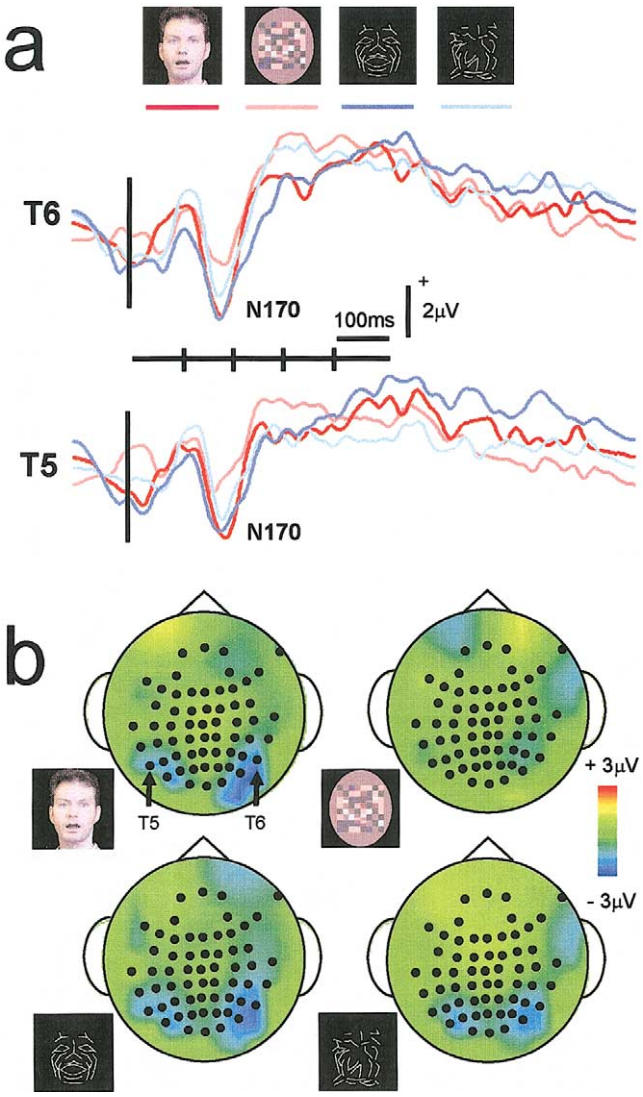

Fig. 4. Group-averaged ERP data from eight subjects. (a) ERP waveforms from left (T5) and right (T6) temporal sites as a function of stimulus type. Note that all stimulus conditions elicit an N170 ERP. N170s to both types of facial motion are larger than the N170s elicited to the motion controls. Calibration markers apply to both ERP waveforms. Vertical black bar on ERP waveforms denotes motion transition. (b) Topographic voltage maps displaying neural activity occuring at the peak of the N170, as viewed in a top-down view, for all four stimulus types. Maps to the two facial motion stimuli are shown on the left half of the panel, whereas maps to control motion stimuli appear on the right half of the panel. Faces elicit a substantial focal negativity on the posterior temporal regions, which is also seen to a lesser extent to the motion controls. A schematic nose appears at the top of the map, and the left ear is displayed on the left side. Black dots superimposed on maps denote electrode locations, as sampled by an electromagnetic electrode digitizer (Polhemus Fastrak). Calibration scale at right applies to all maps.

showed right-sided fMRI activation in the STS. Our scalp electrophysiological recordings in the same subjects indicate that this bitemporal activity peaks at around 170 ms postmotion onset.

Previous ERPs elicited to *static isolated faces* appear to have two generator sources, as indicated by intracranial recordings made from the surface of temporo-occipital cortex (e.g., Allison et al., 1999; McCarthy et al., 1999; Puce et al., 1999). One generator lies on the FG and would produce an effective dipole that would point downward into the neck. This would be consistent with a scalp positive going ERP around the vertex at around 170 ms poststimulus onset (Jeffreys, 1989) and minimal activity at the lateral temporal scalp. The second generator lies on the lateral temporo-occipital cortex (e.g., Allison et al., 1999; McCarthy et al., 1999; Puce et al., 1999) and would most likely be seen as a negative going peak at the temporal scalp (e.g., Bentin et al., 1996). Hence, the bilateral temporal N170s reported in this study would most likely be sampling activity from the lateral temporal cortex and not the FG, although both regions produced fMRI activity to the face stimuli. Since this study's focus was neural activity elicited to facial motion, we did not include ERP activity recorded to the first stimulus in each block in the overall ERP average, so eliminating any facial onset related ERP activity. This procedure would have also minimized any positive activity seen at the vertex at around the same latency. The N170s elicited to mouth motion from real face images and line-drawn faces in the present study were similar to those elicited to viewing mouth or eye movements presented on natural face images (Puce et al., 2000).

The STS activation seen in this study in response to facial motion, and not motion in general, replicates results from a previous fMRI study (Puce et al., 1998). In the previous study MT/V5 was nonspecifically activated by the motion controls, as well as facial motion. The experimental design in the present study focused specifically on the STS. Given the previous study, we assumed that activation of MT/V5 would occur to all conditions and that the statistical comparisons would highlight activity seen only to facial motion. The ERP activity indicates significant neural activity occurred to both the facial motion and control conditions at a similar latency, replicating an earlier ERP study (Puce et al., 2000). Given our previous fMRI study (Puce et al., 1998), it is likely that the ERP activity generated by the motion controls may have come from MT/V5—effectively a gyrus away from the facial motion responses in the STS. ERP (and also magnetoencephalographic) recordings typically have poor spatial resolution relative to fMRI, with sources of neural activity from two regions that are close to one another, for example, STS and MT/V5 being potentially impossible to resolve (see Watanabe et al., 2001).

To our knowledge, this is the first study examining the timing *and* location of neural activity elicited to both natural human motion and line-drawn images of human motion, and our data indicate that both motion types evoked responses in the same cortical region at around the same latency.

We do not claim that the activation seen here is necessarily specific to facial motion. Instead, this region of cortex probably responds to motion of the human body in general, as suggested by ERP studies using natural images of con-

tinuous motion (Wheaton et al., 2001) and fMRI studies using natural human motion stimuli (Beauchamp et al., 2002), Johansson-like displays (Bonda et al., 1996; Grossman et al., 2000; Grossman and Blake 2001, Vaina et al., 2001), as well as static natural images of implied body motion (Kourtzi and Kanwisher, 2000; Downing et al., 2001) (for reviews see Blakemore and Decety, 2001; Puce and Perrett, 2003). Interestingly, fMRI activation to viewing human motion relative to the motion of tools shows a dorsal/ventral segregation, with human motion preferentially activating STS cortex and the tool motion activating a region centered on the middle temporal gyrus (MTG). More importantly, when articulated versus nonarticulated human motion was viewed this dorsal/ventral STS-MTG activation gradient was also observed, suggesting that STS cortex processes complex types of human motion, where the spatial relationships between body parts is changing (Beauchamp et al., 2002). Similarly, a comparison of nonrigid versus rigid motion elicits differential activation in an anterior–posterior gradient in the STS (Grezes et al., 2001). However, the size of the activation in the STS has been shown to be modulated by task requirements and the attention that the observer places on the "human" quality of the motion (Vaina et al., 2001).

The STS, together with the orbitofrontal cortex and the amygdala, are thought to make up a cortical network responsive to social stimuli (Allison et al., 2000; Baron-Cohen, 1995; Brothers, 1997). The STS probably acts as a high-level perceptual processor, sending vital information to the other structures in the network, which enables them to evaluate and interpret affective and social information. Indeed, preliminary fMRI data indicate that activation in these regions can occur when social meaning is gleaned from stimuli that do not involve human (or animal) form (Weisberg and Martin, 2001; Castelli et al., 2000, 2002). Earlier studies in monkeys have identified neurons within the STS, in area STPa, that integrate form and motion (Oram and Perrett, 1994, 1996). For example, neurons that responded to walking motion represented by either natural images or Johansson-like displays were observed (Oram and Perrett, 1994). No response was observed to inversion of the same displays or to other types of motion. Additionally, neurons in the amygdala have also been reported to respond to complex body motion in a social context (Brothers et al., 1990).

The STS/orbitofrontal/amygdala circuit is not the only cortical network that is active during the interpretations of the actions of others. The so-called mirror system (Rizzolatti et al., 2001) of the premotor cortex was first shown to be selectively active in monkeys irrespective of whether they observed acts of grasping or performed these acts themselves (Gallese et al., 1996; Rizzolatti et al., 1996a). Parallel activation in humans has also now been observed in PET (Rizzolatti et al., 1996b), fMRI (Iacoboni et al., 1999), and MEG (Nishitani and Hari, 2000) studies to both the observation and imitation of grasping movements. The activation is usually centered close to or on Broca's area. Interestingly, scalp ERP recordings have indicated that observation of not only hand movements but also movements of the body produces neural activity from centrofrontal regions (Wheaton et al., 2001).

In sum, there appear to be multiple cortical networks in the primate brain that are specialized for the processing of the actions of others. One network, involving the STS (monkey STPa), orbitofrontal cortex, and amygdala might be biased more toward processing social and affective information from the actions of other primates. The other network, centered on the prefrontal cortex (monkey F5) might be biased for the interpretation of actions, including object manipulation of other primates and includes responses to actions such as grasping behaviors of the hand and mouth. Other behaviors, based on the interpretation of potential threats, may well activate more than one of these networks. What is still unknown is how these systems relate to one another and what if any other structures are involved in monitoring or coordinating their activity.

Acknowledgment

This study was completed with a Small Australia Research Council (ARC) grant awarded to Puce by Swinburne University.

References

Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: role of the STS region. Trends Cog. Sci. 4, 267–278.

Allison, T., Puce, A., Spencer, D.D., McCarthy, G., 1999. Electrophysiological studies of human face perception. I. Potentials generated in occipitotemporal cortex by face and non-face stimuli. Cereb. Cortex 9, 415–430.

Baron-Cohen, S., 1995. Mindblindness: An Essay on Austism and Theory of Mind. MIT Press, Cambridge, MA.

Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A., 2002. Parallel visual motion processing streams for manipulable objects and human movements. Neuron 34, 149–159.

Bentin, S., Allison, T., Puce, A., Perez, A., McCarthy, G., 1996. Electrophysiological studies of face perception in humans. J. Cog. Neurosci. 8, 551–565.

Blakemore, S.-J., Decety, J., 2001. From the perception of action to the understanding of intention. Nature Rev. Neurosci. 2, 561–567.

Bonda, E., Petrides, M., Ostry, D., Evans, A., 1996. Specific involvement of human parietal systems and the amygdala in the perception of biological motion. J. Neurosci. 16, 3737–3744.

Brothers, L., 1997. Friday's Footprint: How Society Shapes the Human Mind. Oxford University Press, New York.

Brothers, L., Ring, B., Kling, A.S., 1990. Response of neurons in the macaque amygdala to complex social stimuli. Behav. Brain Res. 41, 199–213.

Castelli, F., Frith, C., Happe, F., Frith, U., 2002. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. Brain 125, 1839–1849.

Castelli, F., Happe, F., Frith, U., Frith, C., 2000. Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage 12, 314–325.

Downing, P.E., Jiang, Y.H., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. Science 293, 2470–2473.

Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. Brain 119, 593–609.

Grezes, J., Fonlupt, P., Bertenthal, B., Delon-Martin, C., Segebarth, C., Decety, J., 2001. Does perception of biological motion rely on specific brain regions? Neuroimage 13, 775–785.

Grossman, E., Blake, R., 2001. Brain activity evoked by inverted and imagined motion. Vision Res. 41, 1475–1482.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., Blake, R., 2000. Brain areas involved in perception of biological motion. J. Cogn. Neurosci. 12, 711–720.

Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C., Rizzolatti, G., 1999. Cortical mechanisms of human imitation. Science 286, 2526–2528.

Jeffreys, D.A., 1989. A face-responsive potential recorded from the human scalp. Exp. Brain Res. 78, 193–202.

Johansson, G., 1973. Visual perception of biological motion and a model of its analysis. Percept. Psychophys. 14, 202–211.

Kourtzi, Z., Kanwisher, N., 2000. Activation in human MT/MST by static images with implied motion. J. Cogn. Neurosci. 12, 48–55.

McCarthy, G., Puce, A., Belger, A., Allison, T., 1999. Electrophysiological studies of human face perception. II. Response properties of face-specific potentials recorded from occipitotemporal cortex. Cereb. Cortex 9, 431–444.

McCarthy, G., Puce, A., Gore, J.C., Allison, T., 1997. Face-specific processing in the human fusiform gyrus. J. Cogn. Neurosci. 9, 604–609.

Nishitani, N., Hari, R., 2000. Temporal dynamics of cortical representation for action. Proc. Nat. Acad. Sci. USA 97, 913–918.

Oldfield, R.C., 1970. The assessment and analysis of handedness: The Edinburgh Inventory. Neuropsychologia 9, 97–113.

Oram, M.W., Perrett, D.I., 1994. Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. J. Cogn. Neurosci. 6, 99–116.

Oram, M.W., Perrett, D.I., 1996. Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. J. Neurophysiol. 76, 109–129.

Perrett, D.I., Smith, P.A.J., Mistlin, A.J., Chitty, A.J., Head, A.S., Potter, D.D., Broennimann, R., Milner, A.D., Jeeves, M.A., 1985. Visual analysis of body movements by neurones in the temporal cortex of the macaque monkey: a preliminary report. Behav. Brain Res. 16, 153–170.

Puce, A., Allison, T., Asgari, M., Gore, J.C., McCarthy, G., 1996. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. J. Neurosci. 16, 5205–5215.

Puce, A., Allison, T., Bentin, S., Gore, J.C., McCarthy, G., 1998. Temporal cortex activation in humans viewing eye and mouth movements. J. Neurosci. 18, 2188–2199.

Puce, A., Allison, T., McCarthy, G., 1999. Electrophysiological studies of human face perception. III. Effects of top-down processing on face-specific potentials. Cereb. Cortex 9, 445–458.

Puce, A., Perrett, D., 2003. Electrophysiology and brain imaging of biological motion. Phil. Trans. Roy. Soc. Lond. B. 358, 435–445.

Puce, A., Smith, A., Allison, T., 2000. ERPs evoked by viewing facial movements. Cogn. Neuropsychol. 17, 221–239.

Rizzolatti, G., Fadiga, L., Fogassi, L., Gallese, V., 1996a. Premotor cortex and the recognition of motor actions. Brain Res. Cogn. Brain Res. 3, 131–141.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., Fazio, F., 1996b. Localization of grasp representations in humans by PET. 1. Observation versus execution. Exp. Brain Res. 111, 246–252.

Rizzolatti, G., Fogassi, L., Gallese, V., 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. Nat. Rev. Neurosci. 2, 661–670.

Vaina, L.M., Solomon, J., Chowdhury, S., Sinha, P., Belliveau, J.W., 2001. Functional neuroanatomy of biological motion perception in humans. Proc. Nat. Acad. Sci. USA 98, 11656–11661.

Watanabe, S., Kakigi, R., Puce, A., 2001. Occipitotemporal activity elicited by viewing eye movements: a magnetoencephalographic study. Neuroimage 13, 351–363.

Weisberg, J., Martin, A., 2001. Different patterns of neural activity associated with social and mechanical knowledge. Neuroimage 13 (6), S485.

Wheaton, K., Pipingas, A., Silberstein, R., Puce, A., 2001. Neuronal responses elicited to viewing the actions of others. Vis. Neurosci. 18, 401–406.

Woods, R.P., Grafton, S.T., Watson, J.D.G., Sicotte, N.L., Mazziotta, J.C., 1998. Automated image registration. II. Intersubject validation of linear and nonlinear models. J. Comput. Assist. Tomogr. 22, 153–165.