Supplementary Material

Materials and Methods

Stimuli

To elicit a relatively natural eye contact impression for the participants, we asked the speakers to imagine that they were talking with a real person (e.g., a friend). In addition, during recording, there was also a real person standing behind the camera with the head position at the same height as the camera lens, in order to make the speakers feel like they were talking with a real person. Similar approaches have been used to elicit direct gaze impressions in non-verbal situations (e.g., Farroni *et al.*, 2002; Burra *et al.*, 2013).

All recordings were done with a digital video camera (Legria HF S10 HD-Camcorder, Canon Inc., Japan) under constant luminance conditions. The video and audio streams were extracted from the original recordings. The video streams were converted to avi format and resized to 1024 × 768 pixels to match the display resolution of the display monitor (1024 × 768 pixels) in the MRI scanner. All videos were processed and cut in Final Cut Pro (version 6, HD, Apple Inc., USA).

The audio tracks of the monologues were post-processed using Matlab (version 8.0, The MathWorks, Inc., USA) to adjust overall sound level with equal root mean square of 0.052. For the “Noise” condition, we mixed the audio track with natural acoustic noise, which consisted of people talking and the clatter of dishes in a cafeteria. We set the auditory signal-to-noise ratio (SNR) level to 0 dB, because a pilot experiment in

the MRI-scanner on one participant showed that this SNR elicited more and longer fixations on the mouth area than the condition without noise, while the speech was still intelligible. The post-processed audio tracks without and with noise were combined with the video streams, forming videos without noise ("Normal videos") and videos with noise ("Noise videos").

Five native Germans (average 23 year-old, 2 females) who did not participate in the actual experiment rated the videos on scales from 1 to 5 regarding their naturalness and emotional content (1 = not emotional/natural at all, 2 = somewhat emotional/natural, 3 = medium emotional/natural; 4 = very emotional/natural, and 5 = extremely emotional/natural). The results showed that the video topics were emotionally neutral and speakers managed to talk in a neutral and natural manner (emotiveness of topic: 1.98 ± 0.49SD, emotiveness of way of talking: 1.70 ± 0.39SD; naturalness: 3.3 ± 0.43SD). The intra-class correlation coefficient based on all ratings was 0.91, indicating a high inter-rater reliability.

MRI data acquisition

The field-map scan consisted of gradient-echo readout (24 echoes, inter-echo time 0.95 ms) with standard 2D phase encoding. The B0 field was obtained by a linear fit to the unwrapped phases of all odd echoes. Before the functional runs, structural images were also acquired for each participant using a high-resolution, T1-weighted 3D MP-RAGE sequence: TI=650 ms, TR=1300 ms, TE=3.93 ms, alpha=10°, spatial

resolution of 1 mm^3, two averages.

Speech recognition analysis

We analyzed reaction time (RT) and percent correct in the speech recognition task. Differences in the performance (percent correct and RT) between the two experimental conditions (normal, noise) were compared using paired *t*-tests. All the statistical analyses were conducted with IBM SPSS (Version 20.0, IBM Corp., USA).

Eye tracking analysis

We used the Tracker software (https://www.cabrillo.edu/~dbrown/tracker/) to estimate the speaker's head position in the video frame by frame. The software determines the speaker's head positions, in terms of coordinates x and y, in each video frame and compares it to the head position at the first frame. We then used a customized matlab script to correct the fixation position in accordance with the relative head positions obtained from the Tracker software. First, we used a simple formula $\frac{\sum_{i=1}^{n} X_i}{n}$ to estimate the average relative head position change (relative x and y change) within a fixation. X_i represents relative head position (either x or y) at frame i, n represents the number of frames for a given fixation. n was computed as the fixation duration divided by the frame duration (40ms). The relative head position change (relative x and y change) were then subtracted from the raw fixation position captured by the eye tracker (raw x and y) to get the corrected fixation position (corrected x and y). The corrected fixation positions were then used for the AOI analysis and dwell analysis

described in the main text.

fMRI analysis

Pre-processing

The first two volumes of the functional images were discarded prior to data analysis to allow the magnetic field to stabilize. We performed standard pre-processing procedures including slice-time correction, realignment and unwarp, coregistration, normalization to Montreal Neurological Institute (MNI) standard stereotactic space, and spatial smoothing at 6 mm full width half maximum (FWHM).

Region of interest (ROI)-based Psychophysiological Interactions (PPI) analysis

Definition for source ROIs

We defined regions for the Eigenvariate extraction for the PPI analyses based on the functional MRI results, or if that was not possible, anatomically (Fig. 3).

Functionally defined ROIs. Several regions could be determined based on higher responses to Eyes than Mouth in the whole brain *standard GLM* analysis (cuneus: x = -6, y = -99, z = 18 and x = 12, y = -93, z = 21; pSTS: x = 54, y = -54, z = 15; mPFC: x = -3, y =21, z = 30; dlPFC: x = 45, y =30, z = 36). For these regions, we extracted the Eigenvariate from spheres (10 mm radius) centered on the subject-specific statistical peaks in the Eye vs. Mouth contrast. Subject-specific coordinates are listed in Table S6.

Standard anatomical maps. Standard anatomical maps were available for bilateral Amy, Pulv, ITC, OFC (including inferior, middle and superior parts of orbital frontal gyrus) and LOC (including anterior and posterior parts of lateral occipital cortex) in the WFU_PickAtlas (Maldjian et al., 2003) or the SPM Anatomy toolbox (v2.1) (Eickhoff et al., 2005). Here we extracted the Eigenvariates from each of the anatomical areas for both hemispheres.

Customized anatomical masks. The FG, the right aSTS and the SC showed no significant responses in the GLM analysis and no standard anatomical maps were available for them. We therefore made customized masks for these regions. The fast-track modulator model assumes involvement of the fusiform gyrus (FG) in eye contact because of its role in face identity processing (Fig. 1C). As the Fusiform Face Area (FFA) has been implicated in face identity processing (Rotshtein *et al.*, 2005; Nestor *et al.*, 2011), we restricted the ROI of FG to bilateral FFA. We used the FSL software (Version 5.0.8, FMRIB, Oxford, UK, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/) to extract the probabilistic map of the FG and intersected it with the probabilistic atlas for face processing (Engell and McCarthy, 2013) (threshold at 0.25). For the right aSTS, we first extracted probabilistic maps of the temporal pole (TP) in the right hemisphere (thresholded at 0.1) with FSL software. We then restricted the TP map to the STG/S region by confining the medial boundary to the medial extent of the STS, the inferior boundary to the lower bank of the STS and the superior boundary to the

lateral fissure. Because SC is a clearly visible and discrete anatomical region on a standard anatomical template, bilateral SC masks were defined as 4-mm-radius spheres with reference to a brain atlas (Duvernoy, 1991) using the MRIcron 3D ROI tool (http://www.mccauslandcenter.sc.edu/mricro/mricro/3droi.html). For the FG and SC we extracted the Eigenvariates combined over both hemispheres. Since the fast-track modulator model makes specific predictions about the role of the right aSTS, we restricted the Eigenvariate extraction to the aSTS in the right hemisphere.

Definition for target ROIs

The target ROIs were defined similarly as the source ROIs. For the functionally defined ROIs we used the group statistical maximum peak coordinates based on higher responses to Eyes than Mouth in the whole brain *standard GLM* analysis (cuneus: x = -6, y = -99, z = 18 and x = 12, y = -93, z = 21; pSTS: x = 54, y = -54, z = 15; mPFC: x = -3, y =21, z = 30; dlPFC: x = 45, y =30, z = 36). Around these statistical maxima, we created 10-mm-radius spheres using the MRIcron 3D painting tool. The maximum image intensity difference was set from origin to 100 and maximum brightness difference at edge to 80 to restrict the ROIs to the gray matter. All other target ROIs were identical with the source ROIs (see section *Definition for source ROIs* and Fig. 3).

Results

Speech recognition task

The participants performed the speech recognition task with high accuracy both in the normal condition (85% correct responses) and the noise condition (80% correct responses). There was no significant accuracy difference between the two conditions ($t_{(17)}$ = 1.54, p = 0.143). Reaction time (RT) was significantly shorter during the normal condition (1.43 s ± 0.20) as compared to that during the noise condition (1.57 s ± 0.25) ($t_{(17)}$ = 2.54, p = 0.021).

Eye Gaze Patterns

NE and IEI within condition –comparisons with Off events

NE (Table S1): Participants fixated more on the eyes than on the other parts of the video (Off) in both conditions (Normal: t = 5.08, p < 0.001; Noise t = 3.44, p = 0.009). They fixated more on the mouth than on the other parts of the video (Off) in the Noise condition (t = 3.41, p = 0.009), but equally in the Normal condition (t = 1.97, p = 0.193).

IEI (Table S2): Participants looked longer at the eyes and mouth areas than at other parts of the video in both conditions (Normal: Eyes > Off, t =9.18, p < 0.001, Mouth > Off, t = 5.47, p < 0.001; Noise: Eyes > Off, t =7.16, p < 0.001, Mouth > Off, t = 4.19, p = 0.002).

References

Burra, N., Hervais-Adelman, A., Kerzel, D., Tamietto, M., De Gelder, B. & Pegna, A. J. (2013) 'Amygdala activation for eye contact despite complete cortical blindness', *The Journal of neuroscience*, **33**(25), pp. 10483-10489.

Duvernoy, H. (1991) 'The human brain: surface, blood supply, and three-dimensional anatomy', New York: Springer-Verlag.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K. & Zilles, K. (2005) 'A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data', *Neuroimage*, **25**(4), pp. 1325-1335.

Engell, A. D. & McCarthy, G. (2013) 'Probabilistic atlases for face and biological motion perception: an analysis of their reliability and overlap', *Neuroimage*, **74**, pp. 140-151.

Farroni, T., Csibra, G., Simion, F. & Johnson, M. H. (2002) 'Eye contact detection in humans from birth', *Proc Natl Acad Sci U S A*, **99**(14), pp. 9602-9605.

Maldjian, J. A., Laurienti, P. J., Kraft, R. A. & Burdette, J. H. (2003) 'An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets', *Neuroimage*, **19**(3), pp. 1233-1239.

Nestor, A., Plaut, D. C. & Behrmann, M. (2011) 'Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis', *Proceedings of the National Academy of Sciences*, **108**(24), pp. 9998-10003.

Rotshtein, P., Henson, R. N., Treves, A., Driver, J. & Dolan, R. J. (2005) 'Morphing Marilyn into Maggie dissociates physical and identity face representations in the

brain', *Nature neuroscience*, **8**(1), pp. 107-113.

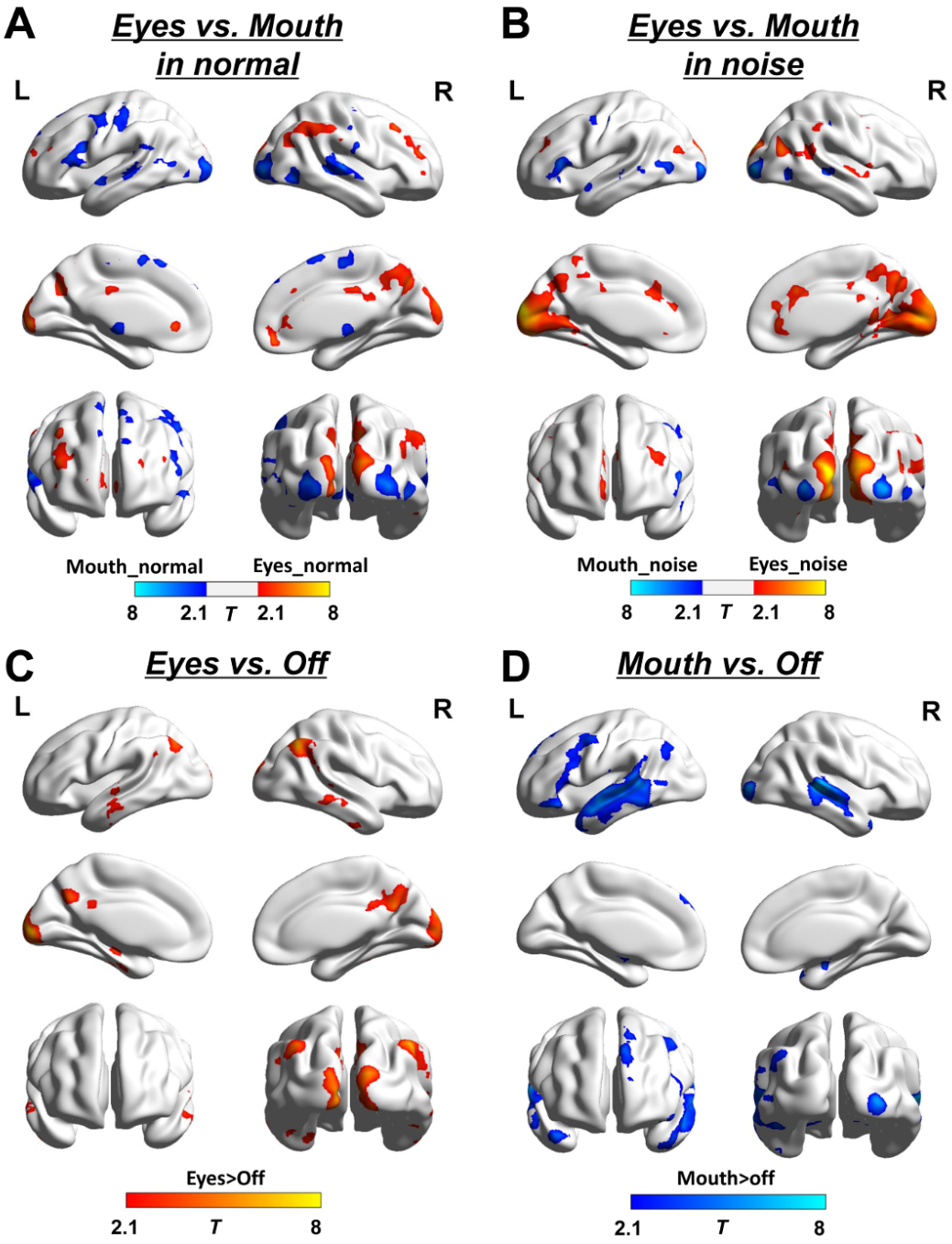

Figure S1. ***A-B,*** **Brain regions showing BOLD response differences between eye contact and mouth fixation in the normal and noise conditions separately.** For visualization only, clusters surviving with a voxel-level threshold of $p < 0.05$ and a minimum size of 50 voxels are shown. MNI coordinates of significant brain regions are listed in **Table S1**. ***C-D,*** Brain regions showing higher BOLD response for eye contact and mouth fixation as compared to Off fixation respectively. For visualization

only, clusters surviving with a voxel-level threshold of $p < 0.05$ and a minimum size of 160 voxels are shown. MNI coordinates of significant brain regions are listed in **Table S2**.

Table S1. The number of event across conditions and within condition for each participant

Subjects	Across Conditions			Normal Condition			Noise Condition			Total
	Eyes	Mouth	Off	Eyes	Mouth	Off	Eyes	Mouth	Off	
sub01	407	808	629	200	355	265	207	453	364	1844
sub02	517	467	355	283	231	202	234	236	153	1339
sub03	499	600	594	301	302	308	198	298	286	1693
sub04	848	674	290	340	224	129	508	450	161	1812
sub05	435	399	227	222	201	114	213	198	113	1061
sub06	659	535	370	317	252	205	342	283	165	1564
sub07	886	501	556	355	161	239	531	340	317	1943
sub08	881	523	461	462	236	269	419	287	192	1865
sub09	505	562	348	299	316	204	206	246	144	1415
sub10	445	465	243	283	279	156	162	186	87	1153
sub11	521	551	349	352	349	229	169	202	120	1421
sub12	551	287	428	258	106	191	293	181	237	1266
sub13	838	625	317	410	264	198	428	361	119	1780
sub14	1011	915	197	468	416	95	543	499	102	2123
sub15	275	210	90	130	85	38	145	125	52	575
sub16	528	196	459	257	95	233	271	101	226	1183
sub17	920	771	459	431	354	193	489	417	266	2150
sub18	416	288	249	210	131	124	206	157	125	953
sub19	541	526	606	292	217	278	249	309	328	1673
Average	614.89	521.21	380.37	308.95	240.74	193.16	305.95	280.47	187.21	1516.47
SD	213.55	193.81	149.61	90.42	95.59	69.48	136.50	116.11	90.14	419.19

Table S2. The inter-event interval across conditions and within condition for each participant

Subjects	Across Conditions			Normal Condition			Noise Condition		
	Eyes	Mouth	Off	Eyes	Mouth	Off	Eyes	Mouth	Off
sub01	1.62	1.68	1.08	2.11	1.72	1.09	1.15	1.65	1.06
sub02	2.53	2.31	0.74	2.82	1.76	0.77	2.19	2.86	0.70
sub03	1.31	2.48	0.81	1.48	2.05	0.72	1.05	2.92	0.91
sub04	2.27	0.83	0.69	3.37	0.64	0.61	1.53	0.93	0.75
sub05	2.34	2.07	0.64	2.38	1.93	0.58	2.31	2.22	0.71
sub06	2.58	1.33	0.73	2.79	1.18	0.61	2.39	1.47	0.88
sub07	2.54	0.53	0.32	3.24	0.50	0.34	2.07	0.54	0.31
sub08	2.32	0.76	0.53	2.40	0.71	0.51	2.25	0.80	0.56
sub09	1.76	2.87	0.55	1.95	1.86	0.52	1.47	4.17	0.59
sub10	2.00	3.42	0.83	1.99	2.14	0.71	2.02	5.33	1.03
sub11	1.19	2.86	0.58	1.24	2.03	0.60	1.09	4.31	0.54
sub12	3.78	0.56	0.42	4.53	0.44	0.34	3.13	0.62	0.48
sub13	2.52	0.72	0.41	2.80	0.67	0.45	2.26	0.76	0.34
sub14	1.84	0.81	0.46	2.17	0.76	0.41	1.55	0.86	0.51
sub15	7.38	2.98	0.39	9.95	0.70	0.42	5.07	4.53	0.36
sub16	4.25	0.73	0.63	4.49	0.71	0.62	4.02	0.76	0.64
sub17	1.93	0.79	0.60	2.03	0.80	0.71	1.85	0.78	0.52
sub18	4.48	1.80	0.74	5.03	1.73	0.77	3.92	1.86	0.71
sub19	2.42	1.83	0.67	2.93	1.55	0.65	1.82	2.02	0.69
Average	2.69	1.65	0.62	3.14	1.26	0.60	2.27	2.07	0.65
SD	1.44	0.95	0.18	1.94	0.62	0.18	1.07	1.53	0.22

Table S3. Coordinates and *p*-values for brain regions showing significant response differences in simple main effects.

Region	Side	p-value (FWE corrected)	cluster volume (mm^3)	T value	MNI coordinates x	y	z	Brodmann Area
Eyes_normal > Mouth_normal[a]								
Cun&Cal&Prec	B	0.000	2326	7.22	-3	-78	57	17/18,7
&TPJ&pSTS	R			4.39	39	-57	33	39/40
vmPFC& dlPFC	R	0.000	417	5.56	9	48	-12	10/24/32
				4.89	39	45	18	9/46
Eyes_noise > Mouth_noise[a]								
Cun&Cal&Prec	B	0.000	4516	9.93	-6	-96	18	17/18,7
				7.81	15	-87	9	
				4.55	6	-63	42	
vmPFC	L	0.002	482	5.76	-3	21	30	10/24/32
				3.36	6	36	3	
TPJ&pSTS	R	0.024	336	5.09	39	-69	21	39/40/42
				3.51	60	-42	21	
Mouth_normal > Eyes_normal[a]								
STG/S	R	0.000	676	5.91	48	-30	9	21/22
MOG&MT	R	0.007	418	5.22	30	-96	9	18/19/37
				3.41	45	-63	-3	

STG/S	L	0.018	384	4.80	-48	-42	21	21/22
IFG&PCG	L	0.000	669	4.50	-42	9	21	6/45
				-3.15	-39	-3	51	
dmPFC&SMA	L	0.000	342	4.10	-15	36	36	6/9
				4.08	3	6	63	
Mouth_noise > Eyes_noise[b]								
MOG	R	0.000	136	6.57	33	-93	0	18
MOG	L	0.000	134	6.23	-30	93	6	18
IFG	L	0.000	202	4.91	-54	27	12	45
MT	R	0.000	69	4.60	48	-57	0	37
STG/S	R	0.000	98	4.55	54	-30	3	21/22
MT	L	0.000	99	4.39	-48	-66	6	37
PCG	L	0.000	147	4.36	-42	0	51	6
aSTG/S	L	0.000	156	4.07	-63	0	-6	21/22

Threshold: a, voxel-level $p < 0.05$, k > 50 voxels, FWE cluster-corrected $p < 0.05$ across whole brain; b, voxel-level $p < 0.001$ uncorrected and k > 50 voxels were used here due to no significant area being found with correction.

Abbreviations: a, anterior; p, posterior; Cun, cuneus; Cal, calcarine; Prec, precuneus; TPJ, temporoparietal junction; STS/G, superior temporal sulcus/gyrus; vmPFC, ventral medial prefrontal cortex; dlPFC, dorsolateral prefrontal cortex; MOG, middle occipital cortex; IFG, inferior frontal gyrus; PCG, precentral and/or postcentral gyrus; dmPFC, dorsomedial prefrontal cortex; SMA, supplementary motor area; MT, middle temporal gyrus/sulcus; L, left hemisphere, R,

right hemisphere, B, bilateral hemispheres.

Table S4. Coordinates and *p*-values for brain regions showing significant higher responses in Eyes/Mouth vs. Off contrasts.

Region	Side	p-value (FWE corrected)	cluster volume (mm^3)	T value	MNI coordinates x	y	z	Brodmann Area
Eyes > Off								
Cun&Cal&Prec	B	0.000	1067	7.76	-6	-93	6	17/18,7
				7.57	18	-90	-3	
				4.98	6	-57	30	
TPJ	R	0.000	158	6.88	45	-60	42	39
TPJ	L	0.019	119	4.77	-39	-66	45	39
Mouth > Off								
STG/S	R	0.000	674	9.16	57	-33	9	21/22,48
STG/S	L	0.000	880	7.00	-48	-21	-15	21/22,48
Cerebellum	R	0.000	365	6.52	30	-72	-33	
MOG	L	0.017	112	5.76	-27	-99	3	18
dmPFC	L	0.001	177	5.36	-12	36	54	8/9
PCG	L	0.024	105	4.84	-48	6	48	6

Threshold: voxel-level $p < 0.01$, k > 50 voxels, FWE cluster-corrected $p < 0.05$ across whole brain.

Abbreviations: Cun, cuneus; Cal, calcarine; Prec, precuneus; TPJ, temporoparietal junction; STS/G, superior temporal sulcus/gyrus; MOG, middle occipital cortex; PCG, precentral and/or

postcentral gyrus; dmPFC, dorsomedial prefrontal cortex; L, left hemisphere; R, right hemisphere; B, bilateral hemispheres.

Table S5. Coordinates for target regions showing significant connection with source regions in PPI analyses.

		Target regions											
		bCun	bSC	bAmy	bPulv	bITC	bLOC	rdlPFC	bOFC	rpSTS	lmPFC	raSTS	bFFA
Source regions	rCun			30,0,-18	6,-24,6		36,-84,-6			54,-51,9			42,-60,-12
	bSC											54,12,-15	
	bAmy												
	bPulv										0,24,27		
	bITC		-3,-27,-3							57,-57,6		54,15,-15	-39,-54,-15
	bLOC			27,-6,-12	-15,-27,6				-33,-27,12	54,-57,6	-6,15,30	54,9,-21	42,-60,-12
	rdlPFC						-48,-78,15			54,-48,9	-6,21,24		
	bOFC										-6,12,27		
	rpSTS						-30,-90,0						
	lmPFC												
	raSTS									51,-51,15			
	bFFA			21,-6,-15			-30,-90,0			51,-48,9		57,12,-21	

Cun, cuneus; SC, superior colliculus; Amy, amygdala; Pulv, pulvinar; ITC, inferior temporal cortex; LOC, lateral occipital cortex; dlPFC, dorsolateral prefrontal cortex; OFC, orbitofrontal cortex; a/pSTS, anterior/posterior superior temporal sulcus; mPFC, medial prefrontal cortex; FFA, fusiform face area; b, bilateral; l, left; r, right.

Table S6. Subject-specific coordinates for source regions in ROI-based PPI analyses

Subjects	rCun				lCun				pSTS				mPFC				dlPFC			
	MNI coordinates			*Distance to group peak coordinate*	MNI coordinates			*Distance to group peak coordinate*	MNI coordinates			*Distance to group peak coordinate*	MNI coordinates			*Distance to group peak coordinate*	MNI coordinates			*Distance to group peak coordinate*
	x	***y***	***z***		***x***	***y***	***z***		***x***	***y***	***z***		***x***	***y***	***z***		***x***	***y***	***z***	
sub01	9	-93	6	15.28	-3	-98	9	9.64	57	-51	19	5.60	-6	17	24	8.05	52	25	35	8.68
sub02	20	-92	18	8.94	-6	-98	9	9.50	42	-54	15	12.19	-3	30	27	9.07	41	30	41	6.68
sub03	9	-99	19	7.35	-9	-93	14	7.82	51	-53	19	5.39	-3	21	36	6.00	35	29	31	10.90
sub04	10	-95	17	5.01	-12	-95	17	7.11	54	-57	21	6.85	-6	24	34	5.51	44	35	37	5.14
sub05	18	-86	24	9.35	-9	-93	24	9.11	58	-56	12	5.53	4	24	29	7.63	46	32	39	4.00
sub06	15	-95	20	3.62	-12	-101	14	7.35	53	-53	13	2.71	0	22	30	2.94	39	30	34	6.31
sub07	6	-93	18	6.77	-9	-96	18	4.47	55	-51	9	6.56	0	18	24	7.41	39	33	30	9.04
sub08	15	-94	18	4.71	-12	-97	15	7.03	53	-46	32	18.90	-1	21	23	7.60	50	28	35	5.79
sub09	9	-89	12	10.13	-15	-93	21	11.30	51	-51	15	4.41	-9	18	30	6.55	49	29	35	4.09
sub10	11	-89	21	3.56	3	-93	9	14.07	48	-60	13	8.78	-9	18	27	7.36	43	34	34	4.94
sub11	18	-95	20	6.01	-12	-98	12	8.54	57	-54	15	2.84	-3	16	27	5.35	46	24	45	10.84
sub12	12	-96	12	9.44	-8	-99	20	2.16	53	-62	24	12.21	-6	21	32	3.53	43	29	43	6.82
sub13	12	-92	24	2.79	-6	-100	9	9.53	48	-56	13	6.37	-6	24	27	5.48	44	32	35	2.31
sub14	17	-89	19	6.65	-6	-98	12	6.60	48	-48	15	8.75	0	21	32	3.97	46	32	35	2.40
sub15	15	-97	14	8.64	0	-93	23	9.73	48	-63	18	11.11	0	24	24	6.98	39	27	30	9.10
sub16	12	-93	21	0.37	-3	-97	20	4.07	54	-54	21	6.17	-9	27	29	8.96	47	30	33	3.85
sub17	9	-90	21	4.59	-15	-87	24	16.01	51	-48	12	7.38	-2	15	32	6.80	39	24	37	8.74
sub18	18	-90	21	6.74	-14	-92	24	12.06	56	-59	18	6.30	1	14	35	9.07	43	28	41	5.30
sub19	9	-93	18	4.30	-6	-96	18	3.27	55	-48	8	9.34	0	27	33	7.33	44	34	37	4.01

Abbreviations: r/lCun, right/left cuneus; STS, superior temporal sulcus; mPFC, medial prefrontal cortex; dlPFC, dorsolateral prefrontal cortex. Distance unit in mm.