OXFORD

Functional organization of face processing in the human superior temporal sulcus: a 7T high-resolution fMRI study

Anne-Kathrin Schobert,[1,2] Corrado Corradi-Dell'Acqua,[1,2,3]
Sascha Frühholz,[2,4,5,6] Wietske van der Zwaag,[7,8] and Patrik Vuilleumier[1,2]

[1]Department of Neurosciences and Clinic of Neurology, University Medical Center, CH-1211 Geneva, Switzerland, [2]Swiss Centre for Affective Sciences, [3]Department of Psychology, FPSE, University of Geneva, CH-1211 Geneva, Switzerland, [4]Department of Psychology, University of Zurich, Zurich, Switzerland, [5]Neuroscience Center Zurich, Zurich, Switzerland, [6]Zurich Center for Integrative Human Physiology (ZIHP), Zurich, Switzerland, [7]CIBM, EPFL, Lausanne, VD, Switzerland, and [8]Spinoza Centre for Neuroimaging, Amsterdam, The Netherlands

Correspondence should be addressed to Patrik Vuilleumier, Laboratory for Neurology and Imaging of Cognition, Department of Neurosciences, 1 rue Michel-Servet, 1211 Geneva, Switzerland. E-mail: patrik.vuilleumier@unige.ch

Anne-Kathrin Schobert and Corrado Corradi-Dell'Acqua contributed equally to this study.

Abstract

The superior temporal sulcus (STS) is a major component of the human face perception network, implicated in processing dynamic changeable aspects of faces. However, it remains unknown whether STS holds functionally segregated subdivisions for different categories of facial movements. We used high-resolution functional magnetic resonance imaging (fMRI) at 7T in 16 volunteers to compare STS activation with faces displaying angry or happy expressions, eye-gaze shifts and lip-speech movements. Combining univariate and multivariate analyses, we show a systematic topological organization within STS, with gaze-related activity predominating in the most posterior and superior sector, speech-related activity in the anterior sector and emotional expressions represented in the intermediate middle STS. Right STS appeared to hold a finer functional segregation between all four types of facial movements, and best discriminative abilities within the face-selective posterior STS (pSTS). Conversely, left STS showed greater overlap between conditions, with a lack of distinction between mouth movements associated to speech or happy expression and better discriminative abilities (for gaze and speech *vs* emotion conditions) outside pSTS. Differential sensitivity to upper (eye) or lower (mouth) facial features may contribute to, but does not appear to fully account for, these response patterns.

Key words: face processing; high-resolution fMRI; superior temporal sulcus

Introduction

The superior temporal sulcus (STS) of the human brain is engaged by various kinds of emotional and social (Haxby *et al.*, 2000) information. Its posterior part (pSTS) is a major component of 'core' networks subserving face perception. It is consistently modulated by not only facial expressions of emotion (Kujala *et al.*, 2009) but also eye-gaze shifts (Calder *et al.*, 2007) and speech-lip reading (Calvert *et al.*, 1997). It is therefore

assumed to process changeable aspects of faces with similar meaning across individuals, unlike more ventral extrastriate visual cortex such as the fusiform face area (FFA) and occipital face area (OFA), which process facial shape information that is invariant across movement and thus crucial for identity recognition (Gobbini and Haxby, 2007). Moreover, pSTS and neighboring cortical areas also respond to biological motion displayed by limb actions or whole-body movements, e.g. walking (Pelphrey *et al.*, 2003).

Despite the many conditions activating STS (Hein and Knight, 2008), little is known about the exact functional organization of this region, in particular whether responses to different kinds of information recruit the same cortical modules or distinct subregions implicating different neuronal populations. Most imaging studies examined different stimulus types in isolation, e.g. gaze direction only. The role of pSTS in processing changeable or socially significant information from faces has generally been considered in a global fashion (Haxby *et al.*, 2002), without seeking to define any systematic anatomical organization. However, although responses to different kinds of facial cues appear to overlap in pSTS across studies, some degree of non-overlap may actually exist and reflect a regional segregation of different functions along the whole-STS region. A few pioneer studies have assessed STS activation to multiple types of stimuli. Pelphrey *et al.* (2005) compared responses with eye, mouth and hand movements and suggested both overlapping and non-overlapping activations across these conditions. However, this work did not allow determining whether regional variations within and around STS were due to differences in the visual categories of stimuli (e.g. mouth and eyes are perceptually different body parts) or in their functional significance (e.g. mouth/eye movements have different communication purposes).Furthermore, responses to facial emotional expressions were not examined, despite the key role of pSTS in face processing. Engell and Haxby (2007) compared activation with averted gaze and emotional facial expressions in right STS and found distinct (but partially overlapping) clusters, with a posterior–anterior division for activations to averted gaze and expression, respectively. Finally, a recent study by Deen *et al.* (2015) also found a posterior-to-anterior organization when comparing responses to a multitude of socially relevant paradigms, with the posterior portion mostly implicated in theory of mind and biological motion, the middle portion in face and voice perception and the anterior portion in language (see also Allison *et al.*, 2000, for an earlier review of STS activation sites). Taken together, previous studies converge in suggesting a sensitivity of STS to social, affective and general visual features of faces, with a posterior-to-anterior organization noted across distinct stimulus classes. However, the heterogeneity of stimuli and paradigms employed, compounded with the low spatial resolution of traditional neuroimaging approaches, leave open the question about the exact nature of this organization and the degree of functional segregation within the face-processing domain. To the best of our knowledge, no study has investigated with fine spatial detail how different portions of STS (or more face-selective pSTS) respond to specific facial movements, carefully controlled for anatomical localization (eye *vs* mouth) and for functional significance (emotional *vs* non-emotional expression).

Here, we exploited the high resolution of 7T functional magnetic resonance imaging (fMRI) to identify the precise topological organization of human STS for different facial features. We predicted the existence of distinct subregions responding to gaze shifts, lip speech and emotion expression, respectively.

Based on previous studies and anatomical proximity with potential projection sites, we hypothesized that activation to gaze may reside in more posterior–superior STS which is connected with parietal areas controlling spatial attention (Hoffman and Haxby, 2000; Vuilleumier, 2002; Ethofer *et al.*, 2011), possibly through the superior longitudinal fasciculus (Gschwind *et al.*, 2012); whereas lip movements should recruit more anterior STS subregions located near auditory areas involved in voice and language processing (Calvert *et al.*, 1997). On the other hand, emotional expressions might recruit distinct subparts associated with medial temporal lobe systems mediating social cognition and memory (LeDoux, 2000; Adolphs, 2001). Because functional specialization might arise from selectivity to different visual features in faces, we used emotion expression involving predominant changes in the mouth (happiness) or eyes (anger) that could be compared with non-emotional cues in the same face parts (speech and gaze, respectively). Finally, because previous research suggests that dynamic facial features are processed not only in pSTS but also its more anterior portions (Wright *et al.*, 2003; Pelphrey *et al.*, 2005), our analysis involved the whole extent of STS (anatomically defined) in addition to face-selective regions (functionally defined by a separate face localizer).

Materials and methods

Participants

We recruited 16 healthy volunteers with normal vision (8 females, mean age 24.5 years, range 19–39 years, all right handed). They provided written informed consent and were paid for their participation. The protocol was approved by the local ethics committee. Each participant underwent 2 runs of the main experiment, 1 face localizer run, 1 anatomical scan as well as 3 runs of an auditory experiment not reported here.

Stimuli and design in localizer experiment

To identify face-selective areas in STS (pSTS) as defined in previous research (Ishai *et al.*, 2005), we acquired a separate face localizer scan (4.5 min) in which participants saw 4 blocks of faces and 4 blocks of houses (gray scale photographs, all ~5° visual angle), plus scrambled versions of these. Blocks lasted 22 s and contained 18 images, each presented for 750 ms and separated by a blank screen of 500 ms. Participants were asked to detect an immediate repetition of the same image (20 times throughout the run), across all image conditions (face, house or scramble). All faces used in this localizer task were photographs of static, neutral and realistic faces with direct gaze and closed mouth.

Stimuli and design in main experiment

For the main experiment, we created different video-clips of facial movements (Figure 1) using the FACSGEN software developed in our center and used in previous studies (Cristinzio *et al.*, 2010). These included two emotion expressions (angry and happy), two non-emotional movements (eye-gaze shifts and speech-lip motion) as well as two control conditions with a neutral static face with and without background motion (following the procedure of Puce *et al.*, 1998). Control conditions were used for contrast with our main experimental conditions to eliminate unspecific activation to visual motion and face shape (Puce *et al.*, 1998). Angry and happy expressions were chosen because they are conveyed by visual cues predominating in the eye or
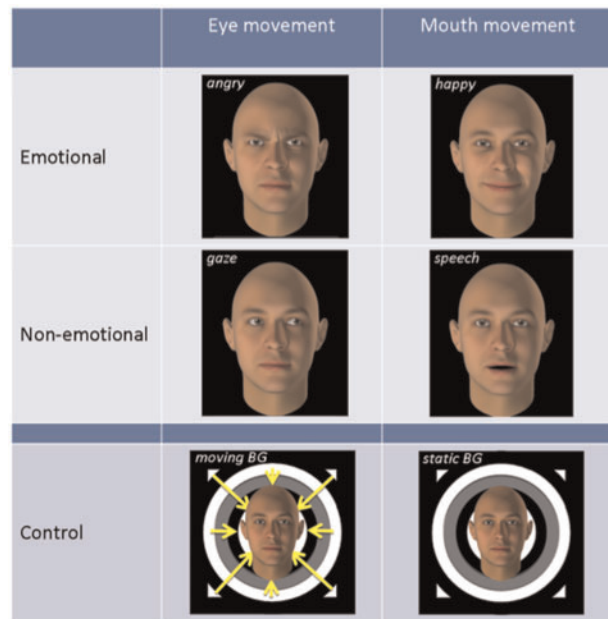
Fig. 1. Stimuli and conditions used in the main experiment. Rows 1 and 2 depict the four experimental conditions angry, happy, gaze and speech. Row 3 shows the two control conditions with static neutral faces and either moving or static backgrounds. For illustration purposes, faces in the control conditions are depicted smaller than in the other four conditions; in the experiment all faces had the same size.

mouth region, respectively, allowing us to compare modulations of STS by information in the upper or lower parts of faces across emotion and non-emotion conditions.

Each condition was presented in a block-design with a one-back repetition detection task for identity (two trials per block). Blocks lasted 20 s each and were repeated six times for each experimental condition, separated by fixation epochs of 5 s. Each block consisted of 12 video-clips (1.7 s). All blocks were pseudorandomly distributed, with equal number of repetitions per run, in two runs of 7 min 45 s each. All visual stimuli were displayed on a screen placed inside the bore of the magnet, viewed via an inbuilt mirror of the tight-fitting rf-coil (Nova Medical). The visual angle of the faces was 10° vertical and 7° horizontal; the stimuli with background expanded 15° × 15°.

Acquisition parameters

High-resolution imaging data was obtained on a head-only 7-T Siemens Magnetom System (Siemens, Erlangen, Germany) in the Center for Biomedical Imaging Center in Lausanne, using a T2*-weighted gradient echo-planar imaging sequence. Twenty-five axial slices were aligned to the superior temporal sulcus along the anterior–posterior orientation (FA 90°, slice thickness/gap = 1.5/1.57 mm, FoV 222 × 222, in-plane 1.5 × 1.5 mm, matrix size 148 × 148). Images were acquired with TR = 2 s and TE = 25 ms. In addition, a high-resolution magnetization pre-pared rapid acquisition gradient echo (MP2RAGE) T1-weighted sequence (slice thickness 1 mm, TR/TE = 5.5 s/2.84 ms, FoV 256 × 240 mm, in-plane 1 × 1 mm) optimized for 7T MRI (Marques *et al.*, 2010) was obtained in sagittal orientation to provide structural brain images from each participant.

Data analysis

Functional images were analyzed with SPM8 software (www.fil.ion.ucl.ac.uk/spm/) and Matlab 7.9 (Mathworks). All images were realigned, corrected for slice timing, normalized to the Montreal Neurological Institute (MNI) template, resliced to a voxel size of 1.5 × 1.5 × 1.5 mm and minimally smoothed (4.5 mm FWHM Gaussian kernel). Spatial filtering techniques (such as the Gaussian smoothing implemented here) have been shown to increase the signal-to-noise-ratio of 7T images (Triantafyllou *et al.*, 2006; Beisteiner *et al.*, 2011), as well as classi-fication performance in multivoxel pattern analysis (MVPA) (Swisher *et al.*, 2010; see also Op de Beeck, 2010, for MVPA effects on simulated data). Realignment parameters were inspected to rule out transient head movements larger than 2 mm (none detected in participants included in our analysis) and then used in the subsequent first-level analysis as covariates of no interest.

Statistical analyses comprised two steps: the univariate gen-eral linear model (GLM) (Friston *et al.*, 1994) and MVPA. The GLM treats each voxel in the brain separately and tests whether it is significantly more activated by the condition of interest than a control condition or baseline. MVPA considers all voxels at the same time and looks for patterns of activation across voxels as a function of conditions.

For the GLM analysis, stimulation blocks were modeled with a box-car function, convolved with a standard hemodynamic response function. Six conditions were defined by separate regressors: angry, happy, gaze, speech, neutral face with mov-ing background and neutral face with static background. Movement parameters from the spatial realignment were entered as covariates of no interest. Statistical parametric maps were generated for each subject from linear contrasts between each of the four conditions (angry, happy, gaze and speech) and the two neutral control conditions [e.g. (2 * angry) − (1 * static control + 1 * motion control)]. We also computed main effects of 'emotional' (happy and angry) *vs* 'non-emotional' (gaze and speech) conditions and of 'eyes' (angry and gaze) *vs* 'mouth' (happy and speech) conditions. Random-effect (RFX) group analyses for the face localizer were performed on the corre-sponding contrast images using one-sample *t*-tests (Friston *et al.*, 1994). In individual analyses of the main experiment we chose to identify voxels activated by different conditions using a threshold of $P < 0.001$ without correction for multiple compari-sons to ensure that differences in the spatial extent and loca-tion of condition-specific response would not be biased by a too strict threshold and primarily driven by differences in response amplitude for one or the other stimulus condition.

ROI analysis

Face-sensitive areas were defined with the localizer experiment in contrasting the face condition *vs* the mean of the house and scrambled picture conditions, separately for each subject. Individual regions-of-interest (ROIs) were delineated using a flexible threshold (ranging between $P < 0.001$ and $P < 0.0001$ uncorrected) to obtain a similar cluster size (at least $k \geq 10$ con-tiguous voxels) across participants and ROIs. In some partici-pants, activation in FFA and OFA partly overlapped (3/16); in these cases we manually separated the two regions using MarsBaR toolbox (http://marsbar.sourceforge.net/). We also used MarsBaR to extract effect sizes for all ROIs and conditions. These values were then subjected to repeated-measure ANOVAs and paired samples *t*-tests.

In addition, we created an anatomical mask of STS for each individual by manually drawing its whole length and depth with MRICron software (Rorden *et al.*, 2007). We included the entire sulcus (i.e. gray matter voxels), from the most anterior to

the most posterior part, as could be defined by visual inspection on native brain anatomy of each participant, and followed the most pronounced, best visible sector along the STS main axis when there were apparent discontinuities or branching in its posterior aspect (see Figure 3 for three examples).

Distribution of activation within STS

To systematically examine the relative location of activated clusters in STS, we computed the distribution of activated voxels along the *x*, *y* and *z* dimensions of the individual anatomical STS mask. Voxels were considered as 'activated' if their signal survived a threshold of $P < 0.001$ (uncorrected) in a condition of interest (relative to the neutral controls). For each dimension of the MNI stereotaxic space (e.g. along *x* axis), we summed all activated voxels over the other two dimensions (e.g. *y* and *z*). This was done for each coordinate along each dimension.

We compared proportions of activated voxels (see Results) with a paired-sample permutation test with 1000 permutations.

Activation in STS segments

In addition to computing activated voxels per coordinate unit, we also counted voxels for three successive segments in STS along the *y*- and *z*-dimensions, allowing us to represent activation distribution at a more global anatomical scale. For each individual STS, we created three equal segments (± one voxel difference when the total voxel number along the axis of interest could not be divided by three), thus yielding posterior, middle and anterior segments along the *y* axis as well as superior, middle and inferior segments along the *z* axis.

Multivoxel pattern analysis

We complemented the univariate analysis with group-wise MVPA, to identify activity patterns discriminative of the different conditions of interest. In particular, we ran two separate analyses, one using as features all voxels within the anatomically defined STS and the other focusing on the face-selective pSTS region only. Keep in mind that, differently from previous analyses carried out on individually defined ROIs (see earlier), here group-wise MVPA searched for coherent distributed representations across the whole population, thus requiring a unique mask for all participants. For this purpose, the anatomical mask was derived by combining all individual STS masks drawn for each participant (see earlier) and including those coordinates present in at least 40% of subjects (see Eickhoff *et al.*, 2005, for a similar cutoff). Instead, the functional pSTS mask was a defined from those coordinates within the anatomical STS mask which were also significantly associated with face stimuli from the localizer session at a threshold corresponding to $P < 0.005$ (uncorrected).

In either analysis, we extracted from the voxels of interest the parameter estimates (*βs*) associated with each condition (contrasted against the two neutral controls) from each subject's first-level analysis. The resulting data [a 16 (subject) × 4 (conditions) matrix for each voxel in the mask] were then fed into a linear kernel support vector machine (SVM) classifier as implemented in the LibSVM (for binary classification) and LibLinear (for four-way classification) toolboxes (Fan *et al.*, 2008). SVMs operate by finding an optimal linear decision boundary (hyperplane) that separates two experimental classes with maximum margin. New data are classified according to which side of the hyperplane they fall onto (Boser *et al.*, 1992). In our study, we performed both binary (e.g. gaze *vs* speech) and four-way classifications. For the four-way classification, we ran the one-*vs*-the-rest strategy with the method of Crammer and Singer, as implemented in LibLinear (Crammer and Singer, 2002; Fan *et al.*, 2008; Keerthi *et al.*, 2008). Essentially, for each class *m*, a binary classifier is developed which distinguishes *m* from the other classes. Then, the new data *x* is fed on all *m* classifiers together and a decision function is obtained. The chosen class is the one that maximizes the *w·x* (i.e. the new data combined with the weights of the previous classification).

In all analyses, classification performance of the SVM was evaluated through leave-one-subject-out cross-validation, which consists in predicting the data from each subject from a model estimated from data of all 15 remaining subjects. This led, for each analysis (anatomical STS and functional pSTS), to a classification accuracy value, reflecting the ability of a linear SVM classifier to discriminate between the conditions of interest in independent subjects. We assessed whether the obtained accuracy values were significantly larger than chance through permutation techniques, by using as cutoff the 95th percentile of a distribution of 1000 control classifications on the same data in which condition labels within each subject were randomly assigned.

In addition, to explore whether the cortical patterns distinguishing between different conditions exhibited any consistent anatomical segregation across subjects, we also analyzed the weight vectors *w* defined during the training phase of the SVM. In the present experimental setup, the weight vector *w* can be represented as a map with the same dimension of the original data whose intensity is proportional to the discriminant power of each voxel (Mourao-Miranda *et al.*, 2005; Gori *et al.*, 2015). Thus, binary classifications led to only one weight-map whose values reflected the voxels' bias in favor of one (positive values) or the other (negative values) condition. Instead, the four-way classifications led to four weight-maps per voxel, whose values reflected the voxels' relative bias in favor (positive values) or against (negative values) a given condition as compared with the other three. Each of these weight-maps was estimated in an unique training fold (independent of the cross-validation procedure) in which data from all 16 subjects were fed to a SVM. The estimation was then repeated under 1000 bootstrap resamples of the population to identify coordinates associated with weights values significantly different from 0 at the group level. These voxels represent regions with anatomically coherent discriminative response between conditions across subjects and were mapped within STS using a height threshold of $P = 0.005$ at the voxel level and a cluster extent of ≥ 5 contiguous voxels.

Results

Behavioral performance

The incidental stimulus repetition detection task (one-back) in the main experiment was performed almost flawlessly. There was no difference in reaction times between the six stimulus categories ($F_{5, 75} = 0.73$, $P = 0.606$) but a modest effect for accuracy (percent correct; $F_{5, 75} = 2.34$, $P = 0.049$). *Post hoc t*-tests revealed that performance for speech lip movements was slightly better than the average of all other conditions (93.92 *vs* 91.25%, $t = 2.83$, $P = 0.013$). All other comparisons were not significant. Moreover, a 2 × 2 ANOVA of the main experimental conditions using face movement type (emotional *vs* nonemotional) and face part (eye *vs* mouth movement) showed no significant effect or interaction.

Activation of the face perception network and functional ROIs

Before focusing on our main experimental questions concerning STS, we first verified that our stimuli effectively engaged the 'core' network associated with face perception. Using the localizer scan data, we contrasted blocks with faces *vs* houses and scrambled pictures. As expected, this revealed strong activations in right ($n=16$; see Table 1 for mean coordinates) and left ($n=15$) FFA, bilateral OFA ($n=15$) as well as right ($n=16$) and left pSTS ($n=15$). No reliable activation was found in left FFA, left STS and bilateral OFA for one participant each.

We then examined how these areas responded to the dynamic face stimuli used in our main experiment by extracting parameters of activity (betas) from each of these six ROIs. Not surprisingly, pSTS was more activated to all dynamic stimuli than FFA and OFA (main effect ROI, $F=19.47$, $P<0.001$), demonstrating differential responses to changeable face cues as manipulated in our study. Remarkably, pSTS showed a different profile than FFA and OFA across emotional and non-emotional stimuli (Figure 2), confirmed by a significant interaction of ROI and stimulus type ($F=15.49$; $P<0.001$). *Post hoc* paired *t*-tests indicated that both FFA and OFA were generally more activated to emotional than non-emotional faces (left FFA: $t=3.52$, $P=0.003$; right FFA: $t=3.88$, $P=0.001$; left OFA: $t=3.32$, $P=0.005$; right OFA: $t=3.13$, $P=0.007$). In contrast, left and right STS were not differentially activated to emotional and non-emotional information in dynamic faces (left pSTS: $t=-0.995$, $P=0.337$; right pSTS: $t=-1.397$, $P=0.183$). This differential pattern of activation between ROIs and conditions validates our stimuli as successfully conveying different kinds of facial movement, with distinct selectivity profiles in ventral and dorsal sectors of the face-processing network.

In addition, our whole-brain analysis of the main experiment identified a posterior temporo-occipital region consistent with the human MT/V5 when contrasting the control conditions with moving *vs* static backgrounds. This area was also more activated by dynamic face stimuli than static neutral faces overall (left $P<0.002$, right $P<0.001$) and by emotional more than non-emotional dynamic faces (left: $t=4.88$, $P<0.001$; right: $t=4.93$, $P<0.001$), a pattern similar to FFA and OFA but different from pSTS. These data further indicate that activation patterns observed in STS are not simply driven by general visual motion cues that drive extrastriate visual areas but by the distinctive social meaning of stimuli.

Activation along STS

Our subsequent analyses focused on the anatomical distribution of activations across the different stimulus categories in our main experiment (angry, happy, gaze and speech), exploiting individual high-resolution obtained at 7T. We first inspected each participant's activation maps along the whole extent of STS. Overall, these maps revealed both overlap and some segregation of activations to the different stimulus conditions. For right STS (Figure 3), the activation cluster for eye gaze movements generally tended to be located in the most postero-superior portion of STS. In contrast, activation to speech generally extended in the more anterior and middle parts of STS. Finally, the two emotion conditions (angry and happy expressions) predominantly activated intermediate areas between the gaze- and the speech-sensitive clusters. Figure 3 shows right STS activations for three representative cases showing examples of this anterior-to-posterior organization ranging from clear and distinct clusters (Subj01) to more overlapping activations (Subj16). This pattern of speech, emotion (angry and happy) and gaze conditions being represented along STS from anterior to posterior portions was present to some degree in 10/16 participants.

Left STS activations (not shown here) were less consistent and exhibited greater overlap (only 5/16 individuals showed the aforementioned organization), especially in the posterior part.

We also addressed the question of how selective STS voxels are by calculating the proportion of those (among all activated STS voxels) that responded to only one condition, or to two, three, or all four conditions (when contrasting each condition to

Table 1. MNI coordinates of face-selective regions

Region		n	x	y	z	Volume
FFA	Left	15	−38 (2.3)	−58 (9.1)	−14 (2.9)	1090
	right	16	40 (3.9)	−55 (8.1)	−12 (2.8)	1677
OFA	Left	15	−41 (5.8)	−81 (5.8)	−4 (6.8)	1622
	Right	15	39 (5.3)	−79 (5.5)	−6 (4.7)	1354
pSTS	Left	15	−50 (6.2)	−48 (7.4)	15 (7.1)	1101
	right	16	50 (5.1)	−47 (11.6)	13 (9.1)	1638

Notes: MNI coordinates are in millimeters and mm³ (volume), std is reported in parentheses. FFA, OFA and pSTS were defined in face localizer with faces > houses, scrambled pictures.
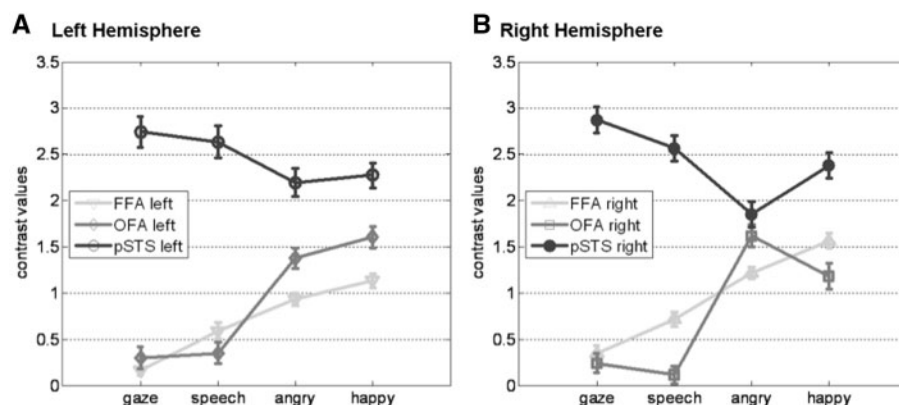
Fig. 2. Activation to emotional and non-emotional stimuli in the face perception network. Plotted are the mean contrast values with standard errors resulting from RFX group analyses of dynamic faces (angry, happy, gaze and speech) *vs* static faces (neutral face with static or moving background). Each line represents average values from a ROI (light gray—FFA, middle gray—OFA and dark gray—pSTS,) in the LH (A) and RH (B).
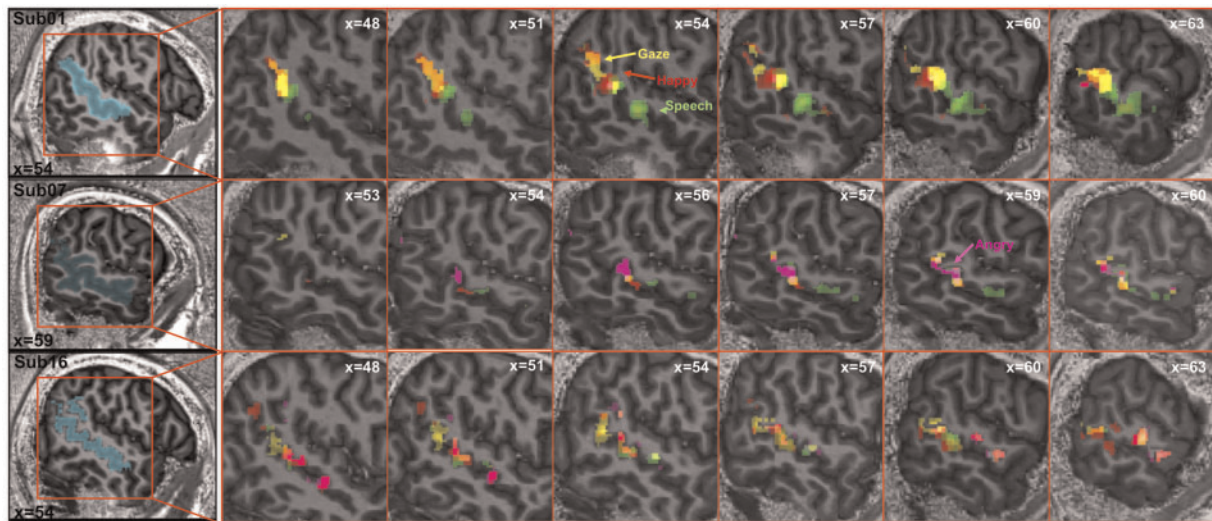
Fig. 3. Activation to angry, happy, gaze and speech conditions in right STS of three representative participants. Both overlapping and segregated representations were consistently seen across all individuals but with different visibility in different individuals as illustrated here.

the control stimuli separately, using a similar threshold of $P < 0.001$). In both hemispheres, for all participants, we found that the vast majority of voxels was selective to one stimulus category (64.2% in right STS, 72% in left STS), which was a significantly higher proportion than voxels activated by two (21.3 and 15%), three (11.4 and 10.5%) or four conditions (3.1 and 2.5%), respectively, (all permutation tests comparing selectivity to one condition with activation to 2–4 conditions were significant at $P < 0.001$).

To obtain a more quantitative group measure of systematic topographic differences while taking individual variation into account, we then examined the spatial distribution of responses along different axes of STS. Activations in each individual anatomical STS mask (see earlier) were quantified by summing the number of activated voxels ($P < 0.001$ uncorrected) for different STS segments, in each stimulus condition. These segments were obtained by splitting STS in three equal parts along either the y or z dimensions (see Materials and Methods). We did not perform this subdivision along the x dimension because the medial to lateral extent of the sulcus is too small to obtain meaningful segments, and because we had no a priori hypothesis about activation clusters along this plane. Subsequently the number of activated voxels per segment was averaged across subjects. Results showed that the four stimulus conditions were unevenly distributed across both the left and right STS and for both the y and z dimensions (Figure 4). This difference was supported by an ANOVA of coordinate values indicating a significant interaction of condition and segment (right STS, y: $F = 6.755$, $P < 0.001$; z: $F = 5.037$, $P < 0.001$; left STS, y: $F_{6, 90} = 3.398$, $P = 0.005$; z: $F_{6, 90} = 3.289$, $P = 0.006$). More detailed ANOVAs and t-tests (Table 2) confirmed that, on both sides, each condition was unevenly distributed throughout STS segments. These effects reflect that speech-responsive voxels were predominantly located not only in the middle segment along the y axis but also in the middle segment along the z axis, with some extension to more anterior and inferior portions on the right side (Figure 4). In contrast, gaze activated both the posterior and middle segments significantly more than the anterior segment. The emotion conditions, angry and happy, also predominated in the middle segment but with larger extension to the posterior segment relative to speech, particularly in left STS.

This analysis also supported the existence of global hemispheric differences in the extent of activations (main effect of hemisphere: $F_{1, 15} = 7.484$, $P = 0.015$; interaction hemisphere × segment: $F_{2, 30} = 4.874$, $P = 0.015$; interaction hemisphere × segment × condition: $F_{6, 90} = 3.621$, $P = 0.003$). Across conditions, activations were generally more widespread in the right than left side. A direct comparison of the number of activated voxels (at $P < 0.001$) showed significant differences between hemispheres not only for the middle (right > left; $F_{1, 15} = 7.103$, $P = 0.018$) and anterior segments ($F_{1, 15} = 6.65$, $P = 0.021$) along the y dimension but also for the inferior (right > left; $F_{1, 15} = 8.453$, $P = 0.011$) and marginally the middle segments ($F_{1, 15} = 4.091$, $P = 0.06$) along the z dimension. Moreover, the relative proportion of responses to different stimulus categories also differed between the two sides, in addition to differences in anatomical distribution (Figure 4). Notably, the proportion of speech-responsive voxels among all those activated was slightly higher in the left hemisphere (LH) than right hemisphere (RH) (60 and 52%, respectively). The latter difference was not significant for the total voxel number ($P = 0.24$, permutation test) but a relative preference for speech in the LH was supported by directly comparing the proportion of speech-responsive voxels vs the mean proportion of angry, happy and gaze voxels, which was significantly different on the left side (permutation test, $P = 0.014$) but not the right ($P = 0.066$).

Multivariate analysis

Although univariate approaches showed a large overlap of activations (Figure 3) with only subtle differences in their 3D distribution and peak location (mainly between gaze and speech conditions in right STS), our qualitative analyses earlier revealed that the majority of voxels (60–70%) preferentially responded to one of the four condition only. This leaves open the possibility that each stimulus recruits specific distributed activity patterns within STS despite gross anatomical overlap (Norman et al., 2006; Corradi-Dell'Acqua et al., 2014, 2016; Deen et al., 2015). Such segregation of functionally distinct cortical population might not be reliably captured with traditional univariate approaches, even in single-subject analyses.
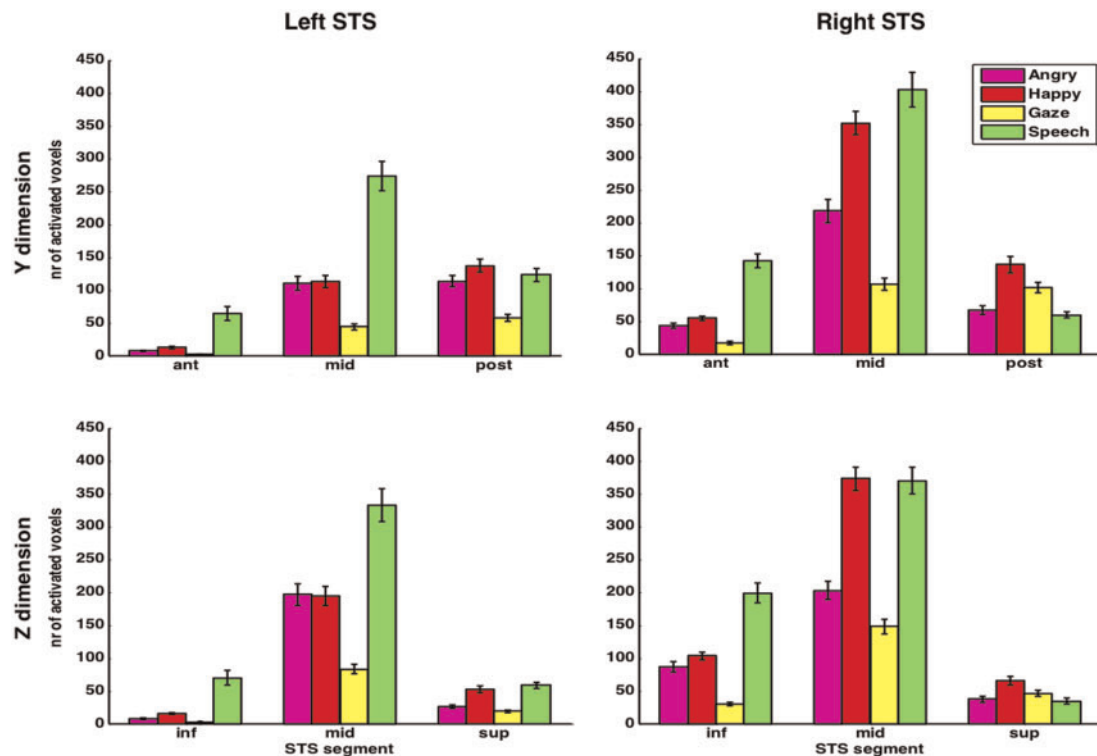
Fig. 4. Spatial distribution of activation in STS. The number of activated voxels is plotted for different segments of left and right STS (first and second columns, respectively). The three segments were obtained by parcellating each subject's STS into three equal parts along either the y axis (anterior to posterior; first row) or the z axis (inferior to superior; second row). Bars represent the group mean; error bars are the SEM. Statistical comparisons are reported in Supplementary Table S1. Magenta bars = angry, red = happy, yellow = gaze and green = speech.

Table 2. Distribution of activation across STS segments

		Left STS					Right STS				
		F	P		t	P	F	P		t	P
y	Angry	5.8	0.007	post–mid	0.093		6.96	0.004	post–mid	−2.467	0.026
				post–ant	3.245	0.005			post–ant	0.978	
				mid–ant	2.598	0.02			mid–ant	2.95	0.01
	Happy	6.2	0.006	post–mid	0.619		10.62	<0.001	post–mid	−2.625	0.019
				post–ant	3.071	0.008			post–ant	1.587	
				mid–ant	3.18	0.006			mid–ant	4.738	<0.001
	Gaze	3.19	0.06	post–mid	0.491		4.44	0.02	post–mid	−0.164	
				post–ant	2.568	0.021			post–ant	2.52	0.024
				mid–ant	2.303	0.036			mid–ant	2.421	0.029
	Speech	5.31	0.01	post–mid	−1.857	0.083	10.16	<0.001	post–mid	−3.204	0.006
				post–ant	0.98				post–ant	−2.113	0.052
				mid–ant	3.796	0.002			mid–ant	3.381	0.004
z	Angry	7.92	0.002	sup–mid	−2.666	0.018	12.43	<0.001	sup–mid	−4.059	0.001
				sup–inf	1.451				sup–inf	−2.555	0.022
				mid–inf	2.997	0.009			mid–inf	3.051	0.008
	Happy	8.13	0.002	sup–mid	−2.535	0.023	18.78	<0.001	sup–mid	−4.526	<0.001
				sup–inf	1.629				sup–inf	−1.239	
				mid–inf	3.302	0.005			mid–inf	4.601	<0.001
	Gaze	7.29	0.003	sup–mid	−2.394	0.03	6.55	0.004	sup–mid	−2.847	0.012
				sup–inf	2.031	0.06			sup–inf	0.812	
				mid–inf	3.023	0.009			mid–inf	2.602	0.02
	Speech	9.35	<0.001	sup–mid	−3.129	0.007	15.8	<0.001	sup–mid	−4.571	<0.001
				sup–inf	−0.268				sup–inf	−2.945	0.01
				mid–inf	3.378	0.004			mid–inf	3.657	0.002

Notes: The table gives the results of ANOVAs (F and P columns), which compare all four conditions, and t-tests (t and P columns), which compare two conditions at a time. Only P-values smaller than 0.1 are reported. P values are not corrected for multiple comparisons. The first 12 rows contain results for the y dimension (anterior–posterior), rows 13–24 list the results for the z dimension (inferior–superior).

Table 3. MVPA: four-class classifications

	Predicted state			
	Gaze	Speech	Angry	Happy
Right whole STS [accuracy **39.06%** (cutoff = 32.81)]				
Gaze	**8**	1	2	5
Speech	2	**6**	2	6
Angry	1	4	**2**	9
Happy	4	1	2	**9**
Left whole STS [**46.88%** (34.38)]				
Gaze	**8**	4	1	3
Speech	1	**11**	2	2
Angry	3	3	**6**	4
Happy	2	6	3	**5**
Right pSTS [**39.06%** (32.81)]				
Gaze	**9**	4	1	2
Speech	4	**4**	1	7
Angry	1	4	**5**	6
Happy	4	4	2	**7**
Left pSTS [25% (32.81)]				
Gaze	**2**	6	5	3
Speech	2	**4**	7	3
Angry	1	6	**9**	0
Happy	4	6	5	**1**

Notes: Values represent the mean accuracy of four-class classification, together with significance cutoff's reported in brackets. Results are shown for the whole anatomical STS mask, and the pSTS functional ROI defined by a separate face localizer. Classification performances that were significantly better than chance are marked in bold. For each classification, the confusion matrix reports how each condition was categorized by the SVM. Bold values refer to correct categorizations (e.g. how frequently gaze parameters were classified as gaze).

Table 4. MVPA: pairwise classifications

	Left	Right
a. Whole STS		
Speech–angry	**78.13 (65.63)**	65.63 (62.50)
Speech–happy	62.50 (62.50)	**71.88 (62.50)**
Gaze–speech	**68.75 (65.63)**	**81.25 (64.06)**
Gaze–angry	**75.00 (65.63)**	**68.75 (62.50)**
Gaze–happy	**65.63 (62.50)**	**68.75 (65.63)**
Angry–happy	56.25 (62.50)	53.13 (62.50)
Emo–non-Emo	62.50 (62.50)	56.25 (62.50)
Eyes–mouth	**68.75 (62.50)**	**71.88 (62.50)**
b. pSTS		
Speech–angry	62.50 (62.50)	43.75 (62.50)
Speech–happy	46.87 (65.63)	62.50 (62.50)
Gaze–speech	43.75 (65.63)	**68.75 (65.63)**
Gaze–angry	**75.00 (65.63)**	**71.88 (62.50)**
Gaze–happy	50.00 (65.63)	**68.75 (65.63)**
Angry–happy	**65.62 (62.50)**	**65.62 (62.50)**
Emo–non-Emo	53.13 (65.63)	59.37 (65.63)
Eyes–mouth	50.00 (62.50)	46.87 (62.50)

Notes: Values represent the mean accuracy of pairwise classifications, together with significance cutoffs reported in brackets. Results are shown for the (a) whole anatomical STS mask and (b) the pSTS functional ROI defined by a separate face localizer. Classification performances were significantly better than chance are marked in bold.

We therefore employed MVPA to further probe for voxel-by-voxel activation profiles across the four stimulus conditions. We first used a multiclass SVM algorithm that tested whether any voxelwise pattern in STS represents one stimulus category more than all others. The algorithm was run with four classes on data from the whole-STS masks, and classification accuracy for each category against the other three was assessed with cross-validation procedures (see Materials and Methods). Results showed that, in both hemispheres, the four-class classification accuracies were highly above chance (Table 3).

The confusion matrix of the four-way classifications reveals good proficiency in detecting gaze (in both hemispheres) and speech (especially in the LH), whereas the two emotional conditions were often confounded with one another (Table 3). This pattern was also confirmed by additional binary classification analyses run on each pair of conditions (Table 4), which showed higher-than-chance accuracies in almost all comparisons involving gaze or speech (except for the speech–happy in left STS), but not when probing for discriminant information between happy and angry.

Figures 5 and 6 depict the location of voxels that were relevant for the classifications. The maps obtained from either four-class and pairwise classifications converged with previous univariate analyses in showing that voxels carrying more information about gaze were most systematically observed in the posterior superior part of right STS. Instead, discriminative speech responses were mostly located in the middle and anterior part of STS. Finally, angry and happy expressions were associated with activity patterns that were more spread out across STS, with a less clear topographic organization. In the LH, this organization was less consistent, with several clusters appearing along STS.

Lastly, we repeated our classification analyses but now reducing the amount of voxels to include only the pSTS subregion that showed a face-selective response in the separate localizer scan. Results for the right pSTS were similar to those obtained with the whole (anatomically defined) STS, with two notable exceptions. First, in pSTS angry and happy expressions could be discriminated (MVPA performance is known to improve when the voxels containing noise or irrelevant information are excluded; see Norman et al., 2006). Second, speech could no longer be discriminated from angry and happy within this smaller but more face-selective subregion. On the other hand, unlike for the right pSTS, classification performance for the left pSTS dropped to chance for all comparisons except for the gaze–angry and angry–happy pairwise discriminations (Tables 3 and 4). Taken together, these data suggest that emotion-specific patterns were reliably present only in the most face-selective part of STS (as identified by the localizer scan). Conversely, speech was best discriminated when taking into account the whole STS, thus pointing a major role played by voxels outside the face-selective pSTS.

These relatively high classification rates observed in MVPA thus agree with the selectivity of voxels as estimated earlier in our more qualitative analysis of single-subject univariate data (see Result section on Activation along STS), which indicated that 64–72% of voxels responded to one condition more than to the three others, independently of their location within STS. This further accords with the notion that this region comprises specialized neuronal subpopulations that may however not be located in contiguous clusters.

Topological organization based on face features or emotional significance

We designed our face stimuli such that they could be organized according to two orthogonal dimensions of facial movements:
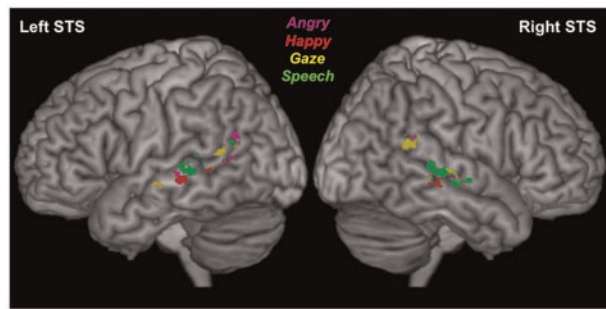
Fig. 5. Localization of voxels contributing most to the four-class classification in left and right STS. Patches in yellow, green or magenta indicate voxels that carry, at the group level, more information about one particular category (gaze, speech, angry or happy) relative to the other three categories (data overlapped across all subjects). For illustrative purpose, maps are thresholded at $P = 0.005$ (uncorrected).

their emotional *vs* non-emotional significance and predominance in the mouth *vs* eye region (happy and speech *vs* angry and gaze). This allowed us to further examine whether the anatomical-functional segregation of face processing along STS could actually be accounted for in terms of supraordinal representations for emotional or featural information in faces. Thus, for the y dimension, a 2×2 ANOVA with face part (eyes, mouth) and STS segment (posterior, middle and anterior) as factors revealed a significant interaction ($F_{2, 30} = 7.04$ and $P = 0.003$), which was absent in a similar analysis using the factor emotional significance ($F_{2, 30} = 0.5$ and $P = 0.61$) instead of face part. The interaction of face part and segment was also significant along the z dimension ($F_{2, 30} = 5.69$ and $P = 0.008$) but again not the interaction of emotional significance and segment ($F_{2, 30} = 0.08$ and $P = 0.92$). Further inspection of these data also indicated that, for the right side, some degree of separation was apparent according to face parts, but not according to emotional significance. Together, these findings point to some separation between sensitivity to dynamic face features in the eye region predominating for the more posterior and superior sectors of right STS but sensitivity to mouth features predominating in more anterior and inferior sectors. For the left STS, no reliable difference in topological distribution was seen according to either face parts or emotional relevance (no significant interaction with segment in y or z dimensions, all $F < 3.32$, $P > 0.05$).

Note, however, that a spatial organization according to face parts does not appear sufficient to fully describe the information encoded in STS, as already indicated by our MVPA results earlier that revealed successful discrimination between gaze and angry expression in right STS (both involving the eyes). Nevertheless, we performed additional MVPA on these supraordinal categories (face part and emotion) and found that the classification performance between emotional and non-emotional movements was not significant in both hemispheres (Table 4). Instead, classification accuracy was above-chance when contrasting eye and mouth movements and was accompanied by a clear dissociation in the anatomical distribution of voxels that contributed most to these two supraordinal classes (Figure 6, bottom). In line with results described in previous sections, the classification weights of voxels, especially in right STS, revealed a general polarity along the y- and z-axes between eye and mouth movements, clustering in the posterior–superior *vs* anterior–inferior regions, respectively (Figure 6). Keep in mind, however, that further comparison between different conditions with
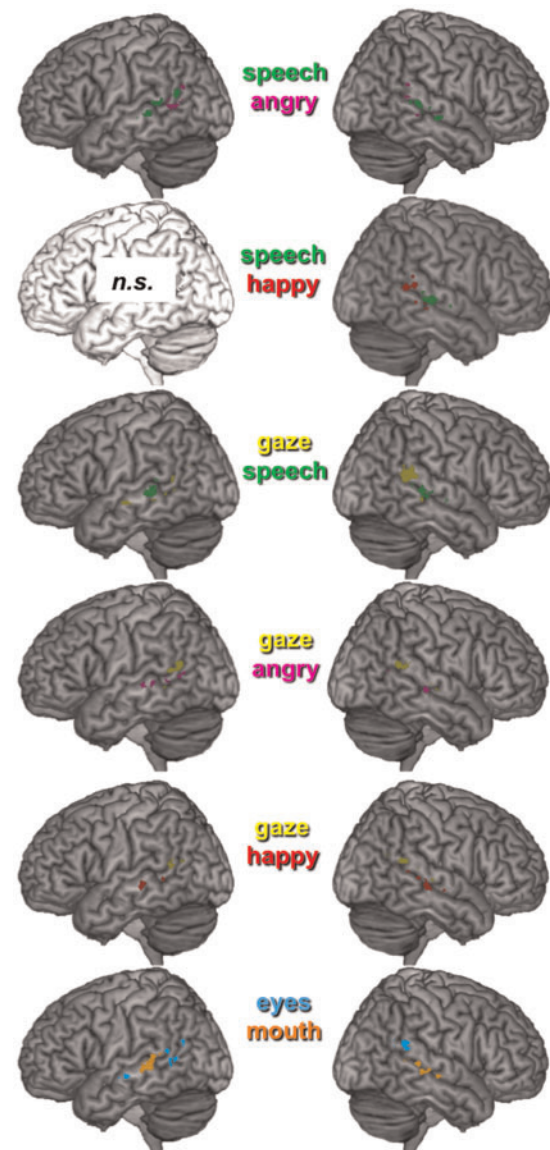

Fig. 6. Localization of clusters contributing most to the binary classification (pairwise comparisons). Note that the comparisons angry–happy and emotional–non-emotional did not yield significant classification results in left and right STS; similarly, speech–happy could not be decoded in left STS. Maps thresholded at $P = 0.005$ (uncorrected), cluster size = 5.

eye movements (gaze *vs* angry) or mouth movements (speech *vs* happy) suggest that the segregation according to face features could be partly driven by particular conditions (e.g. gaze), rather than simply by shared responses to the same face parts.

Together with the multiclass and binary classification data reported earlier, these results converge to suggest that patterns of activity within right STS discriminate between the different kinds of facial movements associated with gaze, expression and speech movements, with a general gradient of sensitivity to eye and mouth information organized from the caudal to rostral sectors of the sulcus. In contrast, patterns of activity within left STS appear to be not only extremely sensitive to the occurrence of speech movement but also widely distributed across the whole sulcus, without the clear-cut organization observed for right STS.

Discussion

Using high-resolution fMRI at 7T, our study is the first to map the functional organization of human STS responses to several dynamic features of face stimuli within the same subjects. Our results reveal partly overlapping, yet distinct subregions in bilateral STS that code for different kinds of facial movements with social and emotional meaning. Thus, despite gross anatomical overlap, consistent differences in cortical distribution between conditions were observed using several analysis approaches. Across individuals, activation to speech-related mouth movements showed a systematic bias to lie toward the more rostral (mid-anterior and inferior) part of STS. In contrast, activation to eye movements was consistently located toward the caudal (mid-posterior and superior) part. Finally, emotion expressions (angry and happy) activated intermediate regions in the middle and inferior segments. This functional segregation was supported by a remarkable convergence of different analyses based on both univariate and multivariate classification approaches. Our findings thus not only agree with, but also extend, other recent results (Deen *et al.*, 2015) suggesting a segregation of different social domains (face, voice, theory of mind and language) along the human STS.

Further, our novel data highlight that the processing of dynamic facial stimuli is accomplished by STS in its entire length not only the posterior part (pSTS) identified by face localizer paradigms in previous fMRI studies (usually using static face pictures). However, whereas the face-selective pSTS on the right side was found to reliably distinguish between all categories of dynamic face features (apart from speech–happy and speech–angry), as demonstrated by our multivariate analysis, this was not the case for the left pSTS. When removing the most anterior/ventral voxels and focusing only on the posterior/dorsal part of the sulcus with face-selective responses in the localizer scan, cortical activity patterns in left pSTS did not discriminate speech and gaze conditions from others (except for gaze–angry). Moreover, the four-class classification (i.e. discrimination of one specific condition against all three others) showed only trend significance in the left pSTS, unlike in the right pSTS. Taken together, this suggests that emotion category-specific activity predominates in the face-selective sector corresponding to pSTS, and that in the RH, this region may hold finer discriminative representations for different types of facial movements. This accords with a general right-hemispheric specialization for face perception (Grand *et al.*, 2003) and emotion processing (Borod *et al.*, 1998). These results also converge with recent studies using MVPA to distinguish between responses to multiple emotion categories from faces (Said *et al.*, 2010) or other stimuli such as bodies or voices (Peelen *et al.*, 2010), where emotion-specific responses were also found in pSTS and its extension toward the temporo-parietal junction. Here, however we were able to map the topological functional organization of STS responses to several other categories of facial motion beyond emotion expression, including gaze and speech.

Other differences were observed between the RH and LH, further supporting some lateralization of functions. First, activations were not only generally more widespread on the right side (i.e. more voxels activated across all conditions) but also spatially more segregated between conditions (Figure 6) as compared with more diffuse organization on the left side. Further, in our multivariate analysis, speech was better discriminated from other conditions in left STS when including voxels from the entire sulcus, rather than only the face-sensitive posterior subregion. These observations also dovetail with a differential specialization for face and language cues in the RH and LH, respectively (see also Deen *et al.*, 2015). In addition, right STS exhibited distinctive cortical patterns that could reliably discriminate between speech and happy expressions (both involving mouth movements); whereas cortical patterns in left STS were less distinctive and unable to discriminate the latter two conditions.

Our results thus not only corroborate but also add to the pioneer work of Pelphrey *et al.* (2005), who reported a subdivision of STS into clusters responding to eye, mouth or hand movements. This was assumed to reflect distinct representations for different body parts or different categories of actions. We also analyzed our data with respect to specific facial movements (eye *vs* mouth), orthogonally to their social meaning. Although our data confirm a topological distinction between responses to specific face features on both sides (eyes and mouth movements generally represented in more posterior and anterior STS, respectively), this distinction appears insufficient to account for the functional segregation observed in STS. For instance, angry expression and gaze conditions both concerned movements in the eye region but were reliably distinguished by MVPA. Thus, an apparent segregation according to different face parts appeared primarily driven by regional selectivity to specific conditions (e.g. gaze *vs* speech rather than eye *vs* mouth). It remains however to be determined whether more subtle distinctions might be found between different kinds of eye movements.

In our dataset, we did not find a reliable discrimination between emotional *vs* non-emotional meaning. At a first glance, this could appear surprising as in the pairwise comparisons emotional states were often well-distinguished from gaze and speech states. However, it should be stressed that the two non-emotional conditions (gaze and speech) are processed by STS through clearly distinct representations. It therefore seems that the STS response to dynamic face stimuli is better explained in terms of a three-category organization (gaze, speech and emotional expressions, with distinction between happy and angry only in pSTS), rather than according to broader supraordinal dimensions of emotional significance. Future research might further determine which facial features contribute to these representations by comparing additional emotion categories (Calder *et al.*, 1996) or different facial action units (Ekman and Friesen, 1978).

Many other studies reported activations in STS during face processing but considered only one or two stimulus categories, making it difficult to extract any systematic organization from this past work. For example, gaze processing has been associated with activity in both anterior (Calder *et al.*, 2007) and pSTS (Hoffman and Haxby, 2000). Face identity and expression were also found to differentially modulate the posterior and anterior STS, respectively (Winston *et al.*, 2004), whereas responses to lip speech were observed in both the middle and pSTS (Wright *et al.*, 2003). In our study, we compared highly controlled facial stimuli that systematically varied in several independent dimensions during the same experiment and were thus able to precisely delineate the spatial layout of STS sensitivity to different dynamic face properties including eye gaze, lip speech and both angry and happy expressions.

The topological organization of STS observed here appears highly consistent with our initial hypotheses based on presumed links with different social and emotional processes. Voxels responding to speech cues in anterior STS were the closest to the auditory cortex, that is, near areas involved in decoding sound and language. This accords with the notion

that visual mouth cues are used to help extract auditory phonemic information (Bristow *et al.*, 2009; Beauchamp *et al.*, 2010). Although some studies reported STS activation to lip-speech in pSTS (Puce *et al.*, 1998; Wright et al., 2003), others found middle and anterior STS areas as here (Allison *et al.*, 2000; Wright et al., 2003; Pelphrey *et al.*, 2005). Furthermore, Pelphrey *et al.* (2005) found peak activation to mouth movements in bilateral STS at locations ($y = -37$ for right side; $y = -44$ for left) very similar to our peaks ($y = -34$ for right side; $y = -47$ for left). In contrast, voxels responsive to eye-gaze in our study clustered in posterior/superior sectors, adjacent to parietal areas implicated in spatial attention and/or eye movements (Corbetta *et al.*, 1998). Again, this accords with the notion that perceived gaze direction can induce reflexive shifts in spatial attention and gaze following in observers (Driver *et al.*, 1999; Vuilleumier, 2002; Kingstone *et al.*, 2004). Activation of pSTS together with parietal and prefrontal areas involved in attention is commonly observed in studies of gaze perception (Hoffman and Haxby, 2000; Cristinzio *et al.*, 2010; Nummenmaa *et al.*, 2010; Ethofer *et al.*, 2011), although anterior activations were also reported (Calder *et al.*, 2007). Finally, the intermediate position of voxels sensitive to emotion expressions suggests that the mid STS may have distinctive projections to limbic areas involved in emotion processing, such as the amygdala, and to frontoparietal areas implicated in emotion regulation, feelings or mimicry (Phelps and LeDoux, 2005; Vrticka *et al.*, 2013).

Note that our findings relate to face stimuli alone but do not directly speak to biological motion involving other body parts. Emotions expressed with the hand (Grosbras and Paus, 2006) or whole body (Kret *et al.*, 2011) might activate partly distinct regions in STS (Pelphrey *et al.*, 2005) or overlap with responses to facial expressions of emotions in a supramodal manner (Peelen *et al.*, 2010). Future studies need to include a large variety of dynamic emotionally and socially relevant stimuli to obtain a more complete understanding of the functional organization of STS.

In sum, our study provides novel evidence that STS carries discriminative information about various dynamic facial cues along its entire spread, and that this information is coded in a spatially organized manner. Critically, this evidence was provided through a convergence of results from both univariate and multivariate approaches. Although each method alone might have its limitation (e.g. see Haufe *et al.*, 2014, for a critical discussion on classification weights), the converging evidence from all approaches used here provides compelling support for an anterior-to-posterior organization of speech, emotional expression and eye gaze movements, with categorical emotion content preferentially represented in the face-selective pSTS subregion, and speech-related responses predominantly located in more anterior STS portions, outside pSTS. Overall, these data shed new light on the functional neuroanatomy of face processing in the human brain.

Funding

References

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, **11**(2), 231–9.

Allison, T., Puce, A., McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, **4**(7), 267–78.

Beauchamp, M.S., Nath, A.R., Pasalar, S. (2010). fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *Journal of Neuroscience*, **30**(7), 2414–7.

Beisteiner, R., Robinson, S., Wurnig, M., Hilbert, M., Merksa, K., Rath, J., Höllinger, I. (2011). Clinical fMRI: evidence for a 7T benefit over 3T. *Neuroimage*, **57**(3), 1015–21.

Borod, J.C., Cicero, B.A., Obler, L.K., *et al.* (1998). Right hemisphere emotional perception: evidence across multiple channels. *Neuropsychology*, **12**(3), 446–58.

Boser, B.E., Guyon, I.M., Vapnik, V.N. (1992). A training algorithm for optimal margin classifiers. Paper presented at the Proceedings of the fifth annual workshop on Computational learning theory, Pittsburgh, PA.

Bristow, D., Dehaene-Lambertz, G., Mattout, J., *et al.* (2009). Hearing faces: how the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, **21**(5), 905–21.

Calder, A.J., Beaver, J.D., Winston, J.S., *et al.* (2007). Separate coding of different gaze directions in the superior temporal sulcus and inferior parietal lobule. *Current Biology*, **17**(1), 20–5.

Calder, A.J., Young, A.W., Perrett, D.I., Etcoff, N.L., Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, **3**(2), 81–118.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., *et al.* (1997). Activation of auditory cortex during silent lipreading. *Science*, **276**(5312), 593–6.

Corbetta, M., Akbudak, E., Conturo, T.E., *et al.* (1998). A common network of functional areas for attention and eye movements. *Neuron*, **21**(4), 761–73.

Corradi-Dell'Acqua, C., Hofstetter, C., Vuilleumier, P. (2014). Cognitive and affective theory of mind share the same local patterns of activity in posterior temporal but not medial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, **9**(8), 1175–84.

Corradi-Dell'Acqua, C., Tusche, A., Vuilleumier, P., Singer, T. (2016). Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. *Nature Communications*, **7**, 10904.

Crammer, K., Singer, Y. (2002). On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, **2**, 265–92.

Cristinzio, C., N'Diaye, K., Seeck, M., Vuilleumier, P., Sander, D. (2010). Integration of Gaze Direction and Facial Expression in Patients with Unilateral Amygdala Damage, *Brain*, **133**(1), 248–61.

Deen, B., Koldewyn, K., Kanwisher, N., Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebral Cortex*, doi: 10.1093/cercor/bhv111.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, F., Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, **6**(5), 509–40.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., *et al.* (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, **25**(4), 1325–35.

Ekman, P., Friesen, W.V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting Psychologists Press.

Engell, A.D., Haxby, J.V. (2007). Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia*, **45**(14), 3234–41.

Ethofer, T., Gschwind, M., Vuilleumier, P. (2011). Processing social aspects of human gaze: a combined fMRI-DTI study. *Neuroimage*, **55**(1), 411–9.

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., Lin, C.-J. (2008). LIBLINEAR: a library for large linear classification. *Journal of Machine Learning Research*, **9**, 1871–4.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S.J. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, **2**(4), 189–210.

Gobbini, M.I., Haxby, J.V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, **45**(1), 32–41.

Gori, I., Giuliano, A., Muratori, F., Saviozzi, I., Oliva, P., Tancredi, R., Cosenza, A. (2015). Gray matter alterations in young children with autism spectrum disorders: comparing morphometry at the voxel and regional level. *Journal of Neuroimaging*, **25**(6), 866–74.

Grand, R.L., Mondloch, C.J., Maurer, D., Brent, H.P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. *Nature Neuroscience*, **6**(10), 1108–12.

Grosbras, M.H., Paus, T. (2006). Brain networks involved in viewing angry hands or faces. *Cerebral Cortex*, **16**(8), 1087–96.

Gschwind, M., Pourtois, G., Schwartz, S., Van De Ville, D., Vuilleumier, P. (2012). White-matter connectivity between face-responsive regions in the human brain. *Cerebral Cortex*, **22**(7), 1564–76.

Haufe, S., Meinecke, F., Görgen, K., *et al.* (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, **87**, 96–110.

Haxby, J.V., Hoffman, E.A., Gobbini, M.I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, **4**(6), 223–33.

Haxby, J.V., Hoffman, E.A., Gobbini, M.I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, **51**(1), 59–67.

Hein, G., Knight, R.T. (2008). Superior temporal sulcus–it's my area: or is it?. *Journal of Cognitive Neuroscience*, **20**(12), 2125–36.

Hoffman, E.A., Haxby, J.V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, **3**(1), 80–4.

Ishai, A., Schmidt, C.F., Boesiger, P. (2005). Face perception is mediated by a distributed cortical network. *Brain Research Bulletin*, **67**(1–2), 87–93.

Keerthi, S.S., Sundararajan, S., Chang, K.-W., Hsieh, C.-J., Lin, C.-J. (2008). A sequential dual method for large scale multi-class linear svms. Paper presented at the Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, Las Vegas, NV.

Kingstone, A., Tipper, C., Ristic, J., Ngan, E. (2004). The eyes have it!: an fMRI investigation. *Brain and Cognition*, **55**(2), 269–71.

Kret, M.E., Pichon, S., Grezes, J., de Gelder, B. (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *Neuroimage*, **54**(2), 1755–62.

Kujala, M.V., Tanskanen, T., Parkkonen, L., Hari, R. (2009). Facial expressions of pain modulate observer's long-latency responses in superior temporal sulcus. *Human Brain Mapping*, **30**(12), 3910–23.

LeDoux, J.E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, **23**(1), 155–84.

Marques, J.P., Kober, T., Krueger, G., van der Zwaag, W., Van de Moortele, P.-F., Gruetter, R. (2010). MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *Neuroimage*, **49**(2), 1271–81.

Mourao-Miranda, J., Bokde, A.L., Born, C., Hampel, H., Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. *Neuroimage*, **28**(4), 980–95.

Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, **10**(9), 424–30.

Nummenmaa, L., Passamonti, L., Rowe, J., Engell, A.D., Calder, A.J. (2010). Connectivity analysis reveals a cortical network for eye gaze perception. *Cerebral Cortex*, **20**(8), 1780–7.

Op de Beeck, H.P. (2010). Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses?. *Neuroimage*, **49**(3), 1943–8.

Peelen, M.V., Atkinson, A.P., Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *The Journal of Neuroscience*, **30**(30), 10127–34.

Pelphrey, K.A., Mitchell, T.V., McKeown, M.J., Goldstein, J., Allison, T., McCarthy, G. (2003). Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. *The Journal of Neuroscience*, **23**(17), 6819–25.

Pelphrey, K.A., Morris, J.P., Michelich, C.R., Allison, T., McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: an FMRI study of eye, mouth and hand movements. *Cerebral Cortex*, **15**(12), 1866–76.

Phelps, E.A., LeDoux, J.E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron*, **48**(2), 175–87.

Puce, A., Allison, T., Bentin, S., Gore, J.C., McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *The Journal of Neuroscience*, **18**(6), 2188–99.

Rorden, C., Karnath, H.O., Bonilha, L. (2007). Improving lesion-symptom mapping. *Journal of Cognitive Neuroscience*, **19**(7), 1081–8.

Said, C.P., Moore, C.D., Engell, A.D., Todorov, A., Haxby, J.V. (2010). Graded representations of emotional expressions in the left superior temporal sulcus. *Frontiers in Systems Neuroscience*, **10**(5), 11.

Swisher, J.D., Gatenby, J.C., Gore, J.C., *et al.* (2010). Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *Journal of Neuroscience*, **30**(1), 325–30.

Triantafyllou, C., Hoge, R.D., Wald, L.L. (2006). Effect of spatial smoothing on physiological noise in high-resolution fMRI. *Neuroimage*, **32**(2), 551–7.

Vrticka, P., Simioni, S., Fornari, E., Schluep, M., Vuilleumier, P., Sander, D. (2013). Neural substrates of social emotion regulation: a FMRI study on imitation and expressive suppression to dynamic facial signals. *Frontiers in Psychology*, **4**, 95.

Vuilleumier, P. (2002). Perceived gaze direction in faces and spatial attention: a study in patients with parietal damage and unilateral neglect. *Neuropsychologia*, **40**(7), 1013–26.

Winston, J.S., Henson, R.N.A., Fine-Goulden, M.R., Dolan, R.J. (2004). fMRI-Adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology*, **92**(3), 1830–9.

Wright, T.M., Pelphrey, K.A., Allison, T., McKeown, M.J., McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex*, **13**(10), 1034–43.