Neural mechanisms of eye contact when listening to another person talking

Jing Jiang,[1,2,3,*] Kamila Borowiak,[1,2] Luke Tudge,[2] Carolin Otto,[1] and Katharina von Kriegstein[1,3]

[1]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig 04103, Germany, [2]Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin 10117, Germany and [3]Institute of Psychology, Humboldt-Universität zu Berlin, Berlin 12489, Germany

Correspondence should be addressed to Jing Jiang, Stephanstr. 1a, Leipzig, 04103, Germany. Email: jiang@cbs.mpg.de

Abstract

Eye contact occurs frequently and voluntarily during face-to-face verbal communication. However, the neural mechanisms underlying eye contact when it is accompanied by spoken language remain unexplored to date. Here we used a novel approach, fixation-based event-related functional magnetic resonance imaging (fMRI), to simulate the listener making eye contact with a speaker during verbal communication. Participants' eye movements and fMRI data were recorded simultaneously while they were freely viewing a pre-recorded speaker talking. The eye tracking data were then used to define events for the fMRI analyses. The results showed that eye contact in contrast to mouth fixation involved visual cortical areas (cuneus, calcarine sulcus), brain regions related to theory of mind/intentionality processing (temporoparietal junction, posterior superior temporal sulcus, medial prefrontal cortex) and the dorsolateral prefrontal cortex. In addition, increased effective connectivity was found between these regions for eye contact in contrast to mouth fixations. The results provide first evidence for neural mechanisms underlying eye contact when watching and listening to another person talking. The network we found might be well suited for processing the intentions of communication partners during eye contact in verbal communication.

Key words: eye contact; verbal communication; fixation-based event-related fMRI; eye tracking; rapid event-related design

Introduction

Eye contact is a powerful visual cue for building social links between communicating partners. It has strong influences on several perceptual processes in communication, such as face detection, gender perception and facial expression recognition (for reviews, see Itier and Batty, 2009; Senju and Johnson, 2009; Madipakkam *et al.*, 2015; Schilbach, 2015). Impairments in eye contact are common and might significantly contribute to communication difficulties, e.g. in autism spectrum disorder (ASD) (Dalton *et al.*, 2005; Pelphrey *et al.*, 2005a), in schizophrenia (Choi *et al.*, 2010; Tso *et al.*, 2012) or in social anxiety disorder (Horley *et al.*, 2003; Schneier *et al.*, 2011).

Eye contact is an important means of non-verbal communication (for reviews, see Itier and Batty, 2009; Senju and Johnson, 2009; Schilbach, 2015), but also occurs very frequently during

verbal communication (Lewkowicz and Hansen-Tift, 2012; Macdonald and Tatler, 2013). Two important features characterize eye contact in verbal communication. First, eye contact occurs voluntarily while the partners are listening or talking to each other (Macdonald and Tatler, 2013). Second, eye contact is intermittent and of varying duration (Argyle and Dean, 1965), because communication partners freely and actively fixate on and scan between different regions of the partner's face. The two salient fixation regions are the eyes and the mouth (Vatikiotis-Bateson *et al.*, 1998; Jack *et al.*, 2009; Lusk and Mitchel, 2016). To our knowledge, currently no study has investigated the neural mechanisms underlying eye contact in scenarios that contain these two major features of eye contact in verbal communication. Thus, the first aim of this study was to identify brain regions and networks involved in eye contact with an

approach that simulates eye contact when watching and listening to a speaker talking.

Currently there are three accounts of the neural mechanisms of eye contact (for a review, see Senju and Johnson, 2009) (Figure 1). The first account, the 'affective arousal model' (Figure 1A), postulates that eye contact elicits responses in brain areas involved in arousal, particularly the amygdala (Kawashima *et al.*, 1999; for a review, see Senju and Johnson, 2009; Conty *et al.*, 2010; von dem Hagen *et al.*, 2014). The second account, the 'communicative intention detector model' (Figure 1B), assumes that eye contact signals the intention to communicate with others and involves the theory-of-mind (ToM) network (Kampe *et al.*, 2003; for reviews, see Senju and Johnson, 2009; von dem Hagen *et al.*, 2014), including posterior superior temporal sulcus and/or temporoparietal junction (pSTS&TPJ), medial prefrontal cortex (mPFC) and temporal pole (TP). The most recent account is the 'fast-track modulator model' (Figure 1C) (for a review, see Senju and Johnson, 2009). It proposes that a rapid subcortical visual processing route [including superior colliculus (SC), pulvinar (Pulv) and amygdala (Amy)] and a slow cortical visual processing route [including lateral occipital cortex (LOC) and inferior temporal cortex (ITC)] interact with brain regions of the so-called 'social brain network'. The social brain network comprises the Amy and orbitofrontal cortex (OFC) for emotion, pSTS and mPFC for intentionality, right anterior STS for gaze direction, and fusiform gyrus (FG) for face identity. In addition, the regions of the social brain network are modulated by the dorsolateral prefrontal cortex (dlPFC) according to task demands and social context. There is solid evidence that these accounts can explain neural mechanisms involved in eye contact in non-verbal situations in which participants view faces with averted or direct gaze (Calder *et al.*, 2002; Kampe *et al.*, 2003; Pelphrey *et al.*, 2004; von dem Hagen

et al., 2014; Cavallo *et al.*, 2015; Oberwelland *et al.*, 2016). To date, it is unknown whether the three accounts are also suitable to explain neural mechanisms underlying eye contact in verbal communication. Therefore, the second aim of the present study was to test the hypothesis that eye contact in verbal communication may rely at least partly on the mechanisms proposed in the three models.

We developed a novel experimental paradigm to simulate eye contact when listening to a speaker talking. We recorded videos of speakers talking about daily life topics with direct gaze at the camera. Participants were instructed to listen to the pre-recorded speakers and freely look at different regions of the speaker's face as they would do naturally. At random intervals the videos were stopped and participants were asked to report the last word they heard. We recorded functional magnetic resonance imaging (fMRI) and eye tracking data from the participants simultaneously. The fixations obtained from the eye tracking data were used to define the events for the fMRI data analysis. The feasibility of this fixation-based event-related (FIBER) fMRI approach has been successfully shown in recent studies on natural viewing behavior in visual scene perception (Marsman *et al.*, 2012; Henderson and Choi, 2015). In our study, this approach allowed us to distinguish when participants made voluntary eye contact with the speaker (i.e. fixated on the eyes region of the speaker whose gaze was directed towards the participant, 'Eyes' events) from when they looked at the mouth region ('Mouth' events) or other regions of the video ('Off' events). We assumed that the natural gaze patterns of viewing a speaker, fixation shifts between different regions of the speaker's face and variable fixation durations (Argyle and Dean, 1965; Henderson, 2011), would meet basic rules for rapid event-related designs (Burock *et al.*, 1998; Friston *et al.*, 1999).
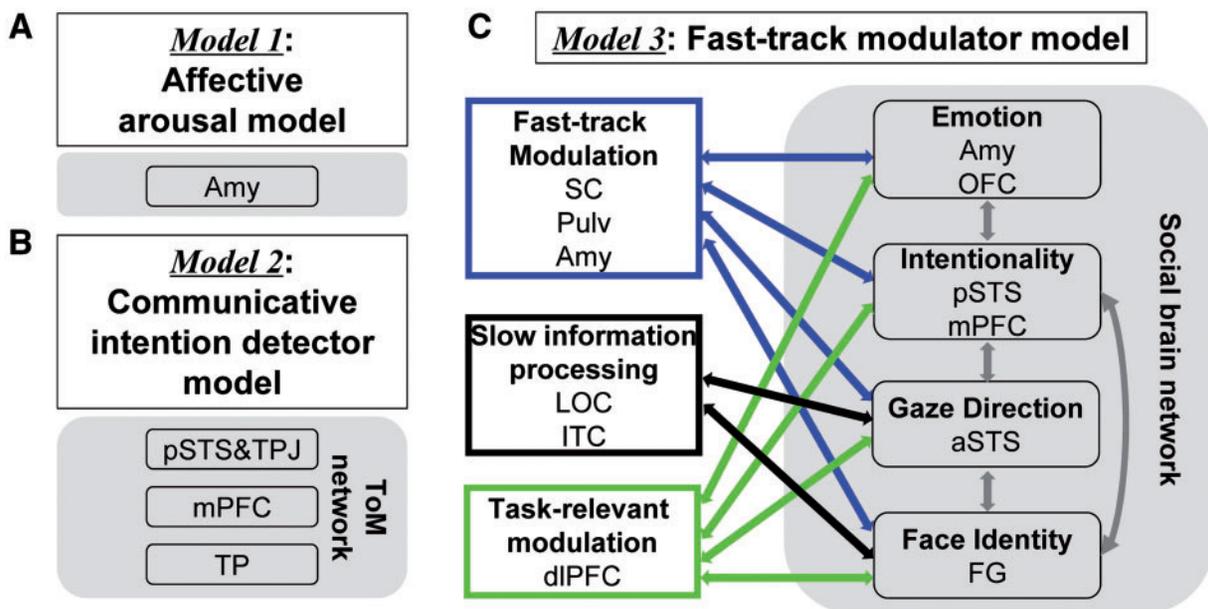

Fig. 1. Neuroscientific models of eye contact processing. **(A)** The affective arousal model postulates that eye contact elicits responses in the brain arousal system and/or emotional system, especially in the amygdala (Amy) (Kawashima *et al.*, 1999; Hooker *et al.*, 2003; Sato *et al.*, 2004). **(B)** The communicative intention detector model assumes that eye contact signals the intention to communicate with others and involves cerebral cortex regions of the ToM network. They are the pSTS&TPJ, the medial mPFC and the TP. This model (for review, see Senju and Johnson, 2009) is based on an earlier model proposed by Baron-Cohen (1997). **(C)** The fast-track modulator model (for review, see Senju and Johnson, 2009) proposes that eye contact is processed via a rapid and a slow information processing route. The rapid route (blue arrows) corresponds to a subcortical pathway involving SC, Pulv and Amy. Information processed in this route modulates processing in different regions of a so-called 'social brain network' including regions processing emotion (Amy; OFC), intentionality (pSTS and mPFC), gaze direction (right anterior STS, aSTS) and face identity (FG). At the same time these regions are modulated via the dlPFC according to task demands and context (green arrows). The slow information route is a visual cortical route including LOC and ITC that projects to regions analysing gaze direction and face identity (black arrows). Figure 1C is adapted from the model figure in Senju and Johnson (2009).

Materials and methods

Participants

In total, 30 healthy volunteers [15 female, 15 male; 27.5 ± 3.6 SD year-old; all right handers (Edinburgh questionnaire; Oldfield, 1971)] who reported normal vision without correction participated in this study. Written informed consent was provided by all participants. The study protocol was approved by the Research Ethics Committee of the University of Leipzig (AZ: 192-14-14042014). Nine participants were excluded due to difficulties with obtaining eye tracking data (e.g. difficulties with calibration before the experiment or eye tracking during the experiment). Two participants were excluded because of head movement in the MRI scanner (>3 mm). Furthermore, one subject's behavioral data were excluded because of technical problems with the response box. Therefore, eye tracking and fMRI data analyses were based on 19 subjects and behavioral data analyses were based on 18 subjects. All participants filled in the Autism-Spectrum Quotient questionnaire (Baron-Cohen *et al.*, 2001) before the fMRI experiment. All participants scored below the cut-off value (32) that is indicative of a manifestation of autistic traits typical for ASD (17.95 ± 4.99 SD).

Stimuli

We recorded 8 monologue videos from 2 female and 2 male German speakers (20, 22, 24 and 24-years–old, respectively). Each monologue lasted ∼6 min and was about daily life topics, such as a description of a typical week or how to learn a foreign language. All speakers had received speaking training in their study or career. They were asked to speak in a natural and emotionally neutral manner. Their naturalness and emotional content were rated by another five native German speakers (see Supplementary Materials). Eye contact is an event in which two partners look at each other's eyes at the same time (also called mutual gaze) (Schilbach, 2015). Here, we simulated such a situation by pre-recording the speakers in front view with their gaze constantly directed towards a person behind the camera (for recording details, see Supplementary Materials).

Our daily communications occur either in relatively quiet situations or noisy situations. Therefore, we made two types of videos corresponding to two conditions: the original recordings

with high signal-to-noise ratio (SNR) ('Normal' condition) and videos with low SNR ('Noise' condition) by mixing the audio tracks of the videos with background noise, (i.e. people talking and the clatter of dishes in a cafeteria). The Noise condition also allowed us to obtain balanced fixations between eyes and mouth. Because when watching and listening to speakers, listeners fixate more on the eyes. However, if the speech signal is noisy, fixations on the mouth of the speaker increase (Vatikiotis-Bateson *et al.*, 1998; Yi *et al.*, 2013). For video post-processing procedures, see Supplementary Materials.

Experimental procedure

The fMRI experiment consisted of four sessions. In each session one 'Normal' and one 'Noise' video were shown (Figure 2A). These two videos were from two different speakers of different genders. The orders of speaker gender, speaker identity and conditions were counterbalanced across sessions and participants. At the beginning of each video, the speaker's mouth was closed for 3 s. Between videos, a black screen was presented for 4 s to indicate a short pause. We instructed the participants to watch the speaker and to listen carefully. To make sure participants attended to what the speaker said, they performed a speech recognition task: The videos were stopped at random intervals and participants were asked 'Which is the last word you heard?' (Figure 2A). They chose the answer from 3 words listed on the screen by pressing one of three corresponding buttons on a response-box. The video continued once the participant pressed a button or after 4 s if no button was pressed. In total, 40 questions were asked, i.e. 5 in each monologue. Participants performed the speech recognition task with high accuracy in both conditions (see Supplementary Materials). Participants were instructed outside the MRI-scanner before the experiment. Between sessions, the participants were asked to close their eyes and rest for 1 min. The experiment was implemented in Presentation software (version 14.5, Neurobehavioral Systems Inc., USA).

Eye tracking

During fMRI scanning, participants' eye movements were recorded using a 120 Hz monocular MR-compatible eye tracker (EyeTrac 6, ASL, USA) simultaneously. The optical path was
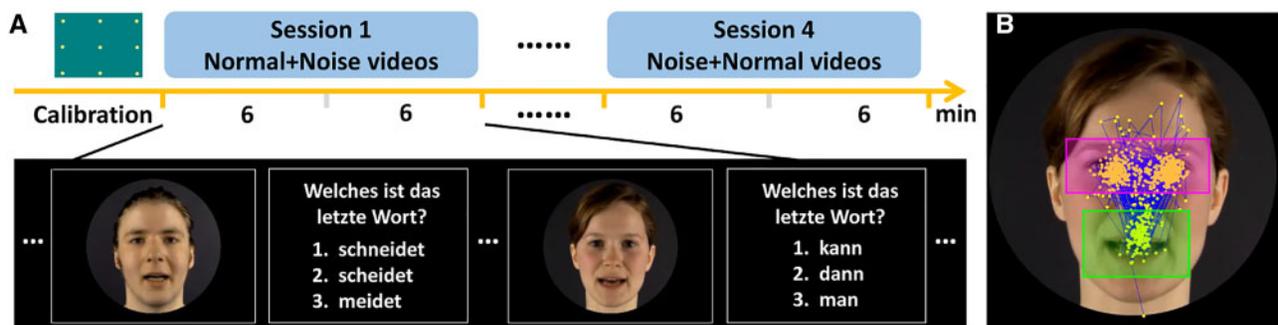

Fig. 2. Experimental procedure and the AOIs. **(A)** Before the experiment there was a nine-point calibration procedure for the eye tracking. In each session, participants viewed videos from two different speakers. The videos were stopped randomly and participants were asked 'Which is the last word you heard?' presented on the screen. They chose the answer from three words listed below the question. The experiment contained videos with auditory background noise ('Noise videos') and videos without noise ('Normal videos'). **(B)** Example of the AOIs that were used to define the events for the fMRI analysis. 1) Eyes (magenta rectangle): the left/right boundary of the rectangle was located 100 pixels to the left/right of the left/right pupil, the upper boundary 70 pixels above the pupils (near the upper border of the eyebrow), and the lower boundary 70 pixels below the pupils. 2) Mouth (green rectangle): the left and right boundaries of the rectangle were located 130 pixels left and right of the center of the mouth, the upper and lower boundaries 80 pixels above and below the center. We made the mouth AOI relatively large, because the size and shape of speaker's mouth changes during talking. 3) Off (all areas outside the two rectangles). Fixations are marked with yellow dots; saccade paths are represented by blue lines.

reflected over a mirror placed on top of the head coil in order to capture the eye image of the participant. Prior to the fMRI experiment, the eye tracking system was calibrated using a standard nine-point calibration procedure for each participant (Figure 2A). Before each session, the accuracy of eye tracking was checked. If necessary, the eye tracking system was recalibrated (ca. 1 time per participant).

fMRI data acquisition

Functional images and structural T1-weighted images were obtained using a 3 T Siemens Tim Trio MR scanner (Siemens Healthcare, Erlangen, Germany), equipped with a 12 channel head coil. A gradient-echo EPI (echo planar imaging) sequence was used for the functional MRI (TE 30 ms, flip angle 90°, TR 2.79 s, whole brain coverage with 42 slices, acquisition bandwidth 116 kHz, 2 mm slice thickness, 1 mm inter-slice gap, in-plane resolution 3 × 3 mm). Geometric distortions were characterized by a B0 field-map scan. For further details, see Supplementary Materials.

Eye tracking analysis

We used EyeNal software (ASL, USA) and customized MATLAB scripts for the eye tracking data analysis. A fixation was defined as having a minimum duration of 100 ms and a maximum visual angle change of 1°. Natural speaking is accompanied by head movements of the speaker. We therefore corrected the position of the participants' fixations based on the head movements of the speaker in the videos (for details see Supplementary Materials). This correction was necessary, because the eye tracker captures positions on the screen rather than real positions of the speaker's head. We defined two areas of interest (AOIs), i.e. eyes region and mouth region (see definition in Figure 2B). We labeled fixations falling inside the AOIs of eyes and mouth as 'Eyes' and 'Mouth' respectively, and fixations outside these AOIs as 'Off'. Fixations occurring consecutively within the same AOI were concatenated into one fixation, resulting in one event trial for the fMRI analyses. The event onset was the start time of the first fixation falling into the respective AOI.

Eye gaze patterns. To examine whether our design was suitable for fMRI analysis as a rapid event-related design, we checked the number of events (NE), the duration of the inter-event-interval (IEI) and the event order. First, we calculated the NE for each event type, to ensure there were a sufficient number of events to get a good and stable estimate of the hemodynamic response (≥ 25) and to obtain sufficient power (≥ 100) in the fMRI analyses (Huettel and McCarthy, 2001; Desmond and Glover, 2002). Second, we calculated the IEI (the time interval between the onset of one event and that of the next event), to check whether the mean IEI was of suitable length, i.e. around 2 s, which has been reported as optimal mean interval for rapid event-related designs (Dale and Buckner, 1997). Additionally, we checked whether the NE and IEI were balanced across conditions (Normal and Noise) and event types (Eyes, Mouth, Off) by performing a 2 × 3 repeated measures ANOVA for each index and *post-hoc* t-tests. The P values were Bonferroni-corrected for multiple comparisons between event types ($n = 3$). Finally, we tested whether the gaze patterns met two further recommendations for rapid event-related design, i.e. a jittered IEI and a variable event order (Burock *et al.*, 1998; Friston *et al.*, 1999). We computed the IEI distribution for each event type to illustrate that the IEI was jittered rather than fixed. We also checked

whether the events occur in a variable order by calculating the transition probability of every possible combination of event order.

Pupil diameter. In addition, we measured pupil dilation as a measure of arousal (Libby *et al.*, 1973; Kampe *et al.*, 2003). To examine whether eye contact leads to larger pupil diameter than mouth or off fixations, we computed pupil diameter for each participant and compared it between event types using a one-way ANOVA.

Potential confounding variables. The different shapes and sizes of the AOIs (see Figure 2B) might influence the saccade distance between sub-fixations and the number of saccades (NS) within each event, thus confounding the fMRI results. To test whether this was the case, we calculated the distance between sub-fixations as inter-fixation distance (IFD) in degrees of visual angle and the NS during each event. We then compared IFD and NS between event types using a one-way ANOVA.

fMRI analysis

Analyses of BOLD responses. All fMRI analyses were performed using Statistical Parametric Mapping software (SPM8, Wellcome Trust Centre for Neuroimaging, UCL, UK, http://www.fil.ion.ucl. ac.uk/spm). We performed standard pre-processing procedures (see Supplementary Materials). At first-level, the general linear model (GLM) analysis included six events of interest that were defined by the eye tracking events (Eyes_normal, Eyes_noise, Mouth_normal, Mouth_noise, Off_normal, Off_noise) and seven regressors of no interest (speech recognition task, 6 spatial movement parameters estimated during realignment). The onset of the speech recognition task regressor corresponded to when the question started to appear (Figure 2A). All events were modeled with a duration of 0 s (referred to as 'standard GLM').

Because there were significant differences between event types on both IFD (Eyes = 1.20°, Mouth = 0.52°, Off = 0.27°; $F(2,17) = 57.95$, $P < 0.001$) and NS (Eyes = 2.86, Mouth = 1.39, Off = 0.21; $F_{(2,17)} = 22.19$, $P < 0.001$), we conducted a 'control' GLM analysis. In this GLM, IFD and NS were entered as two additional parametric modulators for each event type into the same design matrix as that of the standard GLM. Each event had one corresponding IFD value and one corresponding NS value.

To identify the brain regions involved in eye contact when watching another person talking, we computed at the first-level the main effect 'Eyes *vs* Mouth' for both 'standard' and 'control GLM'. In addition, we conducted the simple main effects 'Eyes_noise *vs* Mouth_noise' and 'Eyes_normal *vs* Mouth_normal' and other contrasts (Eyes *vs* Off and Mouth *vs* Off) for 'standard GLM'. The contrast maps obtained from the first-level analysis were entered into second-level random-effects analyses using one-sample t-tests. The SPM statistical maps were thresholded at voxel-wise $P < 0.01$ (for main effects) or $P < 0.05$ (for simple main effects) with a cluster-wise family-wise error (FWE) correction of $P < 0.05$ for the whole brain. We labeled the cluster locations based on anatomical information provided by the Brodmann areas (BAs) and Automated Anatomical Labeling atlas (Tzourio-Mazoyer *et al.*, 2002) and/or probabilistic cytoarchitectonic maps (Eickhoff *et al.*, 2005).

Region of interest-based psychophysiological interactions analysis. The fast-track modulator model (Figure 1C) makes not only predictions about regions involved in eye contact processing, but also about the connectivity between these regions. To test such connectivity in our data set, we conducted psychophysiological interaction (PPI) analyses (Friston *et al.*, 1997) with SPM8. We extracted the physiological variable (first

Eigenvariate) from all regions included in the fast-track modulator model (see Figures 1C and 3). In addition, we included the cuneus in the PPI analyses; it is not part of the fast track modulator model, but showed highly significant responses for the Eyes *vs* Mouth contrast in the present study (see 'Results' section).We defined these regions (Figure 3) either functionally, or if that was not possible, anatomically (see Supplementary Materials).

The first-level GLM analysis included the interaction term (physiological × psychological variable), the physiological variable and the psychological variable (Eyes *vs* Mouth contrast) and was conducted for each source region, i.e. the region of Eigenvariate extraction. To test whether these regions showed enhanced connectivity with each other during eye contact in contrast to mouth fixation, we performed region of interest (ROI) analyses for all regions in the fast track modulator model and the cuneus (target ROIs). The target ROIs were again defined either functionally or anatomically (see Supplementary Materials). They were used to correct for multiple comparisons at the voxel-wise FWE of $P < 0.05$ using small-volume corrections in the second-level analyses.

Results

Eye gaze patterns

NE and IEI across conditions. The average NE across participants was 614.89 (Eyes), 521.21 (Mouth) and 380.37 (Off) (Figure 4A,

Supplementary Table S1). Also each participant had > 100 events for Eyes and Mouth (Supplementary Table S1). The mean IEI across participants was 2.44s (Eyes), 1.65s (Mouth), 0.62s (Off) (Figure 4B, Supplementary Table S2). There were significant main effects of the event types on both NE [$F_{(2,36)} = 11.27$, $P < 0.001$] and IEI [$F_{(2,36)} = 21.84$, $P < 0.001$]. Both Eyes and Mouth were significantly more and longer than the Off events (NE: Eyes >
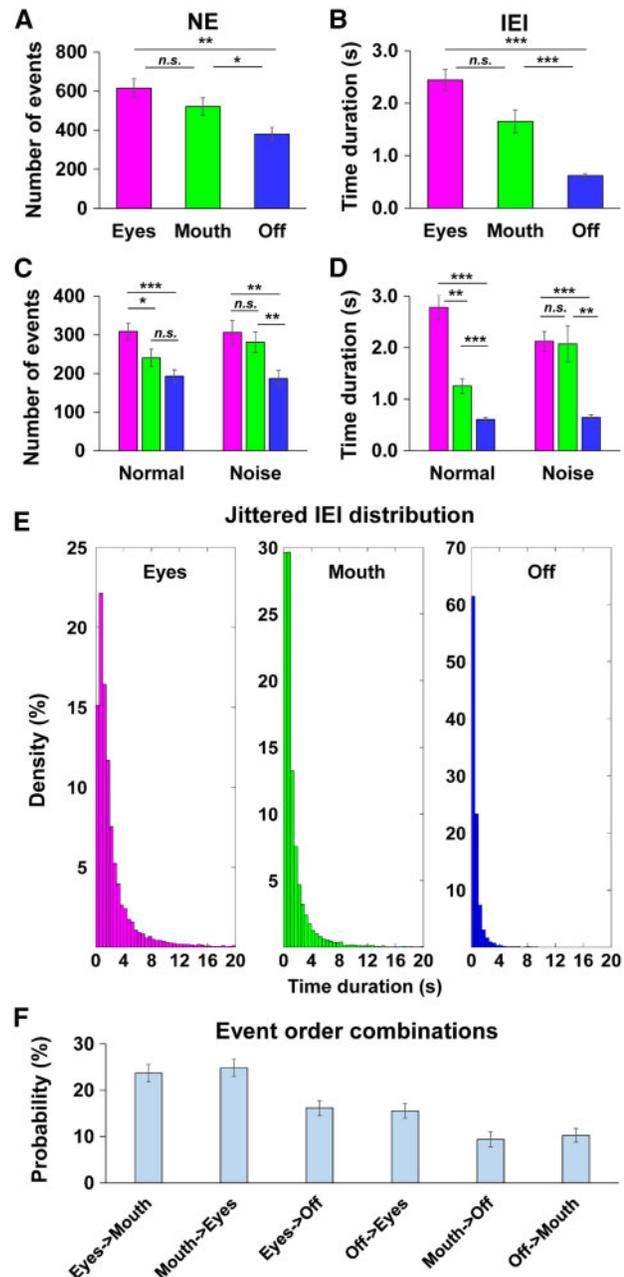
Fig. 4. Eye gaze patterns during freely viewing the monologue videos. **(A,B)** The NE and IEI pooled over the normal and noise condition. **(C,D)** NE and IEI for the normal and noise condition separately. The IEI is defined by the time interval between the onset of one event and that of the next one. Error bars display the standard error of the mean. **(E)**Density histograms indicating jittered IEI distribution for Eyes, Mouth and Off pooled over the whole experiment. Eyes, Mouth and Off represent fixations falling into eyes region, mouth region and outside of these regions respectively.**(F)**Probability of all possible combinations of event orders, i.e., Eyes->Mouth means the fixation shifts from eyes region to mouth region.
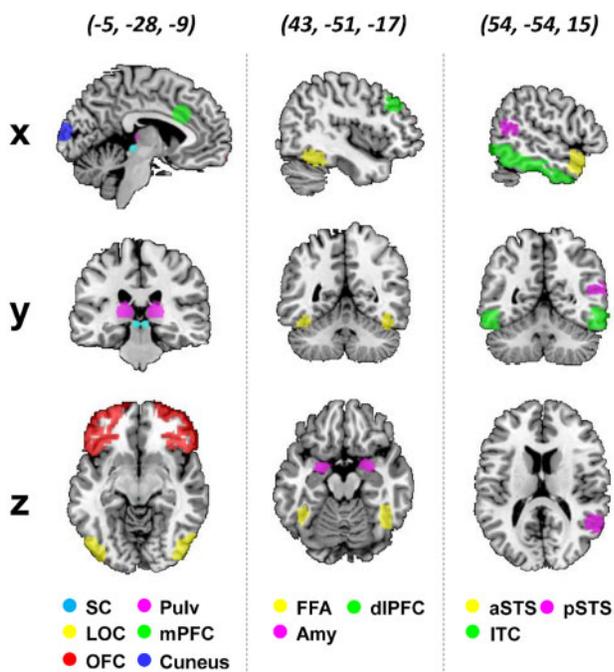

Fig. 3. Regions of interest. SC, superior colliculus; Pulv, pulvinar; LOC, lateral occipital cortex; OFC, orbitofrontal cortex; mPFC, medial prefrontal cortex; FFA, fusiform face area; dlPFC, dorsolateral prefrontal cortex; Amy, amygdala; a/pSTS, anterior/posterior superior temporal sulcus; ITC, inferior temporal cortex. The figure displays the target regions at which we conducted small volume correction for the PPI analyses. Of these, cuneus, pSTS, mPFC and dlPFC were 10-mm-radius spheres centered at maximum peak coordinates in group GLM analysis and were restricted to the gray matter within the brain; others were standard anatomical maps or customized anatomical masks (see Supplementary Materials). For source regions, we used the same regions as above, except that for cuneus, pSTS, mPFC and dlPFC, we used the subject-specific coordinates in the GLM analysis (Supplementary Table S1).

Off, $t = 4.15$, $P = 0.002$; Mouth > Off, $t = 2.85$, $P = 0.032$; IEI: Eyes > Off, $t = 8.42$, $P < 0.001$, Mouth > Off, $t = 4.94$, $P < 0.001$) (Figure 4A and B). However, there was no significant difference between Eyes and Mouth in the NE ($t = 2.21$, $P = 0.120$) or the IEI ($t = 2.13$, $P = 0.139$), indicating these two indexes were roughly balanced between event types.

NE and IEI within condition. The average NE across participants was again well above 100 for all event types in the noise and normal condition (Figure 4C, Supplementary Table S1). There were significant event types × conditions interactions [NE: $F_{(2,36)} = 7.21$, $P = 0.002$, Figure 4C; IEI: $F_{(2,36)} = 20.83$, $P < 0.001$, Figure 4D]. Eyes and Mouth events differed significantly in the Normal condition (NE: $t = 3.39$, $P = 0.010$; IEI: $t = 4.68$, $P < 0.001$) (Figure 4C and D, Supplementary Table S2), but were balanced in the Noise condition as expected (NE: $t = 1.11$, $P = 0.852$; IEI: $t = 0.11$, $P = 1.000$) (Figure 4C and D, Supplementary Table S2).

IEI distribution and event order. The IEI distribution showed a jittered duration ranging from 0.1 to 80.1 s for all the events. The main part of the range (0.1–20 s) was plotted in Figure 4E. The order of events (fixations on different regions) was variable. All combinations of event orders were used and the transition probability for each combination ranged from 10 to 25%, which is close to 16.7% (equal chance for a strict randomized order) (Figure 4F).

Taken together, the gaze patterns showed that our design met the requirements for a rapid event-related design: There were sufficient NE and suitably long IEI for the Eyes and Mouth events and both indices were roughly balanced between event types, particularly across conditions and in the Noise condition.

In addition the IEI was jittered and the events occurred in a variable order.

Pupil diameter

We did not find significant differences in average pupil diameter in response to Eyes fixation and Mouth fixation ($P = 0.892$) or Off fixation ($P = 0.95$), indicating no increased arousal during eye contact as compared to fixating on the mouth when people watch a speaker talking.

Brain regions involved in eye contact *vs* Mouth fixation

When participants fixated on the eyes of the speaker in contrast to the mouth (Eyes > Mouth), we found increased responses in a large brain network (Figure 5A, Table 1): (i) visual cortices including cuneus (Cun, BA 17/18), bilateral calcarine sulcus (Cal, BA 17/18) covering V1, V2 and V3 and extending to bilateral precuneus (Prec, BA 7), (ii) ToM-related brain regions: the right temporal-parietal junction including angular gyrus and supra-marginal gyrus and extending into the posterior superior temporal sulcus (TPJ&pSTS, BA 39/40), mPFC including the anterior cingulate cortex extending to medial orbital frontal cortex (BA 10/24/32) and (iii) dlPFC (BA 9/46) (FWE cluster-wise corrected, $P < 0.05$). Results for the 'standard GLM' (Figure 5A) and 'control GLM' (Figure 5B) were qualitatively the same, indicating little effect of the distance or NS within the event on the fMRI results.

To check whether the differential responses between Eyes and Mouth events were caused by a higher NE and longer duration of eye contact during the normal condition, we analysed
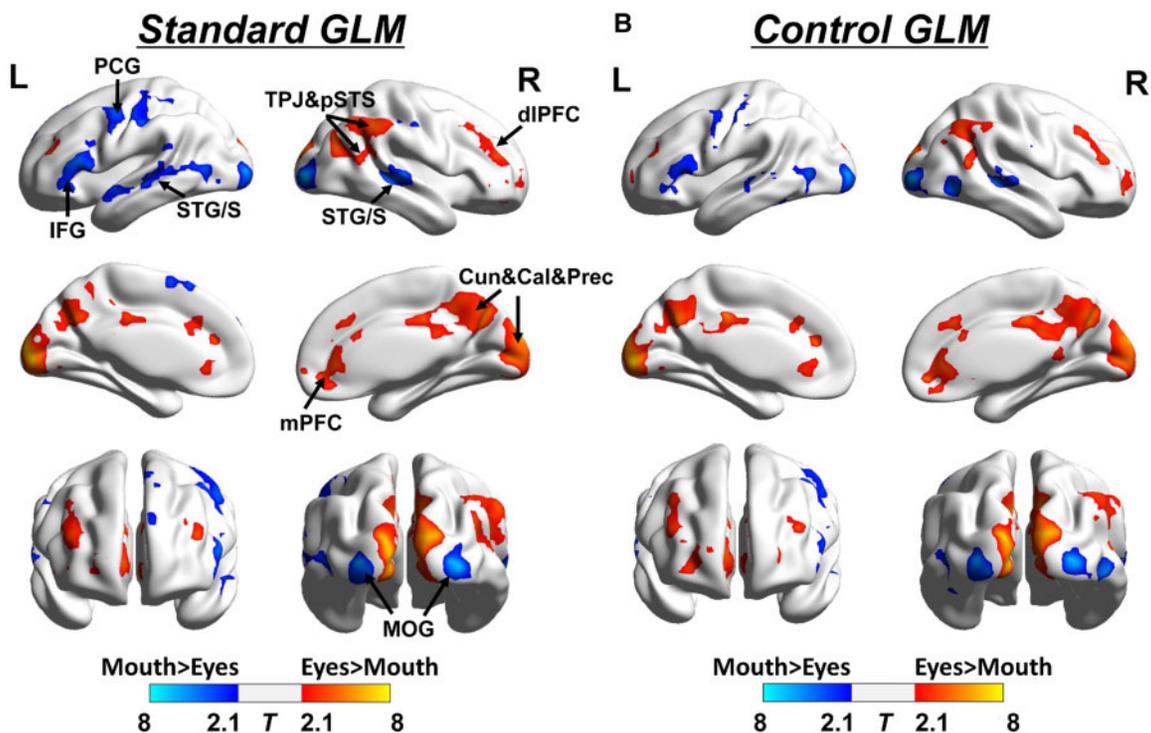
Fig. 5 Brain regions showing BOLD response differences between Eye contact and Mouth fixation. **(A)** Standard GLM; **(B)** Control GLM (including IFD and the NS as additional regressors). Hot colors (red to yellow) indicate brain areas showing higher response to Eye contact than mouth fixation. Cold colors (blue to cyan) showed higher response to Mouth fixations than eye contact. L, left hemisphere, R, right hemisphere. For visualization purposes only, voxels surviving with a voxel-level threshold of $P < 0.05$ and 160 voxels are shown. MNI coordinates of significant brain regions are listed in Table 1. 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.

Table 1. Coordinates and *p*-values for brain regions showing significant response differences in the Eyes *vs* Mouth contrasts

Region	Side	P-value (FWE corrected)	cluster volume (mm³)	T value	MNI coordinates			BA
					x	y	z	
Eyes > Mouth								
Cun&Cal&Prec	B	0.000	1729	8.37	−6	−99	18	17/18/7
				6.93	12	−93	21	
				4.64	−3	−72	39	
TPJ&pSTS	R	0.000	276	6.13	39	−57	30	39/40/42
				5.45	51	−42	33	
				3.47	54	−54	15	
mPFC	B	0.000	255	4.98	−3	21	30	10/24/32
				4.06	6	42	−3	
dlPFC	R	0.016	106	4.19	45	30	36	9/46
Mouth > Eyes								
MOG	R	0.013	127	6.80	33	−90	3	18
MOG	L	0.011	131	6.34	−30	−96	6	18
STG/S	R	0.003	169	5.83	54	−33	6	21/22
IFG	L	0.002	175	5.63	−48	15	9	44/45
PCG	L	0.010	133	5.23	−45	−3	51	4/6
STG/S	L	0.009	98	5.08	−63	−30	3	21/22

Threshold: voxel-level *P* < 0.01, *k* > 50 voxels, FWE cluster-corrected *P* < 0.05 across whole brain.
Abbreviations: p, posterior; Cun, cuneus; Cal, calcarine; Prec, precuneus; TPJ, temporoparietal junction; STS/G, superior temporal sulcus/gyrus; mPFC, medial prefrontal cortex; dlPFC, dorsolateral prefrontal cortex; MOG, middle occipital cortex; IFG, inferior frontal gyrus; PCG, precentral and/or postcentral gyrus; L, left hemisphere; R, right hemisphere; B, bilateral hemispheres.

the simple main effect of eye contact in the normal condition and the noise condition separately. The response patterns for eye contact in the two conditions (Supplementary Figure S1A and B, Supplementary Table S3) were similar to that obtained in the main effects, except that the response in the right dlPFC was absent in the noise condition after FWE correction, but was present at the threshold of P = 0.001 uncorrected. Results for the contrasts Eyes *vs* Off are displayed in Supplementary Figure S1C and listed in Supplementary Table S4.

Brain regions involved in mouth fixation vs Eye contact. For completeness, we here also report regions that were engaged in mouth fixation in contrast to eye fixation. These included (i) bilateral middle occipital gyrus (MOG, BA 18), (ii) bilateral superior temporal gyrus/sulcus (STG/S, BA 21/22), (iii) left inferior frontal gyrus (IFG, BA 44/45) and (iv) left precentral and postcentral gyrus (PCG, BA4/6) (FWE cluster-wise corrected, P < 0.05) (Figure 5 and Table 1). This network of brain regions overlapped well with previous reports on brain mechanisms supporting audiovisual speech perception and lip reading (Skipper *et al.*, 2007; Blank and von Kriegstein, 2013; for a review, see Erickson *et al.*, 2014). Results for the contrast Mouth *vs* Off are displayed in Supplementary Figure S1D and listed in Supplementary Table S4.

Effective connectivity (EC) between brain regions

Our PPI results were consistent with several of the predictions of the fast track modulator model (Figure 6, Supplementary Table S5). There were significant EC increases for eye contact *vs* mouth fixation (i) between the rapid (SC, Pulv, Amy) and slow (LOC, ITC) routes of visual processing and parts of the social brain network; (ii) between the dlPFC and regions involved in intentionality processing (pSTS and mPFC); and (iii) within the social brain network (Figure 6A and B, Supplementary Table S5). However, some of the predicted connectivity was not found (see Figure 6B, dashed arrows). For example, the dlPFC did not show EC increase to emotion, gaze direction, and face identity

processing regions. In addition, we also found EC not predicted by the fast track modulator model: First, there was strong connectivity between the slow visual processing systems and the fast visual processing systems, the dlPFC, and other parts of the social brain network (Figure 6B, cyan solid arrows). Second, we also found that the visual cortex (i.e. the cuneus) that was not part of the fast-track modulator model showed EC with most of the brain regions in the fast-track modulation model (Figure 6B, pink solid arrows). Note that the effective connectivity here did not reflect real direction of influence between regions.

Discussion

We used a novel paradigm to explore the neural mechanisms of eye contact when listening to another speaker talking. We identified a network of brain regions that showed higher responses when looking at the eyes than the mouth of a speaker. The network included visual cortices, brain regions that have been related to ToM/intentionality processing (TPJ&pSTS and mPFC), and the dlPFC. Effective connectivity was enhanced between regions within this network, as well as further regions mainly involved in visual processing and regions that serve functions in social settings, i.e. the so-called social brain network (Figures 1C and 6B).

Several of the findings were in agreement with the present models of eye contact, particularly the communicative intention detector model and the fast-track modulator model (for a review, see Senju and Johnson, 2009), which both assume that the ToM/intentionality processing network (TPJ&pSTS and mPFC) is involved in eye contact. The validity of the predictions of the models in the context of verbal communication was not self-evident. The way of eliciting eye contact in the current study differed in multiple aspects from the ways in previous studies (Conty *et al.*, 2007; Ethofer *et al.*, 2011; von dem Hagen *et al.*, 2014; Cavallo *et al.*, 2015; Oberwelland *et al.*, 2016) on which the models are largely based. First, in our study eye contact occurred in a verbal listening context. Second, it was elicited by the participants themselves rather than by stimuli with averted
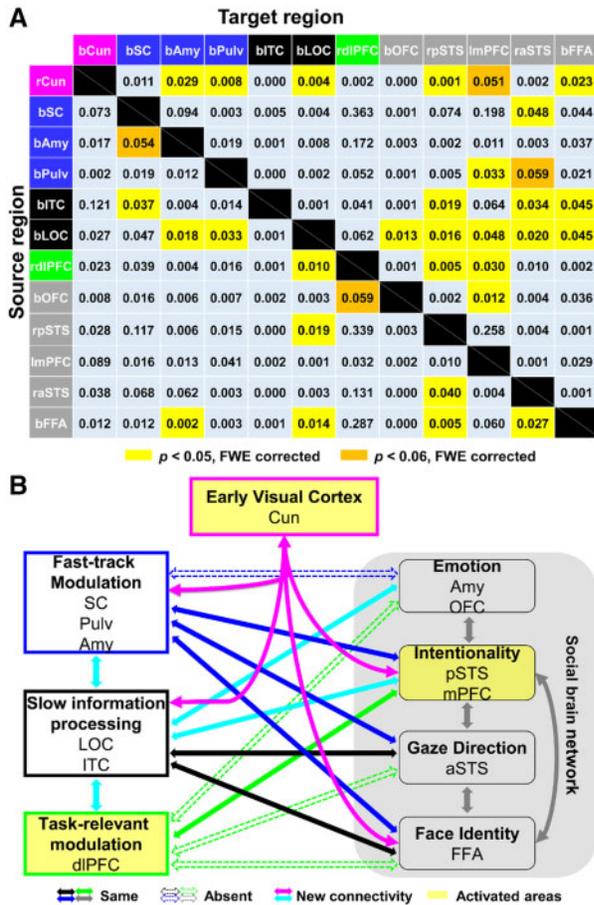
A Target region

Source region	bCun	bSC	bAmy	bPulv	bITC	bLOC	rdlPFC	bOFC	rpSTS	lmPFC	raSTS	bFFA
rCun		0.011	0.029	0.008	0.000	0.004	0.002	0.000	0.001	0.051	0.002	0.023
bSC	0.073		0.094	0.003	0.005	0.004	0.363	0.001	0.074	0.198	0.048	0.044
bAmy	0.017	0.054		0.019	0.001	0.008	0.172	0.003	0.002	0.011	0.003	0.037
bPulv	0.002	0.019	0.012		0.000	0.002	0.052	0.001	0.005	0.033	0.059	0.021
bITC	0.121	0.037	0.004	0.014		0.001	0.041	0.001	0.019	0.064	0.034	0.045
bLOC	0.027	0.047	0.018	0.033	0.001		0.062	0.013	0.016	0.048	0.020	0.045
rdlPFC	0.023	0.039	0.004	0.016	0.001	0.010		0.001	0.005	0.030	0.010	0.002
bOFC	0.008	0.016	0.006	0.007	0.002	0.003	0.059		0.002	0.012	0.004	0.036
rpSTS	0.028	0.117	0.006	0.015	0.000	0.019	0.339	0.003		0.258	0.004	0.001
lmPFC	0.089	0.016	0.013	0.041	0.002	0.001	0.032	0.002	0.010		0.001	0.029
raSTS	0.038	0.068	0.062	0.003	0.000	0.003	0.131	0.000	0.040	0.004		0.001
bFFA	0.012	0.012	0.002	0.003	0.001	0.014	0.287	0.000	0.005	0.060	0.027	

p < 0.05, FWE corrected   p < 0.06, FWE corrected

Fig. 6. Effective connectivity results. (**A**) *P* values for effective connectivity between source regions (for physiological variable extraction) and target regions (for small volume correction) for eye contact in contrast to mouth fixation. Values marked with yellow and orange represent P < 0.05 and 0.06, respectively, with FWE correction for multiple comparisons. Values in light blue represent uncorrected *P* values. Note that, since the findings for right and left cuneus as source region were qualitatively similar, only the findings of right one are displayed here. In addition, for all the target regions, only the hemisphere with the highest *P* value is shown here. For corresponding coordinates see Supplementary Table S4. b, bilateral; r, right; l, left. B, Comparison of effective connectivity found in the current study and that predicted in the fast-track modulator model. For easier comparison, the colors of arrows here are the same as those in Figure 1C. Solid arrows represent connectivity found in the PPI analyses that were predicted by the fast-track modulator model. Dashed arrows represent connectivity predicted in the model but not found in our results. Pink and cyan solid arrows represent new connectivity found in this study, but not predicted by the model. Brain areas marked with yellow are the areas found to be significantly more responsive to eye contact *vs* mouth fixation in the standard GLM analysis.

and direct gaze. These differences between our approach and designs in previous studies might explain several findings in our study that differed from the models' predictions.

As predicted by the fast-track modulator model, the dlPFC was involved in eye contact in contrast to mouth fixation and had particularly strong connectivity with the pSTS and mPFC, which have been implicated in ToM/intentionality processing (Saxe *et al.*, 2004; Gao *et al.*, 2012), but not with the other regions of the social brain network. We speculate that this pattern of results can be explained by the specific nature of listening to the speaker in an emotionally neutral audiovisual verbal context. Our design emphasized the understanding of what was said in a relatively long listening situation, but not face identity,

emotion or gaze direction processing. The fast-track modulator model assumes that the dlPFC has the role of modulating the social brain network (see Figure 1C) depending on the context. Our findings suggest that the dlPFC might be used to modulate regions involved in ToM/intentionality processing, when we look into the eyes of a speaker we are listening to.

Some studies have suggested that eye contact elicits arousal/emotional responses (Kawashima *et al.*, 1999; Conty *et al.*, 2010; von dem Hagen *et al.*, 2014), an important feature of the affective arousal model (for a review, see Senju and Johnson, 2009). However, in what situations an arousal/emotional response to eye contact occurs is currently unclear (Mormann *et al.*, 2015; for a review, see Hamilton, 2016). We found no differences in average pupil diameter, which is considered a reliable measure of arousal (Libby *et al.*, 1973; Kampe *et al.*, 2003). There was also no amygdala response to eye contact. These results suggest that eye contact initiated by the listener may not act as an arousal cue in the context of listening to someone talking about a neutral topic in a neutral manner.

A major difference between the predictions of current neuroscientific models of eye contact processing and our findings was the strong responses in visual cortices, i.e. responses in cuneus/calcarine sulcus to eye contact and in MOG to mouth fixation. It is difficult to explain these responses as the result of larger IFDs or NS within the eye-region, because the control GLM (with IFD and NS as parametric modulators) still showed qualitatively the same results as the standard GLM. The differential responses in early visual areas might be explained by the different information present at foveal vision. The features (eyes, mouth) and the amount of movement (constant gaze, mouth movement) naturally differ at foveal vision for the two event types. We speculate that the feature difference might not play a major role, because previous fMRI studies comparing responses to eyes and mouth stimuli did not report similar responses in early visual areas as in this study (Puce *et al.*, 1998; Pelphrey *et al.*, 2005b; Liu *et al.*, 2010; Arcurio *et al.*, 2012). In contrast, studies investigating responses to moving *vs* static faces/lips have revealed similar MOG (BA18) responses as the Mouth *vs* Eyes contrast in the present study (Calvert and Campbell, 2003; Schultz *et al.*, 2012). This might be a first indication that it is the difference between movement information at the fovea that leads to the differential responses in cuneus and MOG.

Eye tracking and fMRI have primarily been utilized as independent techniques to investigate the gaze patterns or neural mechanisms involved in eye contact processing (for a review, see Schilbach *et al.*, 2012). The current study provides evidence that the FIBER-fMRI method (Marsman *et al.*, 2012; Henderson and Choi, 2015) is an innovative and promising technique balanced between ecological validity and methodological constraints for investigating eye contact in verbal communication. This technique can shed light on the relatively unexplored question of how our brain 'sees' the complex and dynamic world in a voluntary manner. Our experimental paradigm involved several important features of eye contact during verbal communication. However, it missed some other features that are often present in genuinely reciprocal face-to-face communication between partners, such as eye gaze shifts of the talker, verbal turn-taking and the knowledge of interacting with a real partner (Vertegaal *et al.*, 2001; Wilson and Wilson, 2005; Jiang *et al.*, 2012, 2015). We expect that our experimental paradigm together with other interactive approaches (e.g. Oberwelland *et al.*, 2016) will be a solid foundation for integrating these features in future studies.

In this study, FIBER-fMRI allowed us to investigate the neural mechanisms underlying eye contact with a paradigm that is

ecologically valid for situations in which we listen to another person talking. The network that we found for eye contact might be well suited for processing the intentions of communication partners when listening to them during face-to-face verbal communication.

Funding

Supplementary data

Supplementary data are available at *SCAN* online.

Conflict of interest. None declared.

References

Arcurio, L.R., Gold, J.M., James, T.W. (2012). The response of face-selective cortex with single face parts and part combinations. *Neuropsychologia*, **50**(10), 2454–9.

Argyle, M., Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry*, **28** 289–304.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, **31**(1), 5–17.

Blank, H., von Kriegstein, K. (2013). Mechanisms of enhancing visual–speech recognition by prior auditory information. *Neuroimage*, **65**, 109–18.

Burock, M.A., Buckner, R.L., Woldorff, M.G., Rosen, B.R., Dale, A.M. (1998). Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *Neuroreport*, **9**(16), 3735–9.

Calder, A.J., Lawrence, A.D., Keane, J., et al. (2002). Reading the mind from eye gaze. *Neuropsychologia*, **40**(8), 1129–38.

Calvert, G., Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. *Journal of Cognitive Neuroscience*, **15**(1), 57–70.

Cavallo, A., Lungu, O., Becchio, C., Ansuini, C., Rustichini, A., Fadiga, L. (2015). When gaze opens the channel for communication: Integrative role of IFG and MPFC. *Neuroimage* **119**, 63–9.

Choi, S.H., Ku, J., Han, K., et al. (2010). Deficits in eye gaze during negative social interactions in patients with schizophrenia. *Journal of Nervous and Mental Disorders*, **198**(11), 829–35.

Conty, L., N'Diaye, K., Tijus, C., George, N. (2007). When eye creates the contact! ERP evidence for early dissociation between direct and averted gaze motion processing. *Neuropsychologia*, **45**(13), 3024–37.

Conty, L., Russo, M., Loehr, V., et al. (2010). The mere perception of eye contact increases arousal during a word-spelling task. *Social Neuroscience*, **5**(2), 171–86.

Dale, A.M., Buckner, R.L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, **5**(5), 329–40.

Dalton, K.M., Nacewicz, B.M., Johnstone, T., et al. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*, **8**(4), 519–26.

Desmond, J.E., Glover, G.H. (2002). Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. *Journal of Neuroscience Methods*, **118**(2), 115–28.

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.

Erickson, L.C., Heeg, E., Rauschecker, J.P., Turkeltaub, P.E. (2014). An ALE meta-analysis on the audiovisual integration of speech signals. *Human Brain Mapping*, **35**(11),5587–605.

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

Friston, K., Buechel, C., Fink, G., Morris, J., Rolls, E., Dolan, R. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, **6**(3), 218–29.

Friston, K.J., Zarahn, E., Josephs, O., Henson, R., Dale, A.M. (1999). Stochastic designs in event-related fMRI. *Neuroimage*, **10**(5), 607–19.

Gao, T., Scholl, B.J., McCarthy, G. (2012). Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus. *The Journal of Neuroscience*, **32**(41), 14276–80.

Hamilton, A. F d C. (2016). Gazing at me: the importance of social meaning in understanding direct-gaze cues. *Philosophical Transactions of the Royal Society B*, **371**(1686), 20150080.

Henderson, J. (2011). Eye movements and scene perception. In: Liversedge, S., Gilchrist, I. D., Everling, S., editors. *The Oxford Handbook of Eye Movements*, pp. 593–606. Oxford: Oxford University Press.

Henderson, J.M., Choi, W. (2015). Neural Correlates Of Fixation Duration During Real-World Scene Viewing: Evidence From Fixation-related (FIRE) fMRI'. *Journal of Cognitive Neuroscience*, **27**(6), 1137–45.

Horley, K., Williams, L.M., Gonsalvez, C., Gordon, E. (2003). Social phobics do not see eye to eye: a visual scanpath study of emotional expression processing. *Journal of Anxiety Disorders*, **17**(1), 33–44.

Huettel, S.A., McCarthy, G. (2001). The effects of single-trial averaging upon the spatial extent of fMRI activation. *Neuroreport*, **12**(11), 2411–6.

Itier, R.J., Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neuroscience and Biobehavioral Reviews*, **33**(6), 843–63.

Jack, R.E., Blais, C., Scheepers, C., Schyns, P.G., Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, **19**(18), 1543–8.

Jiang, J., Chen, C., Dai, B., et al. (2015). Leader emergence through interpersonal neural synchronization. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(14), 4274–9.

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., Lu, C. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience*, **32**(45), 16064–9.

Kampe, K.K., Frith, C.D., Frith, U. (2003). "Hey John": signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality. *The Journal of Neuroscience*, **23**(12), 5258–63.

Kawashima, R., Sugiura, M., Kato, T., et al. (1999). The human amygdala plays an important role in gaze monitoring. *Brain*, **122**(4), 779–83.

Lewkowicz, D.J., Hansen-Tift, A.M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United states of America*, **109**(5), 1431–6.

Libby, W.L., Lacey, B.C., Lacey, J.I. (1973). Pupillary and cardiac activity during visual attention. *Psychophysiology* **10**(3), 270–94.

Liu, J., Harris, A., Kanwisher, N. (2010). Perception of face parts and face configurations: an fMRI study. *Journal of Cognitive Neuroscience*, **22**(1), 203–11.

Lusk, L.G., Mitchel, A.D. (2016). Differential gaze patterns on eyes and mouth during audiovisual speech segmentation. *Frontiers in Psychology*, **7**, 52.

Macdonald, R.G., Tatler, B.W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision*, **13**(4), 6.

Madipakkam, A.R., Rothkirch, M., Guggenmos, M., Heinz, A., Sterzer, P. (2015). Gaze direction modulates the relation between neural responses to faces and visual awareness. *The Journal of Neuroscience*, **35**(39), 13287–99.

Marsman, J.B.C., Renken, R., Velichkovsky, B.M., Hooymans, J.M., Cornelissen, F.W. (2012). Fixation based event-related fmri analysis: using eye fixations as events in functional magnetic resonance imaging to reveal cortical processing during the free exploration of visual images. *Human Brain Mapping*, **33**(2), 307–18.

Mormann, F., Niediek, J., Tudusciuc, O., *et al.* (2015). Neurons in the human amygdala encode face identity, but not gaze direction. *Nature Neuroscience*, **18**(11), 1568–70.

Oberwelland, E., Schilbach, L., Barisic, I., *et al.* (2016). Look into my eyes: Investigating joint attention using interactive eye-tracking and fMRI in a developmental sample. *Neuroimage*, **130**, 248–60.

Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory'. *Neuropsychologia*, **9**(1), 97–113.

Pelphrey, K.A., Morris, J.P., McCarthy, G. (2005a). Neural basis of eye gaze processing deficits in autism. *Brain*, **128**(5), 1038–48.

Pelphrey, K.A., Morris, J.P., Michelich, C.R., Allison, T., McCarthy, G. (2005b). 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.

Pelphrey, K.A., Viola, R.J., McCarthy, G. (2004). When strangers pass processing of mutual and averted social gaze in the superior temporal sulcus. *Psychological Science*, **15**(9), 598–603.

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.

Saxe, R., Xiao, D.K., Kovacs, G., Perrett, D., Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, **42**(11), 1435–46.

Schilbach, L. (2015). Eye to eye, face to face and brain to brain: novel approaches to study the behavioral dynamics and neural mechanisms of social interactions. *Current Opinion in Behavioral Sciences*, **3**, 130–5.

Schilbach, L., Timmermans, B., Reddy, V., *et al.* (2012). Toward a second-person neuroscience. *Behavioral Brain Research*, **36**, 393–414.

Schneier, F.R., Rodebaugh, T.L., Blanco, C., Lewin, H., Liebowitz, M.R. (2011). Fear and avoidance of eye contact in social anxiety disorder. *Comprehensive Psychiatry*, **52**(1), 81–7.

Schultz, J., Brockhaus, M., Bülthoff, H.H., Pilz, K.S. (2012). What the human brain likes about facial motion. *Cerebral Cortex*, doi: 10.1093/cercor/bhs106.

Senju, A., Johnson, M.H. (2009). The eye contact effect: mechanisms and development. *Trends in Cognitive Sciences*, **13**(3), 127–34.

Skipper, J.I., van Wassenhove, V., Nusbaum, H.C., Small, S.L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, **17**(10), 2387–99.

Tso, I.F., Mui, M.L., Taylor, S.F., Deldin, P.J. (2012). Eye-contact perception in schizophrenia: relationship with symptoms and socioemotional functioning. *Journal of Abnormal Psychology*, **121**(3), 616.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., *et al.* (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, **15**(1), 273–89.

Vatikiotis-Bateson, E., Eigsti, I.M., Yano, S., Munhall, K.G. (1998). Eye movement of perceivers during audiovisualspeech perception. *Perception and Psychophysics*, **60**(6), 926–40.

Vertegaal, R., Slagter, R., Van der Veer, G., Nijholt, A. (2001). Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems ACM, Seattle, Washington, USA.

von dem Hagen, E.A., Stoyanova, R.S., Rowe, J.B., Baron-Cohen, S., Calder, A.J. (2014). Direct gaze elicits atypical activation of the theory-of-mind network in autism spectrum conditions. *Cereb Cortex* **24**(6), 1485–92.

Wilson, M., Wilson, T.P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review*, **12**(6), 957–68.

Yi, A., Wong, W., Eizenman, M. (2013). Gaze patterns and audiovisual speech enhancement. *Journal of Speech, Language, and Hearing Research*, **56**(2), 471–80.