Research report

Attention to emotion modulates fMRI activity in human right superior temporal sulcus

Jin Narumoto\textsuperscript{a,}*\textsuperscript{a,}, Tomohisa Okada\textsuperscript{b}, Norihiro Sadato\textsuperscript{b}, Kenji Fukui\textsuperscript{a}, Yoshiharu Yonekura\textsuperscript{c}

\textsuperscript{a}Department of Psychiatry, Kyoto Prefectural University of Medicine, 465 Kaji-cho, Kawaramachi-Hirokoji, Kamigyo-ku, Kyoto 602-8566, Japan
\textsuperscript{b}Psychophysiology Section, Department of Cerebral Research, National Institute for Physiological Sciences, Aichi 444-8585, Japan
\textsuperscript{c}Biomedical Imaging Research Center, Fukui Medical University, Fukui 910-1193, Japan

Accepted 27 March 2001

Abstract

A parallel neural network has been proposed for processing various types of information conveyed by faces including emotion. Using functional magnetic resonance imaging (fMRI), we tested the effect of the explicit attention to the emotional expression of the faces on the neuronal activity of the face-responsive regions. Delayed match to sample procedure was adopted. Subjects were required to match the visually presented pictures with regard to the contour of the face pictures, facial identity, and emotional expressions by valence (happy and fearful expressions) and arousal (fearful and sad expressions). Contour matching of the non-face scrambled pictures was used as a control condition. The face-responsive regions that responded more to faces than to non-face stimuli were the bilateral lateral fusiform gyrus (LFG), the right superior temporal sulcus (STS), and the bilateral intraparietal sulcus (IPS). In these regions, general attention to the face enhanced the activities of the bilateral LFG, the right STS, and the left IPS compared with attention to the contour of the facial image. Selective attention to facial emotion specifically enhanced the activity of the right STS compared with attention to the face per se. The results suggest that the right STS region plays a special role in facial emotion recognition within distributed face-processing systems. This finding may support the notion that the STS is involved in social perception. © 2001 Elsevier Science B.V. All rights reserved.

Theme: Sensory systems

Topic: Visual psychophysics and behavior

Keywords: Emotion; Face; Attentional modulation; Superior temporal sulcus

1. Introduction

The face conveys various types of information, such as identity, emotion, intention (of another person), and gender. Bruce and Young [8] proposed that the different attributes of the face are processed by a parallel information processing system. Further, several lines of evidence suggest that there are distinct neural substrates for processing different aspects of the face [19]. In particular, the cerebral cortex in and near the superior temporal sulcus (STS) has been well studied as an important structure for the processing of information relevant to interpersonal communication, such as eye gaze, direction, expression, and mouth movement [2]. A body of evidence suggests that the STS region is involved in the perception of eye gaze direction [31,34,35,45]. Recently, a neuroimaging study [21] showed that the STS region has distinct neural representation of eye gaze direction within the distributed face-responsive areas.

Similarly, several lines of evidence have suggested that expression, which is another important social signal, is also processed in the STS. Single-cell recordings in monkeys [18] and humans [30] showed that there are independent cell populations in the STS that respond specifically to expression. Moreover, an electrical stimulation study in humans [15] showed that stimulation of the right STS disturbed labeling of facial emotions. A recent magnetoencephalography study [40] also showed an expression-specific response in the STS. Although these previous studies suggested that there might be a distinct neural
representation for emotional expression in the temporal neocortex, neuroimaging studies mainly focused on the subcortical structures [7,27,28,33,44]. Furthermore, the reported activation of the STS during facial emotion recognition [3,11,17,32,39] was anterior extension of diffusely activated visual association cortices including fusiform face areas. These findings indicate that the STS is a part of the face-responsive areas, but it is still inconclusive whether the STS region plays a special role in processing the emotional expression.

In the present study, we sought to determine whether selective attention to facial emotion specifically enhanced activity in the STS. We hypothesized that selective attention to facial emotion, compared with attention to other aspects of the face, should elicit a stronger response of the STS than of other face-responsive areas. To test this, we adopted a delayed match to sample procedure with face or non-face samples. First face-responsive areas were functionally defined by comparing face tasks with a non-face one. Within these areas, tested was the effect of explicit attention to the characteristics of the face such as identity and emotional expressions. Considering the result of multidimensional scaling of emotional facial expressions [38], happy and fearful expressions were used in a valence discrimination task, and sad and fearful expressions were used in an arousal discrimination task.

2. Materials and methods

2.1. Subjects

Twelve right-handed healthy volunteers (nine men and three women), aged 19–35 years, with no history of neurological or psychiatric illness, participated in the study. The study was conducted at the Biomedical Imaging Research Center, Fukui Medical University, where the protocol was approved by the ethics committee. All subjects gave their written informed consent for the study. One man’s data were excluded from the analysis owing to task performance error.

2.2. Stimuli

We used 18 facial expressions from Ekman and Friesen [13]. They consisted of six faces of Caucasian adults (an equal number of men and women) expressing fear, happiness, and sadness. A cross-cultural study [14] has shown that Japanese persons can judge the facial emotions of Caucasian persons with a high level of agreement when they select a single verbal label for each expression. The faces were framed with a rectangular or circular contour.

2.3. Experimental tasks

Subjects performed four different delayed matching-to-sample tasks: a contour-matching task (CO), a face-matching task (FA), an emotion-matching task with happy and fearful expressions (HF), and an emotion-matching task with sad and fearful expressions (SF). For each task, a single sample face was presented for 400 ms, followed by two choice faces presented side by side for 1000 ms (Fig. 1). Trials were presented every 3200 ms. Subjects indicated which selectively attended aspect of the choice stimuli matched that of the sample stimulus by pressing the left or right button of a response box. Each of the four face tasks was administered in a separate imaging series. To control motor response and visual input, we used a non-face task as a control task, in which scrambled images with a rectangular or circular contour were presented in the same way, and subjects matched the stimuli with respect to the contour of the images. Four repetitions of a rest block, a non-face task block, and a face task block were presented to the subject (R-N-F-R-N-F-R-N-F-R-N-F). Each block contained 10 trials of matching tasks, resulting in a total of 40 trials per imaging series. During the rest periods, subjects gazed at a cross (+) on the screen. The presentation order of the four series was counterbalanced across subjects in pseudorandom order. Stimuli were

---

Fig. 1. Design of sample trials in the delayed matching-to-sample task used in the fMRI study. Subjects attended selectively to the contour of the face, the face per se, or to the emotional expression, and, in the control task, to the contour of scrambled images. S, stimulus; ISI, interstimulus interval.
presented on a rear projection screen placed at the foot of the scanner bed and viewed by the subjects through a mirror.

There were 36 samples of three men and three women, with three different emotional expressions (happy, sad and fearful) with different contours (rectangular and circle). In CO condition, 40 trials were performed using all 36 samples (and hence four samples were presented twice). Subjects were asked to choose the same contour. An identical person with or without the same expression was not used as a choice. In FA condition, 40 trials were performed using all 36 samples. Subjects were asked to choose the identical person with the same expression. An identical person with a different expression was not used as a choice. Within the samples and choices 1 and 2, the contour and expression were pseudorandomized. In SF condition, 40 trials were performed using 24 samples, 16 of which were presented twice. Subjects were required to choose the same expression. An identical person with the same expression was not used as a choice. Within the samples and choices 1 and 2, the person and contour were pseudorandomized. The same procedure was adopted in HF.

Happy and fearful expressions presented in HF condition were different in regard to valence (positive versus negative), and sad and fearful expressions presented in SF condition were different in regard to arousal [38]. Then, in the HF and SF conditions, subjects were required to discriminate emotion by valence and arousal, respectively.

2.4. Image acquisition

Functional magnetic resonance imaging (fMRI) was performed with a GE Signa 3T MRI system (General Electric Medical Systems, Milwaukie, WI) equipped with a standard birdcage head coil. A time-course series of 102 volumes was acquired with T2*-weighted, gradient echo, echo planar imaging (EPI) sequences. Each volume consisted of 42 slices, and the slice thickness was 2.7 mm with a 0.3-mm gap to cover the entire cerebral and cerebellar cortex. The interval between two successive acquisitions of the same image was 4000 ms, echo time was 30 ms, and flip angle was 90°. The field of view was 22 cm. The digital in-plane resolution was 64×64 pixels with a pixel dimension of 3.44×3.44 mm. Head motion was minimized by placing comfortably tight foam padding around the subject's head.

2.5. Image analysis

The image data were analyzed with statistical parametric mapping (SPM99 software from the Wellcome Department of Cognitive Neurology, London, UK http://www.fil.ion.ucl.ac.uk/spm) implemented in Matlab (Mathworks Inc., Sherborn, MA). The first six volumes of each fMRI run (prescan period) were discarded because magnetization was unsteady, and the remaining 96 volumes were used for the statistical analysis. Images were realigned, normalized to the standard stereotaxic space [41], and smoothed with an isotropic three-dimensional Gaussian filter of 10 mm.

Using group analysis according to a random effect model [16], we identified regions that showed significant responses to faces, as compared with non-face stimuli, as face-responsive areas. Thus, these regions of activation were defined independently of response differences between attention to face and attention to contour or between attention to face and attention to emotion. The group analysis consisted of two levels. In the first level, the signal time course of each subject was modeled with a delayed box-car function convolved with a hemodynamic response function in the context of a general linear model. One contrast image per subject was created by contrasting the four face task conditions with the non-face condition. In the second level, these images were entered into a one-sample t-test. The statistical threshold was set at P<0.05 corrected for multiple comparisons at the cluster level. Subsequently, we determined the coordinates of the face-responsive areas in each subject by exploring the coordinates that showed maximum response for the face tasks within the spherical region (radius 10 mm) around the maximum coordinates of the group analysis. Finally, for each subject, we measured the size of the response to each face task as the percentage change in relation to the response to the non-face task by using multiple regression in the functionally defined face-responsive areas.

In all areas, measures of the response size were analyzed with a repeated-measures analysis of variance (ANOVA) with planned comparisons for the simple differences between the CO task and the other face tasks (FA, HF, and SF), between the FA task and the HF and SF (emotion) tasks, and between the emotion tasks. The first comparison (explicit face processing, as in the FA, HF, and SF tasks, vs. implicit face processing, as in the CO task) was expected to show attentional modulation. The second comparison (selective attention to emotion, as in the HF and SF tasks, vs. attention to the face per se, as in the FA task) was expected to show emotional modulation. Finally, the third comparison (valence discrimination, as in the HF task, vs. arousal discrimination, as in the SF task) was expected to show the difference in emotional modulation between attention to valence and attention to arousal.

3. Results

3.1. Behavioral data

The mean (±S.D.) percentage of correct responses was 96±2.6% for the CO task, 97±3.1% for the FA task, 92±6.2% for the HF task, and 84±9.4% for the SF task. Repeated-measures ANOVA showed that the accuracies
Table 1
Coordinates of the brain regions significantly activated by viewing faces (four face tasks) as compared with viewing scrambled pictures (non-face task) in the group level

<table>
<thead>
<tr>
<th>Region</th>
<th>Hemisphere</th>
<th>Talairach coordinates (mm)</th>
<th>Z score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>Lateral fusiform gyrus</td>
<td>Left</td>
<td>-42</td>
<td>-58</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>46</td>
<td>-52</td>
</tr>
<tr>
<td>Superior temporal sulcus</td>
<td>Right</td>
<td>52</td>
<td>-48</td>
</tr>
<tr>
<td>Intraparietal sulcus</td>
<td>Left</td>
<td>-28</td>
<td>-58</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>-40</td>
<td>-52</td>
</tr>
</tbody>
</table>

Coordinates are those of the stereotaxic brain atlas of Talairach and Tourneaux [41].

were significantly greater in the CO task than in the other face tasks \((F(1,10)=11.2, P<0.007)\), in the FA task than in the emotion tasks \((F(1,10)=19.8, P<0.001)\), and in the HF task than the SF task \((F(1,10)=5.2, P<0.046)\). Subjects performed at ceiling level during the non-face task (>99% correct).

3.2. Neuroimaging data

Five regions responded more to viewing faces than to viewing scrambled pictures: the bilateral lateral fusiform gyrus (LFG), the right superior temporal sulcus (STS), and the bilateral intraparietal sulcus (IPS) at the population level (Table 1 and Fig. 2A). The averaged coordinates of each region are shown in Table 2. Although no significant activation was found in the left STS, the coordinate of the left STS of each subject was also determined by using the mirror coordinate \((x=-52, y=-48, z=12)\) of the right STS \((x=52, y=-48, z=12)\) as a center. The locations of these areas are consistent with the recent neuroimaging study [21], in which the face-responsive regions were defined functionally.

Table 3 and Fig. 2B show the percentage signal change during the face tasks as compared with the non-face task at each region. Explicit face processing elicited a stronger response than implicit processing in the bilateral LFG \((F(1,10)=17.98, P<0.002)\) on the left; \((F(1,10)=41.04, P<0.001)\) on the right), the right STS \((F(1,10)=7.58, P<0.02)\), and the left IPS \((F(1,10)=12.53, P<0.005)\). By contrast, selective attention to emotion elicited a stronger response than attention to the face per se only in the right STS \((F(1,10)=9.58, P<0.011)\), as we predicted. However, contrary to our prediction, the effect of task failed to show the laterality \((F(1,20)=0.55, P<0.05)\). Finally, in all regions, there was no significant difference in response between the emotion conditions.

4. Discussion

We found that attention to face characteristics diffusely activate the face-responsive areas whereas the right STS was the only area which was activated by the attention to the facial emotion. The results indicate that the STS plays a unique role in processing facial emotion within the distributed human neural system for face perception. This finding also supports the notion that the STS is involved in social perception from visual cues [2]. In addition, our results show that explicit processing of the face enhances activities in the face-responsive areas. This phenomenon may reflect attentional modulation [10,20,22,23].

In monkeys, neurons in the STS respond differentially to facial expressions in static pictures [18]. In humans, neuronal activity in the STS shows significant changes during labeling of facial expression [30]. In addition, electrical stimulation of the STS prevents facial emotion recognition [15]. Lesion studies [1,6,37] also suggest that the region is involved in facial emotion recognition. Thus, the present results are concordant with the findings of these earlier studies.

Previous neuroimaging studies [3,17,32,39] have also revealed activation in the STS during facial emotion processing. However, some of these studies focused not on the attentional aspect but on the effect of expression per se [3,32]. Furthermore, they did not show specific activation in the STS, rather, they showed non-specific activations over the visual association cortices. These activations may reflect combined effect of emotional information and cognitive attention. Unlike these studies, we distinguished emotional modulation from attentional modulation. Thus, we could delineate the specific activation in the STS during explicit processing of emotional expression. A recent neuroimaging study [11], which examined neuro-anatomical dissociation between explicit and implicit processing of emotion, also demonstrated that explicit processing elicited a stronger response in the middle temporal gyri than implicit processing. The result, together with ours, suggests that the visual association cortex around the STS is preferentially recruited when emotional expressions are explicitly processed.

In addition to providing the first evidence for specific involvement of the right STS in facial emotion processing in face-responsive areas, the present results are consistent with the notion that the STS is involved in social perception [2]. The STS is also involved in processing other facial attributes, such as eye gaze direction [21,45] and mouth movement [35], and, therefore, may be responsible for processing the changeable aspects of faces, which is
Fig. 2. (A) Regions of the brain significantly activated during the face-perception tasks. Activations are superimposed on a T1-weighted template. The coronal image is shown at y=-58. Statistical threshold was set at P<0.05 corrected for multiple comparisons. (B) Mean percentage change in signal responses (bar, S.E.) in the face-responsive areas. CO, contour task; FA, face task; HF, emotion-matching task with happy and fearful expressions; SF, emotion-matching task with sad and fearful expressions. *P<0.05, **P<0.01.

Table 2
Averaged coordinates of the brain regions significantly activated by viewing faces as compared with viewing scrambled pictures in the subject (mean±S.D.)

<table>
<thead>
<tr>
<th>Region</th>
<th>Hemisphere</th>
<th>Talairach coordinates (mm)</th>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lateral fusiform gyrus</td>
<td>Left</td>
<td>-43±3</td>
<td>-57±6</td>
<td>-21±3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>46±3</td>
<td>-51±6</td>
<td>-20±3</td>
<td></td>
</tr>
<tr>
<td>Superior temporal sulcus</td>
<td>Left</td>
<td>-55±5</td>
<td>-52±5</td>
<td>10±5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>55±5</td>
<td>-50±5</td>
<td>12±3</td>
<td></td>
</tr>
<tr>
<td>Intraparietal sulcus</td>
<td>Left</td>
<td>-30±4</td>
<td>-59±6</td>
<td>45±6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>38±6</td>
<td>-52±5</td>
<td>45±4</td>
<td></td>
</tr>
</tbody>
</table>

Coordinates are those of the stereotaxic brain atlas of Talairach and Tourneaux [41].

important for interpersonal communication. In line with this notion, activation of the STS is found close to V5/MT, which respond to moving visual stimuli [43,46] and implied motion [24]. This may also reflect the involvement of the region in processing the dynamic changeable aspects of faces. The location of an expression-sensitive area, where connections of the dorsal and ventral visual pathways converge [42], may be advantageous to rapid processing of expressions.

Furthermore, we found the significant emotional modulation in the right STS but not in the left. This is consistent with previous clinical [1,4,6,12] and experimental [5,9,25,29] studies, which suggest right hemisphere dominance in facial emotion processing. However, the interaction between the laterality and the modulation failed to
Table 3
Percentage signal change at each brain region during the face tasks as compared with the control task (mean±S.E.)

<table>
<thead>
<tr>
<th>Region</th>
<th>Hemisphere</th>
<th>CO</th>
<th>FA</th>
<th>HF</th>
<th>SF</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFG</td>
<td>Left</td>
<td>0.77±0.10</td>
<td>1.1±0.12</td>
<td>1.1±0.11</td>
<td>1.0±0.13</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>0.88±0.10</td>
<td>1.1±0.12</td>
<td>1.3±0.13</td>
<td>1.2±0.11</td>
</tr>
<tr>
<td>STS</td>
<td>Left</td>
<td>0.31±0.06</td>
<td>0.26±0.08</td>
<td>0.40±0.07</td>
<td>0.36±0.09</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>0.27±0.05</td>
<td>0.32±0.04</td>
<td>0.50±0.07</td>
<td>0.50±0.06</td>
</tr>
<tr>
<td>IPS</td>
<td>Left</td>
<td>0.38±0.06</td>
<td>0.72±0.08</td>
<td>0.81±0.13</td>
<td>0.82±0.09</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>0.58±0.07</td>
<td>0.75±0.06</td>
<td>0.76±0.12</td>
<td>0.80±0.09</td>
</tr>
</tbody>
</table>

Regions: LFG, lateral fusiform gyrus; STS, superior temporal sulcus; IPS, intraparietal sulcus. Face tasks: CO, contour task; FA, face task; HF, emotion task with happy and fearful face; SF, emotion task with sad and fearful face.

reach significant level. The left STS might be inappropriate to examine the emotional effect because it did not determined functionally. Alternatively, the small sample size of the present study might be a factor.

In addition to the STS and the LFG, we found significant activation in the IPS during face perception tasks. Moreover, activity in the left IPS was stronger in the explicit tasks than in the implicit task. According to Haxby et al. [19], activity in the IPS may be specifically associated with the spatial aspects of perceived eye gaze. However, the explicit tasks used in our study did not need selective attention to gaze direction. Then, the activity might reflect eye movements during the tasks and/or attention to spatial configuration of facial components.

Finally, there are some limitations in this study. First, it is possible that the differences in difficulty level associated with the identity and expression matching confound the result. As the percentage of correct response was significantly lower in the expression tasks than in the identity task, subjects might be required more concentration and motivation on the expression tasks. However, these factors might affect the responses of the face-responsive areas globally rather than locally. Hence the specific activation of the STS cannot be attributed to the difference of the task difficulties. Second, since the present study used a blocked design, we cannot examine time courses associated with individual events. As described by Puce et al. [36], in the fusiform face area, the top-down modulation was found in late components of event-related potential (ERP), specifically P290 and N700. Therefore, the modulation in the right STS observed in the present study might also occur in the late period. However, to address the issue, an event-related design and a more streamlined task would be more appropriate. Third, in the comparison of the emotional conditions with the FA condition (SF+HF vs. FA), implicit effect of fearful faces cannot be completely excluded because of more frequent presentation of fearful faces (one half of the all presentations) in the emotion conditions than in the FA condition (one third). However, previous studies reported that fearful expression per se activated the amygdala and insula, but not the cortical regions including STS [27,44]. Thus, we interpret the modulation in the right STS as a reflection of explicit attention to emotional expressions rather than that of implicit effect of fearful faces.

4.1. Conclusions

The finding of specific enhancement in the right STS while attending to emotion is consistent with the results of previous studies and supports the idea that the STS is involved in social perception. More generally, the present results suggest that human face-responsive areas are separately tuned for different aspects of facial information.

Acknowledgements

The authors wish to thank B.J. Hessle, E.L.S., for skillful editing. This study was supported by a research grant (JSPS-RFTF97L00203) for the ‘Research for the Future’ Program from the Japan Society for the Promotion of Science.

References


