# Supporting Information

von Kriegstein *et al.* 10.1073/pnas.0710826105

## SI Methods

**Participants.** We initially recruited 26 healthy volunteers (16 females, 23 right-handed) and 18 volunteers with hereditary prosopagnosia (11 females, 18 right-handed). All participants gave their written informed consent following the guidelines of the Ethics committee of the J. W. Goethe University. Nine control subjects were excluded to match both groups with respect to age and gender. The first prosopagnosic served as a pilot and was excluded because he knew the hypotheses of our study. Participants were not informed about the purpose of the experiment except that the topic is voice recognition in prosopagnosia.

**Stimuli.** Voice and face stimuli were taken from audiovisual recordings obtained from six male actors using a digital video camera (DCR-PC01E; Sony) (32-kHz sampling rate, 16-bit resolution). The recordings included semantically neutral, phonologically and syntactically homogeneous sentences. For the training phase we recorded, for each speaker, 20 interrogative, second person sentences in German [e.g. "Magst du sie?" ("Do you like her?")]. For the test phase we recorded, for each speaker, 20 declarative, third person sentences [e.g. "Er mag sie." ("He likes her.")]. In total, the same 40 sentences were recorded from each actor. The occupation symbols were presented as pictures taken from the web site http://office.microsoft.com/en-us/clipart. For a vehicle recognition task (see below) sounds of motorbikes, racing cars and trains were taken from several websites. All auditory stimuli were postprocessed using CoolEdit (Syntrillium Software) to adjust overall sound pressure.

Faces for the visual face area localizer experiment were still frames taken from the audiovisual recordings of 36 additional speakers. Visual objects were photographs taken from 36 different objects. All stimuli for the visual localizer were digital color pictures (768 × 576 pixels).

**Experimental Design.** Stimuli were presented and responses were recorded by using Presentation software (http://nbs.neurobs.com).

***Training phase.*** Participants were instructed to learn the association of the three voices, names, and faces or occupations. After each learning session, the level of learning was evaluated. In this evaluation, a voice was presented followed either by a written name or by a face (set 1) or occupation (set 2). Participants reported whether the auditory and visual stimuli were from the same or from a different person by clicking the left or right mouse button. Feedback of the correct combination was provided immediately after each trial. The learning session and tests were repeated twice. If the subjects did not reach a criterion of at least 80% correct responses, the learning and test cycle was repeated a third time. All participants reached the 80% criterion after two or three cycles. A single learning session contained 20 trials of voice-face learning (set 1) or voice-occupation (set 2) learning per speaker. The evaluation session contained four randomly selected trials (of the training set) with audiovisual feedback per speaker. Total exposure to audiovisual information about a speaker was 63 s, for two sessions, or 94 s for three sessions.

*Data acquisition.* Structural and functional MRI was performed on a 3-T Siemens Vision scanner (gradient booster, standard head coil). Functional imaging used an echoplanar imaging sequence covering the whole brain [33 slices; 1-mm gap; voxel size, $3 \times 3 \times 3$ mm$^3$; time to repeat (TR), 2 s; 460 volumes per session per participant in the test phase of the main experiment, and 170 volumes per session per participant for the face area localizer]. Anatomical scans were obtained using a magnetization rapid-acquisition gradient echo sequence (144 slices; TR, 2.3 s; voxel size, $1 \times 1 \times 1$ mm$^3$; 256 matrix). Acoustic stimuli were delivered in the MRI scanner with a commercially available high-quality sound system (mr-confon) (stimuli, 80 dB SPL; scanner noise, 100 dB; passive attenuation by sound system, 40 dB).

*Analysis of MRI data.* MRI data were analyzed with SPM5 (www.fil.ion.ucl.ac.uk/spm) and Matlab 6.5.1 (The MathWorks). Standard spatial preprocessing (realignment and unwarp, normalization to a standard MNI reference brain, and smoothing with an 8-mm full-width at half-maximum (FWHM) Gaussian kernel for group analysis and a 4-mm FWHM Gaussian kernel for single subject analysis) (1) was performed. Statistical parametric maps were generated by modeling the evoked hemodynamic response for the different conditions as boxcars convolved with a synthetic hemodynamic response function in the context of the general linear model (2).

***Face area localizer.*** Population level inferences about blood oxygen level-dependent (BOLD) signal changes in the group analyses of the face area localizer were based on a random effects model that estimated the second-level $t$ statistics at each voxel. The face area localizer was thresholded at $P < 0.001$ uncorrected to define the regions of interest, i.e., posterior STS and FFA.

***Test phase. Categorical analysis.*** BOLD signal changes in the Test Phase in single subjects were considered to be located in a face area (STS or FFA) if they were within 10 mm from the maximum of a face area as defined by the Face Area Localizer in each subject. For the FFA, for four subjects, there was no maximum within this region (at $P < 0.05$ uncorrected). In these cases, we extracted parameter estimates directly from the location of the face area localizer, of these subjects (Table S1 and Table S2).

For the left and right STS, in four subjects each, there was no maximum for the localizer of the STS Face Area ($P < 0.05$ uncorrected) (Table S1 and Table S5). In these subjects, the localizer location was determined by the group maximum location as defined by the face area localizer (STS: left, $x = -52, y = -56, z = 6$; right, $x = 54, y = -40, z = 6$).

*Correlation analysis.* Population level inferences about BOLD signal changes in the group analyses of the test phase were based on a random effects model that estimated the second-level $t$ statistics at each voxel. Comparisons between groups were performed using a two sample $t$ test, with correction for nonsphericity. Conjunctions were used to test for commonalities of the two groups. Group analyses of the test phase were based on region of interest (ROI) analyses and considered significant at $P < 0.05$ corrected for multiple comparisons over voxels.

1. Friston KJ, *et al.* (1995) Spatial registration and normalisation of images. *Hum Brain Mapp* 2:165–189.

2. Friston KJ, *et al.* (1995) Statistical parametric maps in functional imaging: A general linear approach. *Hum Brain Mapp* 2:189–210.

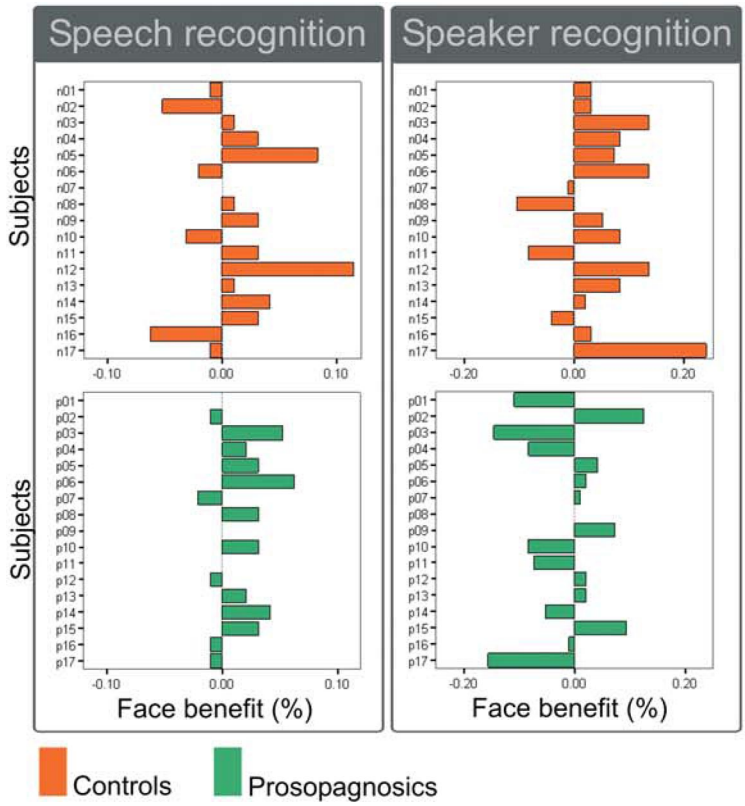


**Fig. S1.** Behavioral results. Face-benefits for speech and speaker recognition for individual subjects of both groups.

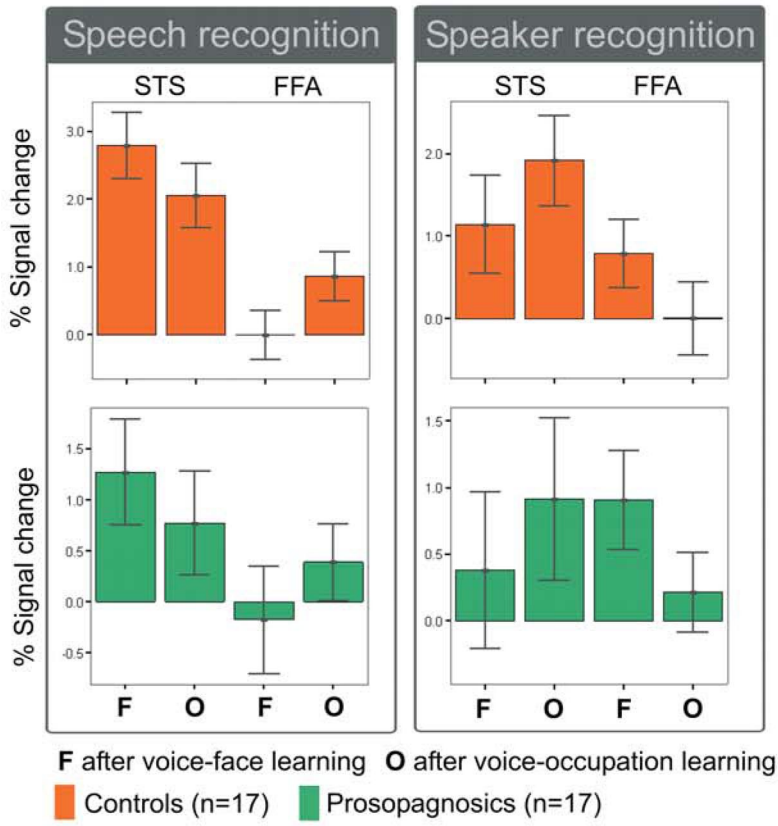


**Fig. S2.** Percentage of signal change of BOLD-responses for each experimental condition, each group and each face-sensitive area, averaged over subjects. The difference contrasts between the two different types of learning are plotted in Fig. 4 *A* and *B*. There were no significant differences in responses, between the two groups, in any of the conditions.

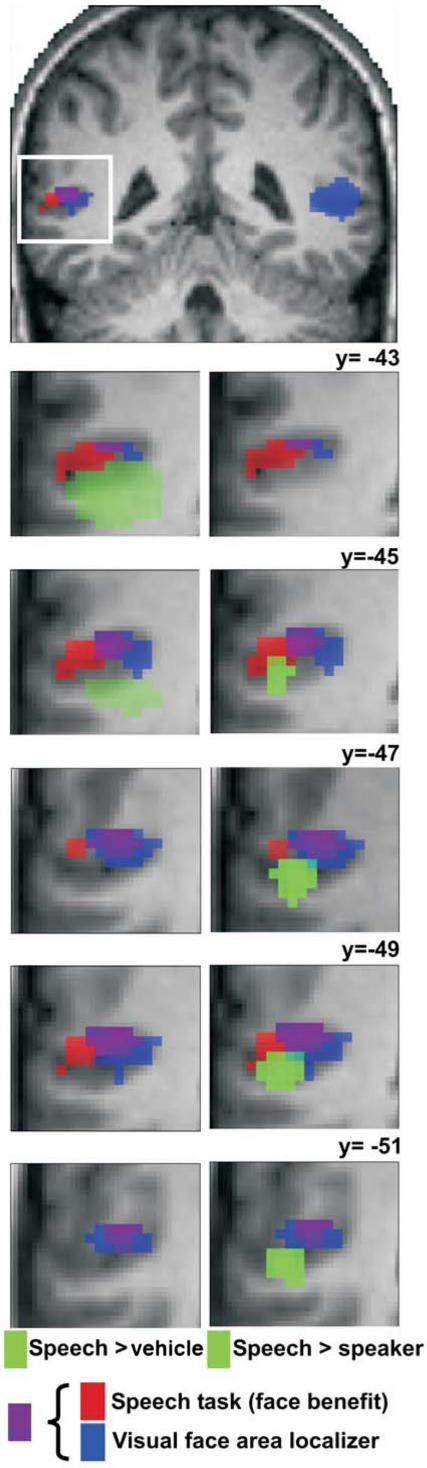


**Fig. S3.** Localization of face-sensitive left STS with respect to speech-responsive areas. To investigate whether the face-movement-sensitive STS is distinct from the auditory-speech-sensitive STS we computed two additional contrasts: (*i*) All speech conditions > vehicle sound condition (green in *Left*) and (*ii*) speech task/voice-occupation learning > speaker task/voice-occupation learning (green in *Right*). The results show that activations in the speech contrast (green) is distinct from activations due to face movement (moving face > static faces, blue). The positive correlation of activity in the STS with the face benefit in the speech task (red) overlaps with the face area (purple) but not with the auditory area (green).

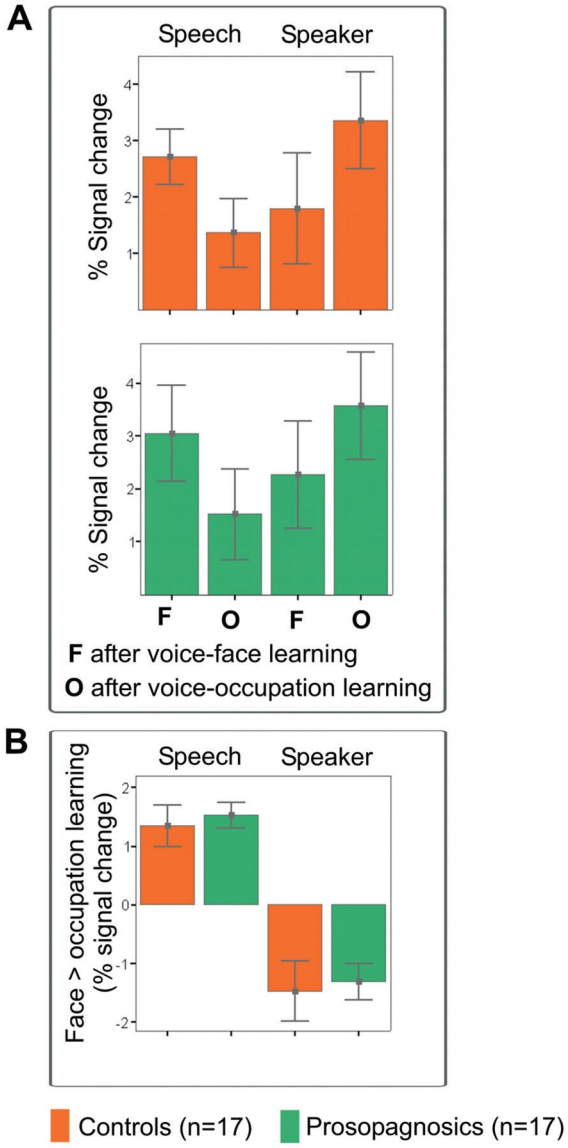


**Fig. S4.** In both groups, activity in the face-sensitive right STS is increased after voice-face learning, for speech recognition [interaction between learning (voice-face vs. voice-occupation) and task (speech vs. speaker)] [ANOVA: $F(1, 32) = 46$, $P < 0.0001$]. Twelve of 17 control subjects and 15 of 17 prosopagnosics (Table S5) showed this effect. (*A*) The percentage of signal change of BOLD-responses for each experimental condition and each group, averaged over subjects. (*B*) The difference contrasts, between the two different types of learning. There were no significant differences in responses, between the two groups, in any of the conditions. The correlation analysis did not show a significant result in this right STS region.

**Table 1. Local activation maxima for single subjects in the left STS**

| | Left STS (visual) | | | | Left STS (auditory) | | | |
|---|---|---|---|---|---|---|---|---|
| Subject | *x* | *y* | *z* | *Z* | *x* | *y* | *z* | *Z* |
| n1 | −54 | −58 | 6 | 2.68 | −60 | −66 | 4 | 2.39 |
| n2 | −48 | −50 | −2 | 2.45 | −48 | −50 | 2 | 2.35 |
| n3 | −58 | −44 | 6 | 4.1 | −48 | −44 | 22 | 2.45 |
| n4 | −52 | −56 | 12 | 3.11 | −46 | −62 | 10 | 2.51 |
| n5 | −62 | −46 | 12 | 4.9 | −66 | −48 | 6 | 2.31 |
| n6 | −60 | −40 | 6 | 2.53 | −58 | −40 | 4 | 2.34 |
| n7 | −52 | −44 | 10 | 3.87 | −48 | −40 | 14 | 2.39 |
| n8 | — | — | — | — | −48 | −50 | 8 | 3.01 |
| n9 | — | — | — | — | — | — | — | — |
| n10 | — | — | — | — | — | — | — | — |
| n11 | −54 | −54 | 4 | 3.59 | −54 | −52 | 6 | 2.77 |
| n12 | −50 | −50 | 14 | 2.52 | −52 | −52 | 14 | 1.68 |
| n13 | −56 | −48 | 4 | 2.53 | −48 | −44 | −4 | 3.64 |
| n14 | −60 | −54 | 18 | 3.82 | −58 | −60 | 16 | 3.86 |
| n15 | −58 | −54 | 10 | 2.47 | −54 | −46 | 10 | 1.91 |
| n16 | −60 | −38 | 10 | 3.16 | −56 | −36 | 18 | 2.37 |
| n17 | −50 | −42 | 10 | 4.06 | −60 | −44 | 0 | 1.95 |
| | | | | | | | | |
| p1 | −48 | −40 | 0 | 1.66 | — | — | — | — |
| p2 | −54 | −52 | 0 | 3.96 | −56 | −58 | −8 | 2.96 |
| p3 | −66 | −46 | 16 | 2.74 | — | — | — | — |
| p4 | −48 | −56 | 8 | 2.39 | −48 | −56 | −2 | 2.95 |
| p5 | −62 | −50 | 8 | 1.93 | −66 | −48 | 8 | 2.36 |
| p6 | −56 | −52 | 22 | 2.88 | −60 | −60 | 26 | 4.23 |
| p7 | — | — | — | — | −54 | −54 | 14 | 3.27 |
| p8 | −60 | −64 | 14 | 3.48 | −64 | −54 | 14 | 2.75 |
| p9 | −52 | −48 | 12 | 2.91 | −50 | −56 | 12 | 4.13 |
| p10 | −58 | −50 | 10 | 2.97 | −66 | −42 | 6 | 2.45 |
| p11 | −40 | −62 | 10 | 2.69 | −42 | −66 | 16 | 2.24 |
| p12 | −58 | −62 | 6 | 2.67 | −54 | −58 | −2 | 3.55 |
| p13 | −62 | −60 | 10 | 2.65 | −66 | −54 | 4 | 1.84 |
| p14 | −48 | −60 | 8 | 3.46 | −42 | −64 | 2 | 3.27 |
| p15 | −56 | −52 | 0 | 2.96 | −50 | −56 | −4 | 2.67 |
| p16 | −62 | −48 | 4 | 3.93 | −60 | −40 | 4 | 3.26 |
| p17 | −54 | −50 | 8 | 2.82 | −48 | −48 | 2 | 2.54 |

Left STS (visual): visual face area localizer (moving face vs. static faces). Left STS (auditory): interaction between task and learning (speech task/voice–face − speech task/voice–occupation) − (speaker task/voice–face − speaker task/voice–occupation). Coordinates *x*, *y*, and *z* are in Montreal Neurological Institute standard and describe local statistical maxima (in millimeters). *Z* indicates the statistical value. Missing values indicate that we could not find a maximum ($P < 0.05$).

**Table 2. Local activation maxima for single subjects in the FFA**

| | FFA (visual) | | | | FFA (auditory) | | | |
|---|---|---|---|---|---|---|---|---|
| Subject | x | y | z | Z | x | y | z | Z |
| n1 | 46 | −44 | −26 | 5.1 | 36 | −46 | −22 | 2.48 |
| n2 | 44 | −42 | −28 | 5.54 | 42 | −50 | −28 | 1.86 |
| n3 | 40 | −52 | −20 | 7.54 | 42 | −44 | −22 | 2.89 |
| n4 | 48 | −60 | −24 | 9.63 | 44 | −60 | −24 | 2.51 |
| n5 | 40 | −54 | −22 | 4.21 | 40 | −56 | −24 | 3.71 |
| n6 | 42 | −46 | −24 | 6.94 | 50 | −42 | −18 | 2.83 |
| n7 | 46 | −56 | −22 | 5.71 | — | — | — | — |
| n8 | 42 | −42 | −24 | 6.21 | 42 | −38 | −22 | 2.05 |
| n9 | 46 | −56 | −28 | 5.1 | 40 | −48 | −26 | 4.78 |
| n10 | 48 | −58 | −18 | 6.78 | 44 | −68 | −22 | 2.34 |
| n11 | 40 | −50 | −24 | 3.37 | 42 | −52 | −28 | 1.71 |
| n12 | 42 | −46 | −30 | 3.28 | 38 | −46 | −32 | 2.8 |
| n13 | 40 | −42 | −32 | 5.36 | 38 | −40 | −26 | 2.38 |
| n14 | 46 | −52 | −28 | 6.07 | — | — | — | — |
| n15 | 38 | −54 | −22 | 8.47 | 38 | −54 | −22 | 2.46 |
| n16 | 44 | −52 | −18 | 6.9 | 36 | −54 | −20 | 2.83 |
| n17 | 40 | −38 | −24 | 6.28 | 42 | −46 | −20 | 1.94 |
| p1 | 42 | −45 | −20 | 5.58 | 42 | −42 | −22 | 2.18 |
| p2 | 44 | −48 | −28 | 9.76 | 50 | −40 | −20 | 3.39 |
| p3 | 46 | −48 | −30 | 5.58 | 48 | −50 | −26 | 2.97 |
| p4 | 36 | −46 | −28 | 2.53 | 44 | −42 | −24 | 2.58 |
| p5 | 48 | −56 | −24 | 5.53 | 50 | −48 | −26 | 3.18 |
| p6 | 40 | −42 | −28 | 5.31 | — | — | — | — |
| p7 | 48 | −60 | −20 | 5.11 | 50 | −60 | −24 | 1.7 |
| p8 | 46 | −50 | −28 | 6.13 | 44 | −44 | −34 | 2.02 |
| p9 | 40 | −58 | −24 | 10.9 | 48 | −48 | −24 | 1.77 |
| p10 | 46 | −52 | −30 | 4.52 | — | — | — | — |
| p11 | 42 | −42 | −28 | 5.12 | 40 | −40 | −30 | 4.03 |
| p12 | 46 | −46 | −22 | 4.61 | 38 | −44 | −22 | 2.4 |
| p13 | 46 | −52 | −20 | 6.29 | 40 | −56 | −20 | 2.41 |
| p14 | 46 | −46 | −32 | 2.58 | 42 | −46 | −32 | 3.37 |
| p15 | 44 | −54 | −26 | 3.8 | 36 | −60 | −22 | 3.59 |
| p16 | 46 | −48 | −30 | 6.16 | 40 | −52 | −28 | 1.74 |
| p17 | 46 | −54 | −20 | 4.51 | 48 | −48 | −22 | 2.35 |

FFA (visual): visual face area localizer (faces vs. objects). FFA (auditory): interaction between task and learning (speaker task/voice–face − speaker task/voice–occupation) − (speech task/voice–face − speech task/voice–occupation). Coordinates *x*, *y*, and *z* are in Montreal Neurological Institute standard and describe local statistical maxima (in millimeters). *Z* indicates the statistical value. Missing values indicate that we could not find a maximum ($P < 0.05$).

**Table 3. List of the most discriminative symptoms for the diagnosis of congenital prosopagnosia and a count of how many of the subjects of both groups displayed these symptoms**

| Symptoms | Prosopagnosics | Controls |
|---|---|---|
| Lasting and irritating subjective uncertainty of face recognition | 17/17 | 0/17 |
| Face recognition deficit especially in crowded places or out-of context encounters | 17/17 | 0/17 |
| False negative and false positive face recognition events | 17/17 | 0/17 |
| Face recognition time longer than socially accepted | 17/17 | 0/17 |
| Face learning time longer than socially accepted | 17/17 | 0/17 |
| Onset in childhood | 17/17 | 0/17 |
| Development of adaptive behaviour | 17/17 | 0/17 |
| No gaze contact necessary | 17/17 | 3/17 |
| Use of explicit learning strategies for visual person recognition | 17/17 | 0/17 |
| Impaired visual recognition of objects and scenes | 17/17 | 0/17 |
| Affected first degree relatives | 17/17 | 0/17 |

Symptoms were assessed based on a written questionnaire and an in-depth interview by an experienced physician (MG) as described in refs. 1 and 2.

1. Gruter M, *et al.* (2007) Hereditary prosopagnosia: The first case series. *Cortex* 43:734–749.
2. Kennerknecht I, *et al.* (2006) First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). *Am J Med Genet A* 140:1617–1622.

**Table 4. Cognitive features usually not impaired in prosopagnosics (1, 2)**

| Symptoms | Prosopagnosics | Controls |
|---|---|---|
| Normal recognition of facial emotions or emotions in general | 15/17 | 16/17 |
| Unimpaired recognition of gender from faces | 17/17 | 17/17 |
| Normal judgment of facial attractiveness | 17/17 | 16/16 |
| Normal semantic memory for persons | 17/17 | 17/17 |

One control did not answer questions regarding facial attractiveness.

1. Gruter M, *et al.* (2007) Hereditary prosopagnosia: The first case series. *Cortex* 43:734–749.
2. Kennerknecht I, *et al.* (2006) First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). *Am J Med Genet A* 140:1617–1622.

**Table 5. Local activation maxima for single subjects in the right STS**

| | Right STS (visual) | | | | Right STS (auditory) | | | |
|---|---|---|---|---|---|---|---|---|
| Subject | *x* | *y* | *z* | *Z* | *x* | *y* | *z* | *Z* |
| n1 | 48 | −40 | 8 | 2.91 | — | — | — | — |
| n2 | 62 | −46 | 14 | 4.25 | — | — | — | — |
| n3 | 54 | −42 | 10 | 4.7 | 56 | −44 | 14 | 1.9 |
| n4 | 50 | −42 | 2 | 2.37 | 58 | −42 | 4 | 2.66 |
| n5 | 56 | −42 | 2 | 3.2 | — | — | — | — |
| n6 | 50 | −44 | 4 | 3.2 | 60 | −54 | 10 | 2.4 |
| n7 | 58 | −44 | 2 | 1.84 | 62 | −52 | 0 | 2.81 |
| n8 | 60 | −52 | −6 | 2.17 | 56 | −44 | −8 | 2.89 |
| n9 | 52 | −44 | 12 | 1.7 | — | — | — | — |
| n10 | — | — | — | — | 46 | −44 | 4 | 2.49 |
| n11 | 54 | −46 | 12 | 4.05 | 54 | −50 | 2 | 2.89 |
| n12 | 50 | −46 | 10 | 4.49 | 50 | −48 | 4 | 2.72 |
| n13 | 58 | −46 | 12 | 2.58 | — | — | — | — |
| n14 | 66 | −54 | 10 | 3.25 | 62 | −56 | 8 | 3.75 |
| n15 | 54 | −60 | 8 | 3.14 | 56 | −54 | 14 | 3.34 |
| n16 | — | — | — | — | 52 | −44 | 8 | 3.03 |
| n17 | 60 | −48 | 4 | 3.79 | 64 | −56 | 0 | 2.06 |
| p1 | 48 | −44 | 8 | 2.1 | — | — | — | — |
| p2 | 50 | −38 | 0 | 2.71 | 50 | −38 | 8 | 2.72 |
| p3 | 66 | −52 | 12 | 3.42 | — | — | — | — |
| p4 | 48 | −44 | 8 | 2.61 | 54 | −52 | 0 | 2.47 |
| p5 | 68 | −46 | 10 | 2.06 | 64 | −56 | 14 | 2.12 |
| p6 | 64 | −40 | 12 | 3.19 | 64 | −40 | 6 | 2.36 |
| p7 | — | — | — | — | 56 | −32 | 4 | 2.96 |
| p8 | 52 | −46 | 6 | 3.15 | 52 | −38 | 8 | 2.69 |
| p9 | 44 | −46 | 10 | 2.76 | 44 | −42 | 6 | 2.86 |
| p10 | 54 | −52 | 4 | 3.23 | 64 | −54 | −2 | 3.75 |
| p11 | — | — | — | — | 50 | −48 | 6 | 1.91 |
| p12 | 60 | −38 | 2 | 3.25 | 58 | −48 | −4 | 3.45 |
| p13 | 62 | −40 | 12 | 2.75 | 60 | −30 | 6 | 1.7 |
| p14 | 62 | −50 | 6 | 4.85 | 66 | −44 | 14 | 2.44 |
| p15 | 50 | −50 | 4 | 3.24 | 52 | −54 | 6 | 2.47 |
| p16 | 54 | −36 | 0 | 3.13 | 64 | −36 | −4 | 2.47 |
| p17 | 54 | −44 | 14 | 3.22 | 54 | −36 | 6 | 2.89 |

Right STS (visual): visual face area localizer (moving face vs. static faces). Right STS (auditory): interaction between task and learning (speech task/voice–face − speech task/voice–occupation) − (speaker task/voice–face − speaker task/voice–occupation). Coordinates *x*, *y*, and *z* are in Montreal Neurological Institute standard and describe local statistical maxima (in millimeters). *Z* indicates the statistical value. Missing values indicate that we could not find a maximum ($P < 0.05$).