Brain Structure and Function (2018) 223:3073–3089
https://doi.org/10.1007/s00429-018-1674-5

ORIGINAL ARTICLE

Spatial–temporal dynamics of gesture–speech integration: a simultaneous EEG-fMRI study

Yifei He[1] · Miriam Steines[1] · Jens Sommer[2] · Helge Gebhardt[4] · Arne Nagels[3] · Gebhard Sammer[4] · Tilo Kircher[2] · Benjamin Straube[1]

Received: 20 October 2017 / Accepted: 27 April 2018 / Published online: 8 May 2018

Abstract

The semantic integration between gesture and speech (GSI) is mediated by the left posterior temporal sulcus/middle temporal gyrus (pSTS/MTG) and the left inferior frontal gyrus (IFG). Evidence from electroencephalography (EEG) suggests that oscillations in the alpha and beta bands may support processes at different stages of GSI. In the present study, we investigated the relationship between electrophysiological oscillations and blood-oxygen-level-dependent (BOLD) activity during GSI. In a simultaneous EEG-fMRI study, German participants ($n = 19$) were presented with videos of an actor either performing meaningful gestures in the context of a comprehensible German (GG) or incomprehensible Russian sentence (GR), or just speaking a German sentence (SG). EEG results revealed reduced alpha and beta power for the GG vs. SG conditions, while fMRI analyses showed BOLD increase in the left pSTS/MTG for GG > GR ∩ GG > SG. In time-window-based EEG-informed fMRI analyses, we further found a positive correlation between single-trial alpha power and BOLD signal in the left pSTS/MTG, the left IFG, and several sub-cortical regions. Moreover, the alpha-pSTS/MTG correlation was observed in an earlier time window in comparison to the alpha-IFG correlation, thus supporting a two-stage processing model of GSI. Our study shows that EEG-informed fMRI implies multiple roles of alpha oscillations during GSI, and that the method is a best candidate for multidimensional investigations on complex cognitive functions such as GSI.

Keywords Simultaneous EEG-fMRI · Gesture · Language · Alpha oscillations · Multisensory integration

Introduction

The ability to integrate visual and auditory information from gesture and speech during face-to-face communication is an important cognitive function (Goldin-Meadow 2005; McNeill 2006, 2008). In comparison to the integration of audio–visual speech (Beauchamp et al. 2004b; Calvert et al. 2004; van Wassenhove et al. 2005), gesture–speech integration (GSI) is a more complex cognitive function that encompasses not only lower order auditory, visual, and action perception, but is also inherently semantic in nature (Andric and Small 2012; Kita and Özyürek 2003; McNeill 2008; Özyürek et al. 2007; Willems et al. 2009). For example, to understand a co-speech gesture formed by a thumb-up gesture and a sentence saying 'this is really good', it is necessary to understand the semantic representations from both input modalities, and then construct a coherent semantic representation in a combinatorial manner.

Recent evidence from functional magnetic resonance imaging (fMRI) research suggests that GSI primarily

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00429-018-1674-5) contains supplementary material, which is available to authorized users.

✉ Yifei He
yifei.he@staff.uni-marburg.de

1 Translational Neuroimaging Lab, Department of Psychiatry and Psychotherapy, Marburg Center for Mind, Brain and Behavior (MCMBB), Philipps-University Marburg, Rudolf-Bultmann-Str. 8, 35039 Marburg, Germany

2 Department of Psychiatry and Psychotherapy, Marburg Center for Mind, Brain and Behavior (MCMBB), Philipps-University Marburg, Rudolf-Bultmann-Str. 8, 35039, Marburg, Germany

3 Department of General Linguistics, Johannes-Gutenberg University Mainz, Jakob-Welder-Weg 18, 55128, Mainz, Germany

4 Cognitive Neuroscience at Centre for Psychiatry, Justus Liebig University Giessen, Am Steg 28, 35385 Giessen, Germany

recruits two core regions within the cerebral neocortex: the posterior superior temporal sulcus/middle temporal gyrus (pSTS/MTG) and the left inferior frontal gyrus (IFG) (Dick et al. 2009, 2012b; Green et al. 2009; He et al. 2015; Holle et al. 2010; Hubbard et al. 2009; Kircher et al. 2009; Skipper et al. 2009; Straube et al. 2011; Willems et al. 2007, 2009). The causal relationship between the two regions and semantic GSI were also supported by recent evidence from transcranial magnetic stimulation (TMS); that is, disrupting activities in both regions were shown to impair the semantic GSI (Zhao et al. 2018). It is further proposed that the two regions may be functionally dissociable during GSI within the framework of a two-stage processing model: while the pSTS/MTG is supposedly related to the multisensory association of visual–auditory modalities in a first stage (multimodal matching stage), the left IFG is primarily responsible for higher order semantic processes, e.g., semantic evaluation and reconstruction in a subsequent stage (semantic integration stage, cf. Dick et al. 2012b; He et al. 2015; Willems et al. 2009). This functional interpretation of the two regions is able to account for a wide range of empirical data. However, to date, due to the coarse temporal resolution of fMRI, no studies have directly tested whether the two substages within the model (1) can be mapped onto distinct brain regions, and (2) can be reflected in an earlier activation of the left pSTS/MTG during GSI compared to the left IFG.

In addition to functional imaging studies, recent EEG (electroencephalography) studies looking at the oscillatory aspects of GSI offer an alternative perspective. Neural oscillations reflect a fundamental mechanism for enabling coordinated activity that underlies cognitive functions (Buzsáki and Draguhn 2004; Uhlhaas and Singer 2010). To date, there are only a few studies that directly investigate the integration between gesture and speech, and these studies suggest that modulations in low-frequency oscillations may at least reflect processes that are relevant to GSI (Biau and Soto-Faraco 2015; Biau et al. 2015; He et al. 2015). For example, Biau and colleagues (2015) investigated the GSI between beat gesture and natural speech. They found phase synchronization in the theta band (5–6 Hz) as well as a desynchronization in the alpha band (8–10 Hz) for an audio–visual bimodal condition in contrast to an auditory-only unimodal condition. In this study, even if the GSI between beat gesture and auditory speech is not semantic in nature, the findings suggest that gesture is able to increase sensitivity to auditory speech in a way similar to that of visual speech (Luo et al. 2010; Schroeder et al. 2008), and this lower order integration process can be reflected by phase-resetting in the theta and alpha bands. To investigate semantic-level GSI, in a previous study from our laboratory (He et al. 2015), we looked at the GSI of intrinsically meaningful gestures (emblems and pantomimes) using separate EEG and fMRI recordings. We compared a bimodal co-speech condition (GG) directly with gestures with foreign Russian speech (GR) and videos containing only comprehensible German speech (SG). In the fMRI experiment, similar to the line of fMRI studies on GSI as well as audio–visual speech integration, we observed enhanced BOLD activity in the left pMTG for the bimodal GG condition ($GG > GR \cap GG > SG$). In the EEG experiment, we found that both alpha (7–13 Hz) and beta power (16–20 Hz) decreased for GG–SG, and that alpha power decreased for GG-GR. Therefore, our finding may suggest a potential role of alpha oscillations during higher order GSI, because they are sensitive to both bimodal-auditory and bimodal-visual comparisons. Of note, although alpha oscillations have been suggested to underlie both visual and auditory attention that are, to some degree, relevant to GSI (Foxe and Snyder 2011; Obleser and Weisz 2012; Thut et al. 2006), it remains unclear why semantic-level GSI is reflected by power modulations in the alpha band. Interestingly, a study from Willems et al. (2008) also reported alpha power decrease (8–12 Hz) for semantic integration between picture and sentence (Willems et al. 2008). This study, together with our previous study, may suggest that alpha oscillations may at least support higher order semantic processes. Therefore, it would be necessary to further investigate the exact functional role of alpha oscillations overarched under GSI. Another confound from He et al. (2015) is the observation of the beta band effect: as we observed both alpha and beta power decrease for the bimodal vs. auditory comparison, it remains elusive whether the effects reflect semantic GSI, or merely observation of an additional visual input, as both alpha and beta band power are sensitive to action observation (Järveläinen et al. 2004; Perry et al. 2011). Clearly, the exact functional interpretation of both alpha and beta oscillations during GSI needs to be further specified.

Nevertheless, the initial evidence from EEG and fMRI investigations on GSI suggests a potential relevance between alpha (and potentially beta) oscillations and brain activity in a left fronto-temporal network including both the left pSTS/MTG and the left IFG. However, to date, there exists no literature that directly explores the relationship between brain oscillations and BOLD signal during the integration of gestures and corresponding speech. Therefore, simultaneous EEG-fMRI can be applied to resolve this issue. Simultaneous EEG-fMRI is a powerful tool to reveal the relation between EEG effects and changes of brain activity as revealed by BOLD signal (Debener et al. 2006). The method is especially informative, because it adds an additional temporal dimension to the relatively static fMRI (Eichele et al. 2005), so that both temporal and spatial precision can be achieved. Recent research with simultaneous EEG-fMRI mostly suggests that alpha band, as well as beta band power often negatively correlates with BOLD activity not only in the resting state (Goldman et al. 2002; Laufs et al. 2003a, 2006), but also during cognitively demanding memory and motor tasks (Arnstein et al. 2011;

Scheeringa et al. 2009). However, positive correlation between alpha power and BOLD was also reported (Sadaghiani et al. 2010). This difference in the direction may suggest diverse roles of alpha oscillations that varies depending on tasks and functions (Sadaghiani and Kleinschmidt 2016). Moreover, although event-related changes in alpha and beta bands very often co-occur, they seem to reflect independent processes (Scheeringa et al. 2011). As the previous EEG studies suggest the potential relevance of both alpha and beta oscillations during GSI, it is thus necessary to evaluate whether they couple with BOLD in an independent and thus dissociable manner.

The purpose of the present study was twofold: first, we intended to identify brain regions showing BOLD activity change that correlates to GSI-related oscillatory activity in the alpha and/or beta bands. Second, we sought to explore the temporal dynamics of the brain regions that are potentially involved in GSI. We thus recorded simultaneous EEG-fMRI while subjects watched video clips of an actor performing either intrinsically meaningful gestures in a familiar (gesture-German, GG) or unfamiliar (gesture-Russian, GR) language context, or familiar speech on its own (speech-German, SG). Similar to our previous study—where we employed the identical experimental paradigm, but recorded EEG and fMRI in separate sessions/subjects (He et al. 2015)—we compared the bimodal GG condition with the two 'unimodal' control conditions (GR and SG), respectively. We then conducted a classical EEG-informed fMRI analysis, by correlating single-trial alpha/beta power (averaged across the time windows showing significant difference between conditions) with the BOLD. Taking a step further, we conducted a novel time-window analysis, correlating power parameters with BOLD within five arbitrarily defined time windows of 200 ms (see "Methods"). Based on the previous findings from fMRI and EEG studies on GSI (Dick et al. 2012b; He et al. 2015; Straube et al. 2011; Willems et al. 2009), we expected a replication of results from He et al. (2015) in both the conventional EEG and fMRI analyses. With regard to the coupling between EEG and fMRI, we first hypothesized a negative link between alpha power and BOLD signal in a left fronto-temporal network including the left pSTS/MTG and potentially the left IFG. Following the assumptions of the two-stage processing model, we further hypothesized that a correlation of pSTS/MTG activity and alpha power may be found in earlier time windows compared with that of the left IFG.

Methods

Participants

Twenty participants (12 females, mean age 22.65 years, range 19–32 years) participated in the study. All participants were native German speakers and had no knowledge of the Russian language. All participants were right-handed (Oldfield 1971) and had normal or corrected-to-normal vision. None of the participants reported any hearing deficits. Exclusion criteria were histories of relevant medical or psychiatric illnesses of the participants. All subjects gave written informed consent prior to participation in the experiment. The study was approved by the local ethics committee. Data from one subject had to be excluded because of incomplete data collection due to a technical malfunction.

Materials

Subjects were presented with video clips of speech together with intrinsically meaningful gestures (IMG) or speech on its own, all selected from a large pool of videos. Identical stimuli were used in the previous studies with separate EEG and fMRI recordings (He et al. 2015). The IMG videos consisted of an equal number (52 each) of emblematic and tool-related gestures (pantomimes). All videos lasted 5 s and showed the same actor speaking SVO sentences. The actor was a highly proficient bilingual speaker of both German and Russian. In these videos, he performed either: (1) an IMG in the context of a corresponding German sentence (GG), (2) an IMG in the context of a Russian sentence (GR), or (3) an isolated German sentence with social (emblematic) or tool-related content, respectively (SG). Crucially, while GG represents the bimodal condition which allows semantic integration, both GR and SG condition contain meaningful input from only one modality (GR for gesture input and SG for speech input). Notably, the GR condition provides comparable auditory input without semantic content (incomprehensible Russian speech). Similar approaches to obtain an integration effect have proven fruitful in the previous studies on iconic/metaphoric gestures from our research group (Green et al. 2009; Kircher et al. 2009; Straube et al. 2011). Overall, we used 312 videos (52 videos per condition $\times$ 3 conditions $\times$ 2 sets) and another 26 filler videos containing Russian sentences with meaningless gestures. For an illustration of the videos, please refer to Fig. 1. Importantly, as suggested by Fig. 1, for videos describing the same gesture–speech event, identical gestures and German sentences were performed in the three conditions, respectively. Moreover, due to counterbalancing, each participant saw each gesture or heard each sentence only once, to avoid repetition effects. For a complete list of experimental items including a detailed description of the videos (picture illustration, speech and gesture durations, understandability and naturalness ratings, and full list of materials), please refer to He et al. (2015).

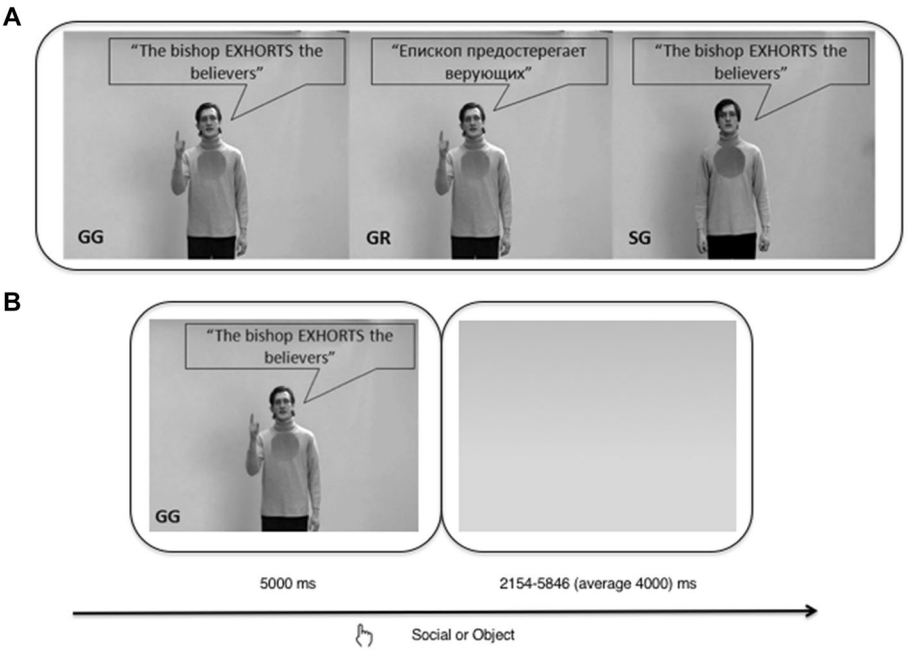

Fig. 1 **a** Picture illustration for each of the three experimental conditions (GG, gesture-German; GR, gesture-Russian; SG, speech-German). For illustrative purposes the spoken German sentences were translated into English and all spoken sentences were written into speech bubbles. **b** Illustration of a sample trial. Participants performed a content judgment task for each video, indicating via button press whether a stimulus was either social- or object-related

Experimental procedure

During EEG-fMRI data acquisition (see below), participants were presented with videos of an actor performing IMG with accompanying German (GG) or Russian (GR) speech, or only speaking meaningful German sentences (SG). An experimental session comprised 182 videos (52 for each condition and 26 additional filler videos) and consisted of two 14-min blocks. Each block contained 91 trials with a matched number of items from each condition (26) and 13 filler trials, which displayed meaningless gestures with Russian speech. The stimuli were presented in an event-related design in pseudo-randomized order and were counterbalanced across participants. Each video clip was followed by a gray background with a variable duration of 2154–5846 ms (jitter average: 4000 ms). Participants performed a content judgment task for each video, indicating via button press whether a stimulus was either person- or object-related: all emblematic gestures/social sentences were person-related, whereas tool-related gestures/sentences were object-related. By performing this task, the participants focused directly on the semantic representation of the IMG and the corresponding speech. Participants were instructed to respond to the task as soon as they had decided on an answer. Of note, this task is independent of experimental conditions, namely, participants were asked to focus on the performance of the actor, and not specifically on either gesture or speech.

EEG acquisition and preprocessing

The EEG was recorded from 31 channels attached to an MR-compatible BrainCap (Brain Products) according to the international 10–20 system. After participants were mounted with the EEG cap, they lay in the scanner tube in a supine position. Their head movement was restricted by a head clamp attached to the head coil. An MRI compatible amplifier together with the 'Syncbox' device was used (Brain Products). The 'Syncbox' synchronizes the clock of the EEG amplifier to the internal clock of the scanner, which is important for the off-line removal of the MR gradient artifact. The reference electrode was located at FCz and the ground electrode was located at the forehead in front of Fz. One additional electrode was placed on the participant's upper back to record and later reject the electrocardiogram (ECG) artifact during EEG preprocessing. Impedances for all electrodes were kept below 5 kΩ and EEG data were sampled at 5 kHz. EEG preprocessing was performed in BrainVision Analyzer 2.0 (Brain Products) and the Fieldtrip toolbox for EEG/MEG analysis (Oostenveld et al. 2011). We first assured that there was no jitter between volume start markers given by the scanner (invariant 2000 ms). Then, gradient and BCG (ballistocardiogram) artifacts were corrected according to the average artifact subtraction method proposed by Allen et al. (1998). Specifically, the gradient artifacts were corrected by constructing an average artifact template over 41 consecutive volumes in a sliding-window approach, and then subtracting this template from the raw EEG data for each volume. The BCG artifacts were removed using a similar approach in which ECG R-waves were averaged to produce a BCG artifact template. The resulting artifact template was then subtracted from the EEG data to correct for BCG contamination. The MR-corrected continuous EEG data were down-sampled to 250 Hz, band-pass filtered between 0.1 and 125 Hz, and re-referenced to the average

of the two mastoids (TP9 and TP10). EOG artifacts were identified and rejected via an infomax independent component analysis (ICA) before being exported to the Fieldtrip toolbox (Maris and Oostenveld 2007) for further analysis. For each subject, a maximum of three ICs were identified and rejected.

The raw EEG was segmented into −0.5–1.5 s segments around the onset of each critical word (the word during which the meaning of gesture and speech coincides, such as *good* in the sentence "the actor did a good job"). The average onset of critical words for all video clips was 1.16 s after the beginning of the video (standard deviation 0.54 s). After the segmentation, additional jump artifacts were automatically detected and rejected based on the amplitude distribution across trials and channels (as implemented in Fieldtrip toolbox). Cutoffs for the jump artifacts were set at $z = 20$. The average number of trials remaining was 44.1 (out of 52), roughly equal between conditions (paired sample t test, $t_{max} = 0.32$, $p_{min} = 0.74$).

fMRI acquisition and preprocessing

All MRI data were acquired on a 3T scanner (Siemens MRT Trio series). Functional images were acquired using a T2-weighted echo planar image sequence (TR = 2 s, TE = 30 ms, flip angle 90 degrees, slice thickness 4 mm with a 0.36 mm interslice gap, 64 × 64 matrix, FoV 230 mm, in-plane resolution 3.59 × 3.59 mm, 30 axial slices orientated parallel to the AC-PC line covering the whole brain). Two runs of 425 volumes were acquired during the experiment. The experiment was synchronized to the scanner pulse, and the onset of each video for each condition was shifted between − 1846 and 1846 ms, leading to a variable jitter of 2154–5846 ms (see Fig. 1). MR images were analyzed using Statistical Parametric Mapping (SPM8) standard routines and templates (http://www.fil.ion.ucl.ac.uk). After discarding the first five volumes to minimize T1-saturation effects, all images were spatially and temporally realigned, normalized into the MNI space using the MNI template (resulting voxel size 2 × 2 × 2 mm^3), smoothed (8 mm isotropic Gaussian filter), and high-pass filtered (cut-off period 128 s).

EEG data analysis

Like in the previous studies (He et al. 2015), time–frequency representations (TFRs) of single-trial data were computed in the frequency range between 2 and 30 Hz, to reveal event-related oscillatory dynamics of the EEG. For time–frequency decomposition of the EEG time-series, a constant Hanning taper of 400 ms was used to compute power changes in frequency steps of 1 Hz and time steps of 0.05 s. For statistical analyses, a mass cluster permutation test was carried out on the baseline-corrected (−0.5 to −0.15 s) relative power changes. The procedure of the statistical analysis is briefly described in the following (for detailed description of the statistical procedure, see Maris and Oostenveld 2007): first, for every data point (time–frequency-channel point), a simple-dependent samples t test is performed and results in uncorrected p values. Second, all significant data points ($p < 0.05$) are grouped as clusters (here: clusters of at least three neighboring electrodes with maximum distance of 2.5 cm). For each cluster, the sum of the t statistics is used in the cluster-level statistics. Finally, a Monte Carlo non-parametrical permutation method with 1000 repetitions is implemented to estimate type I error controlled cluster significance probabilities ($p < 0.05$) in the time–frequency-channel space.

Conventional fMRI data analysis

Statistical whole-brain analysis was performed in a two-level, mixed-effects procedure. On the first level, single-subject BOLD responses were modeled by a design matrix comprising the time points of each event (critical word of each sentence as used in the previous event-related fMRI studies, e.g. Green et al. 2009; He et al. 2015; Kircher et al. 2009) with duration of 1 s for all experimental conditions. The micro-time onset was set to the average time bin (8 of 16) to align the onset vector to the slice in the middle of the brain. This should lead to the highest temporal accuracy of the onsets for fronto-temporal brain regions. As an additional factor, each video was modeled as a mini-block with 5 s duration, to account for general video processing across conditions (Straube et al. 2013). For all conditions, the duration of speech and/or gesture was used as a parameter of no interest on a single-trial level. The six movement regressors (three rotations and three translations) were entered in the single subject's model to account for movement induced effects on fMRI results. The hemodynamic response was modeled by the canonical hemodynamic response function (HRF). Parameter estimate (β-) images for the HRF were calculated for each condition and each subject. Parameter estimates for the three relevant conditions were entered into a within-subject flexible factorial ANOVA.

To identify activity specific to GSI, we first created single contrasts by subtracting both the gesture-Russian and the speech-German condition from the gesture-German condition (GG > GR and GG > SG), as what we did in the previous studies (He et al. 2015). Then, we calculated the conjunction (Nichols et al. 2005) of the bimodal condition in contrast to both 'unimodal' conditions (GG > GR ∩ GG > SG). The rationale behind this procedure is that GG represents a bimodal condition which allows semantic integration of speech and gesture, whereas both GR and SG conditions contain meaningful input from only one input modality. In addition, the GR condition provides comparable auditory

input without semantic content (incomprehensible Russian speech). Similar approaches to obtain an integration effect have proven fruitful in previous studies on iconic/metaphoric gesture from our lab (Green et al. 2009; He et al. 2015; Kircher et al. 2009; Straube et al. 2011).

A Monte Carlo simulation of the brain volume was employed to establish an appropriate voxel contiguity threshold (Slotnick and Schacter 2004). This correction has the advantage of higher sensitivity to smaller effect sizes, while still correcting for multiple comparisons across the whole-brain volume. Assuming an individual voxel type I error of $p < 0.001$, a cluster extent of 50 contiguous resampled voxels was indicated as necessary to correct for multiple voxel comparisons at $p < 0.05$. This cluster threshold (based on the whole-brain volume) has been applied to all difference contrasts and conjunctions of the conventional analyses.

EEG-informed fMRI analysis

In this analysis, according to the classic routine of EEG-informed fMRI analysis (Scheeringa et al. 2009), we first examined power–BOLD coupling in the alpha/beta bands within the shared time–frequency-electrode bins that showed significant contrasts between GG/GR and SG comparison based on the EEG results. Therefore, we averaged the power at 9–11 Hz between 0.2 and 0.6 s for the alpha band and at 17–20 Hz between 0.4 and 1.0 s for the beta band, post-onset of the critical word, across the electrodes that showed significant contrasts (see Fig. 2b), and correlated them with the BOLD signal. In a second step, we run a time-window analysis, by calculating the correlations between BOLD signal and alpha/beta power in five arbitrarily defined continuous time windows of 200 ms, post-onset of the critical word. With this analysis, we were able to explore how power–BOLD coupling evolves in a temporal manner. For EEG-informed fMRI analyses (both classical and time-window-based), in addition to the condition-related regressors in the classical fMRI analysis, correlation effects between power values in both the alpha/beta bands and BOLD were modeled in the GLM using parametric modulations based on single-trial power. The power values were *z*-transformed to minimize between-subject variance, and missing values resulted from artifact rejection were replaced by zeros. Additional regressors of no interest (identical to the conventional fMRI analysis) were also included in the model. In the classical EEG-informed fMRI analysis, alpha and beta power were modeled separately at each time point. In the time-window analysis, for each of the five time windows, two separate models were constructed for alpha and beta band power at each time point.

Different from the conventional analysis where difference contrasts can be straightforwardly interpreted, in the EEG-fMRI analyses, we first run an *F* test to evaluate whether parametrically modulated effects in each condition were significantly different from zero, despite the direction of the correlation. Second, we masked these regions with regions that showed either positive/negative correlations between BOLD and alpha/beta power. The overlap resulted from these two steps can be interpreted as positive (+) or negative (−) correlations for each experimental condition. Based on this set of results, as no negative correlations between alpha/beta power and BOLD were observed, in a third step, similar to the conventional fMRI analysis, we calculated the GG > GR ∩ GG > SG conjunction, to reveal the oscillations-associated BOLD signal change that is specific for GG at each time point (GG > GR ∩ GG > SG ∩ GG +). Importantly, while the conjunction GG > GR ∩ GG > SG alone is statistically confounding, we only focused on regions which are not only significantly modulated in the GG condition, but also specific for the comparison between GG and both the GR and SG conditions.

Typically, EEG-informed fMRI studies are relatively weak when compared to the conventional fMRI analysis, because the analysis measures the residual effects after removing the mean evoked responses. Therefore, uncorrected voxel-wise threshold of $p < 0.001$ or $p < 0.005$ is able to yield highly interpretable statistical brain maps (Debener et al. 2005; Eichele et al. 2005; Laufs et al. 2003a). Here, for the analysis of single conditions (as in steps 1 and 2), we used uncorrected threshold of $p < 0.01$ for the *F* test, and the mask with positive/negative directions for baseline *t* contrasts were created based on a primary threshold of $p < 0.001$ and a cluster size of 50 voxels to correct for multiple comparisons. For the conjunction analyses, we used Monte Carlo simulations with an uncorrected primary threshold ($p < 0.05$) in combination with a larger cluster size (211 contiguous resampled voxels) to correct for multiple comparisons at $p < 0.05$. Comparable cluster-extent correction methods have been applied in the previous studies examining the GSI effects, and have yielded highly interpretable statistical brain maps (Green et al. 2009; Nagels et al. 2015; Straube et al. 2011, 2014).

The reported voxel coordinates of activation peaks are located in MNI space. For the anatomical localization, functional data were referenced to the AAL toolbox (Tzourio-Mazoyer et al. 2002).

Correlation analysis

To explore the relationship between alpha and beta oscillations and specific stimulus parameters as well as task performance that are likely to be related to the difficulty of semantic integration in each experimental condition, we performed several correlation analyses (Pearson's *r*): correlations were computed between the alpha/beta relative power

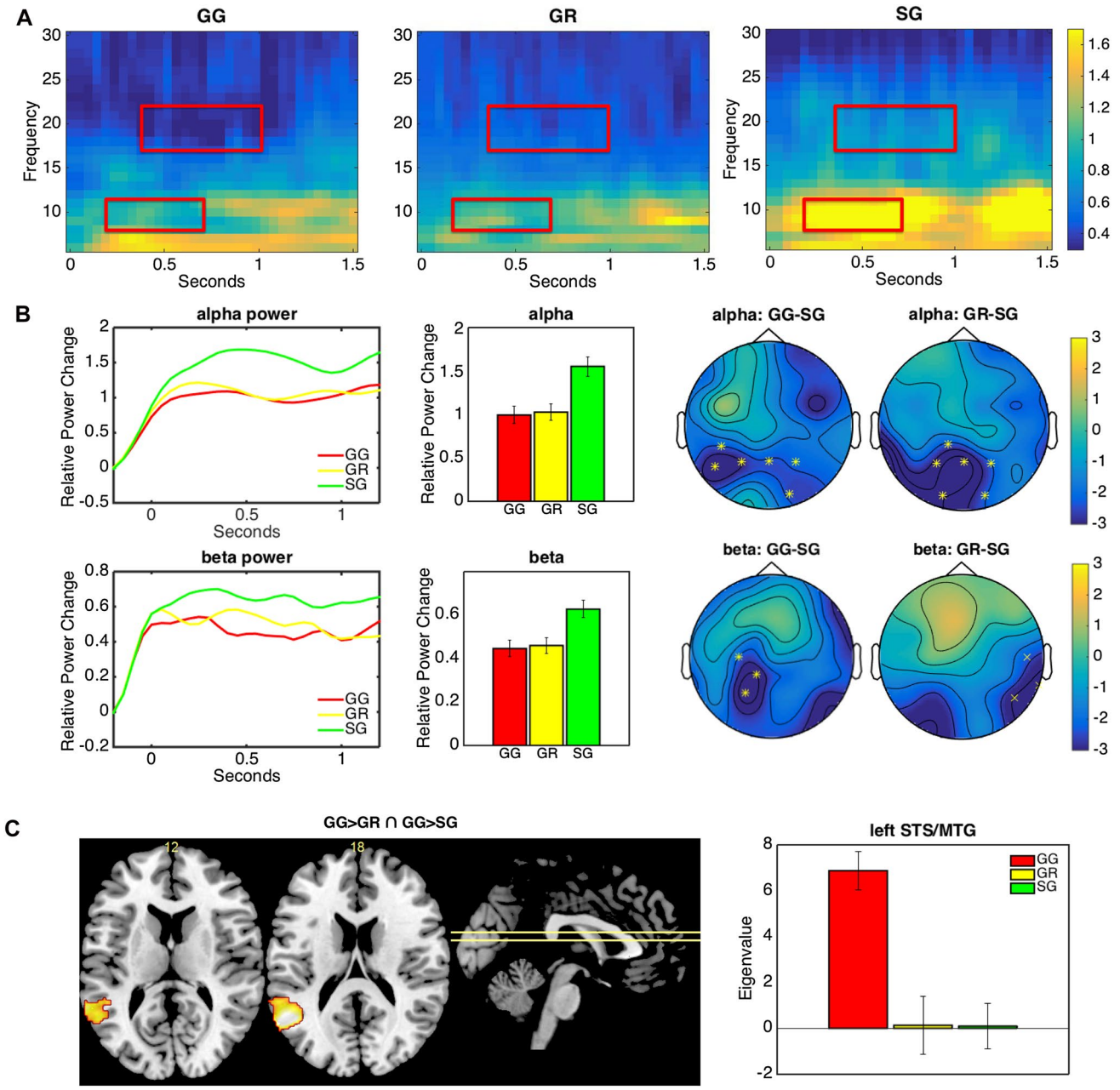

Fig. 2 Results of the EEG and fMRI analysis. **a** Time–frequency representations (TFR) at electrode P3 for the GG (left), GR (middle), and the SG (right) conditions. **b** Temporal evolution of the alpha and beta power between −0.2 and 1.2 s after the critical word onset, for the three experimental conditions (left), the averaged power for both alpha (0.2–0.7 s) and beta (0.4–1.0 s) bands in four parietal electrodes (Pz, Cz, P3, and P7) for all three conditions (middle), and the scalp distribution of the alpha (9–11 Hz) and beta band (17–20 Hz) effects in the GG vs. SG and GR vs. SG comparisons. The electrodes which form the significant cluster between the GG vs. SG comparison ($p < 0.05$, Monte Carlo corrected) are marked with asterisks on the topographic plot. **c** Brain activation for the GSI effect as reflected by conjunction analysis (GG > GR ∩ GG > SG) in the conventional fMRI analysis. The bar graph indicates the eigenvalue of the left pSTS/MTG (810 voxels), standard errors are indicated with the error bar

changes in different time windows and (a) understandability ratings, (b) naturalness ratings, and (c) concreteness ratings (acquired in a different set of subjects for stimulus selection; see He et al. 2015) as well as (d) subjects' reaction times and (e) accuracy rates. To minimize between-subject variance, we *z*-transformed power values within each subject for each frequency band and time window, and then averaged the *z*-transformed power values across subjects for each experimental item (trial). With regard to the behavioral measures, both the reaction times and accuracy rates were

averaged across subjects for each trial. Then, the by-item (trial) correlation analysis was performed separately for all three experimental conditions.

Results

Behavioral results

For each video, participants were instructed to indicate via button press whether the actor's gesture and/or speech referred to a person- or object-related event. Means and standard deviations of correct responses and their reaction times are listed in Table 1. No difference in reaction times was observed. However, the difference of accuracy rates between conditions was significant by within-subjects ANOVA ($F_{(2,38)}=4.71$, $p<0.03$). Post hoc pair-wise t test (Bonferroni-corrected) suggests that only the difference between the GG and SG condition for accuracy rate is significant (t test, $t_{(19)}=4.35$, $p<0.0005$).

EEG results

We directly compared the meaningful bimodal condition (GG) with the two conditions that contained only one comprehensible modality (GR and SG). When comparing the GG and SG conditions, we observed a negative cluster ($p=0.0001$, as illustrated in Fig. 2a) as reflected by power changes in two frequency bands: in the alpha band (9–11 Hz), the power decrease of the GG vs. SG condition occurred between 0.2 and 0.7 s after critical word onset, and has a parietal scalp distribution; in the beta band (17–20 Hz), GG showed a power decrease when compared to SG from 0.3 to 1.2 s and, similar to the alpha band, the effect has a parietal scalp distribution. For an illustration of the alpha and beta time-series and scalp distribution, see Fig. 2b. For the comparison between the GG and GR conditions, no significant cluster was observed ($p_{min}=0.11$). Similar to the separate EEG results in our previous study (He et al. 2015), the comparison for GG–SG yielded significant effects in both the alpha and the beta bands. However, we did not observe reliable effects during the GG–GR comparison. In addition, we compared GR vs. SG conditions. This comparison revealed a significant power decrease for the GR vs. SG comparison ($p=0.0001$) in both the alpha band (0.2–0.6 s, parietal scalp distribution) and the beta bands (0.4–1.0 s, right-parietal scalp distribution).

Table 1 Results for the behavioral task

Condition	Correct response (%)		Reaction times (ms)	
	Mean	SD	Mean	SD
Gesture-German (GG)	89.71	5.38	5362.44	981.40
Gesture-Russian (GR)	86.25	7.60	5387.44	961.88
Speech-German (SG)	85.48	6.73	5377.03	960.87

Reaction times were measured in reference to each video onset (full-video length 5000 ms). *SD* standard deviation

Conventional fMRI results

The bimodal integration (GG > GR ∩ GG > SG) effect in the conventional fMRI analysis is shown in Fig. 2c and Table 2. The whole-brain analysis revealed that this effect emerged in the left pSTS and the left pMTG. The finding is direct replication to the previous fMRI studies concerning the integration of IMG (He et al. 2015); they are also in line with studies on GSI of other gesture types (Dick et al. 2012b; Holle et al. 2010; Hubbard et al. 2009; Straube et al. 2011).

Classical EEG-informed fMRI results

We first analyzed the BOLD signal informed by both alpha and beta power within the latency range (200–600 ms for the alpha band, 400–1000 ms for the beta band) that showed a significant difference between GG/GR and SG. For the beta band, no positive/negative correlations were observed for either single condition baseline contrasts or the conjunction analysis. For the alpha band, however, we observed only positive correlations with BOLD (as in Fig. 3a; Tables 3, 4). In the GG condition, alpha power correlates positively with BOLD in the anterior cingulate gyrus and the middle occipital gyrus; the conjunction analysis revealed additionally the left rolandic operculum, the left IFG, the left insula, the right postcentral gyrus, as well as the cerebellum in both hemispheres. Taken together, GG condition showed condition-specific positive correlation between alpha power and BOLD in the anterior cingulate gyrus and the middle occipital gyrus.

Time-window-based Power–BOLD coupling

Besides the EEG-informed fMRI analysis, we additionally informed fMRI with EEG power in both the alpha (9–11 Hz) and beta bands (17–20 Hz) averaged within five arbitrarily defined time windows. In the time-window analysis, for the alpha band, we only observed significant positive correlations

Table 2 Brain activations for GSI in the conventional fMRI analysis

Contrast	Peak	HF	size	MNI coordinates (*x*, *y*, *z*)			*t*
GG > GR ∩ GG > SG	Middle temporal gyrus	L	810	−52	−58	18	4.75

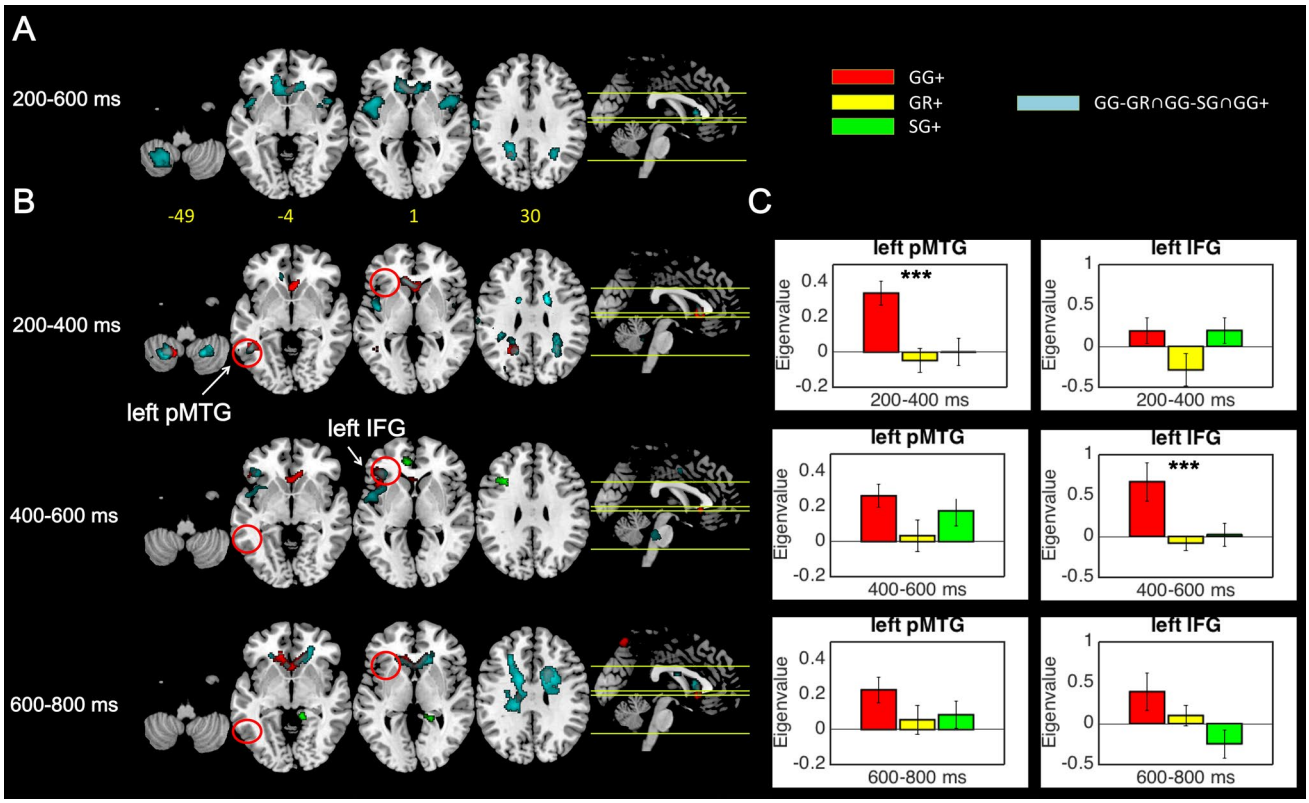

Fig. 3 Alpha power and BOLD coupling. **a** Single-trial correlations between the alpha band power (200–600 ms) and the BOLD signal across conditions. **b** Single-trial correlations between the alpha band power and the BOLD signal in different time windows. **c** Mean eigenvalues in the left pMTG and the left IFG in different time windows for all three experimental conditions in the alpha-BOLD correlation analysis. Asterisks indicate the time window(s) in which the left pMTG/left IFG activation is significantly correlated with alpha power

with BOLD (Fig. 3; Tables 3, 4). In the 0-200 ms time window, no significant correlations were observed. In the 200–400 ms time window, only GG condition showed positive alpha-BOLD correlation. The clusters were located in the anterior cingulate gyrus, the bilateral middle temporal gyrus, the left middle occipital gyrus, and the cerebellum. The conjunction analysis revealed similar regions. In the 400–600 ms time window, GG condition revealed positive alpha-BOLD correlation in the anterior cingulate gyrus and the left IFG. SG also showed positive alpha-BOLD correlation in the left middle frontal gyrus, the anterior cingulate gyrus, as well as the left superior frontal gyrus. The conjunction analysis also activated these regions. In the 600–800 ms time window, GG showed positive alpha-BOLD correlation in the cingulate gyrus, the right postcentral gyrus, and the left precuneus. SG showed positive alpha-BOLD correlation in the right lingual gyrus. The conjunction analysis in this time window revealed similar regions to the GG condition, despite more restricted cluster size. In the 800–1000 ms time window, no significant effects were observed for each single condition.

With regard to the beta band, the single condition baseline contrasts and the conjunction analysis revealed no overlapping regions: this data pattern may suggest that the no beta power–BOLD correlation is related to the bimodal integration in the GG condition. We thus reported the beta-BOLD correlation results in the supplement.

Correlation analysis results

The by-item (trial) correlation analyses between alpha/beta power and behavioral task performance as well as stimuli rating values revealed no significant correlations with beta power (in all time windows). In addition, no significant correlations between relative alpha power change and behavioral task performance were observed. However, we found significant positive correlation between naturalness ratings and alpha power in the 200–400 ms time window ($r = 0.20$, $p = 0.04$), as illustrated in Fig. 4.

Discussion

The previous EEG literature shows that oscillations in the alpha and beta bands may reflect processes that are relevant to gesture–speech integration (Biau and Soto-Faraco 2015; Biau et al. 2015; He et al. 2015). In addition, a plethora

Table 3 Coupling between BOLD and alpha power (single conditions)

Time window (ms)	Condition	Direction	Peak	HF	Size	MNI coordinates			*F*
						x	*y*	*z*	
200–600	GG	Positive	Anterior cingulate gyrus	R	154	2	24	−2	7.99
			Middle occipital gyrus	L	52	−24	−52	30	5.96
0–200									NS
200–400	GG	Positive	Anterior cingulate gyrus	R	221	2	22	−2	9.58
			Middle temporal gyrus	L	106	−46	−48	−6	8.51
			Middle occipital gyrus	L	213	−25	−54	32	8.48
			Middle temporal gyrus	R	58	34	−58	16	7.36
			Cerebellum VIII	L	151	−16	−56	48	7.21
			Middle occipital gyrus	L	78	−36	−74	16	6.77
			Precuneus	R	72	28	−52	28	5.24
400–600	GG	Positive	Anterior cingulate gyrus	R	86	10	30	−4	6.92
			Inferior frontal gyrus pars triangularis	L	155	−40	30	0	6.46
	SG	Positive	Middle frontal gyrus	L	92	−32	20	30	6.98
			Anterior cingulate gyrus	L	67	−10	40	2	6.55
			Superior frontal gyrus	L	146	−26	32	46	6.52
600–800	GG	Positive	Anterior cingulate gyrus	L	462	−12	32	4	8.22
			Middle cingulate gyrus	R	183	16	12	30	7.59
			Middle cingulate gyrus	L	58	−18	−10	34	7.15
			Postcentral gyrus	R	93	28	−26	42	6.92
			Precuneus	L	101	−2	−68	60	6.64
	SG	Positive	Lingual gyrus	R	186	18	−40	−12	5.74
800–1000									NS

This table lists the respective anatomical regions, cluster size, MNI coordinates, and *F* values for significant clusters in each experimental condition

MNI Montreal Neurological Institute, *L* left, *R* right

of fMRI studies suggests that both the left pSTS/MTG and the left IFG are closely related to GSI (Andric and Small 2012; He et al. 2015; Holle et al. 2010; Hubbard et al. 2009; Straube et al. 2011; Willems et al. 2009). By recording simultaneous EEG-fMRI in the present study, we aimed at identifying potential anatomical correlates to GSI-related oscillatory activities. In addition, with a novel time-window analysis, we tried to explore the temporal dynamics of the brain regions involved in GSI. In line with the previous studies employing separate EEG and fMRI recordings (He et al. 2015), we observed an alpha and beta power decrease for the bimodal (GG) vs. auditory-only (SG) comparison in the EEG analysis, and increased activation in the left pSTS/MTG in the conventional fMRI analysis for the GSI effect. In the classical EEG-informed analyses, we observed no correlation between beta power and BOLD, but positive correlation between alpha power and BOLD in several regions in the left hemisphere; however, no correlation effects in the left pSTS/MTG and the left IFG were observed. In a step further, our time-window-based analyses showed differential neural correlates for changes in the alpha and beta band power: single-trial alpha power correlated positively with BOLD signal in both the left pSTS/MTG and the left IFG. However, for the beta band, no beta-BOLD correlations seem to be specifically related to the GG vs. GR/SG conditions. More importantly, our time-window analyses also revealed that the alpha-BOLD correlation in the left pSTS/MTG occurs in an earlier time window (200–400 ms) in comparison to that of the left IFG (400–600 ms), thus providing experimental evidence in support of the two-stage processing model of GSI (Dick et al. 2012b; He et al. 2015; Willems et al. 2009).

The role of the pSTS/MTG during GSI

In the multisensory integration literature, the pSTS/MTG has been established as the most important brain region for integrating inputs from multiple modalities, from low-order audio–visual integration (Beauchamp 2005; Beauchamp et al. 2004a; Callan et al. 2004; Calvert et al. 2000) to higher order (semantic) integration between gesture and speech (Dick et al. 2012b; Green et al. 2009; He et al. 2015;

Table 4 Coupling between BOLD and alpha power (GG > GR ∩ GG > SG ∩ GG +)

Time window (ms)	Peak	HF	Size	MNI coordinates			*t*
				x	*y*	*z*	
200–600	Caudate	L	724	−6	26	4	3.72
	Rolandic operculum	L	917	−46	−4	8	3.29
	Angular gyrus	L	316	−22	−44	30	2.97
	Postcentral gyrus	R	458	32	−30	40	2.86
	Cerebellum VIII	L	355	−20	−54	−50	2.69
0–200							NS
200–400	Middle cingulate gyrus	R	3974	18	8	34	3.76
	Middle temporal gyrus	L	263	−46	−48	−6	3.75
	Caudate	L	525	−6	26	4	3.59
	Cerebellum VIII	L	4258	−22	−56	−46	3.36
	Rolandic operculum	R	235	46	−2	6	2.4
400–600	Inferior frontal gyrus pars triangularis	L	778	−38	30	0	3.21
	Cerebellum III	R	325	18	−26	−34	2.79
	Superior frontal gyrus	L	372	−18	−6	48	2.48
600–800	Middle cingulate gyrus	R	2560	20	12	30	3.42
	Postcentral gyrus	R	297	28	−30	46	2.83
800–1000	Cerebellum VII	L	247	−20	−54	−44	2.41

This table lists the respective anatomical regions, cluster size, MNI coordinates, and *t* values for each significant activation ($p < 0.05$, Monte Carlo cluster threshold corrected)

MNI Montreal Neurological Institute, *L* left, *R* right

Holle et al. 2008). As such, the increased activation in the pSTS/MTG during GSI observed in the present study further strengthens the understanding of the posterior temporal gyrus as multimodal hub. Moreover, the results from the conventional fMRI analysis are a direct replication of our separate fMRI experiment employing the exact same experimental design (He et al. 2015). The results thus support the validity and stability of our experimental paradigm, generating the same effect despite different data collection methods (fMRI-only vs. concurrent EEG-fMRI). In addition, they further corroborate the specific functional role of the pSTS/MTG in the integration of gesture and speech.

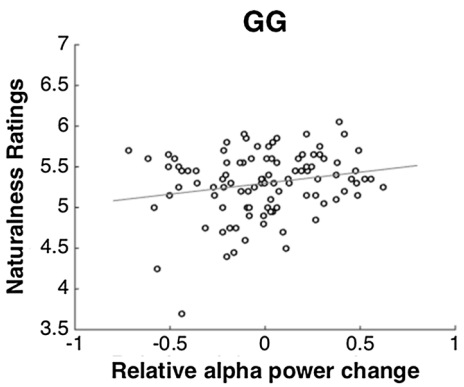

Fig. 4 Positive correlation between normalized power in the alpha band in the 200–400 ms time window and naturalness ratings (scale range 1–7) in the GG condition; each data point represents one trial/video

However, given the complexity of integrating gesture and speech, the role of pSTS/MTG may need to be further elucidated. To this end, the two-stage model of GSI provides a helpful theoretical framework: it has been suggested that, during GSI, the pSTS/MTG may be responsible for relatively low-order combinatorial processes, e.g., the matching of the input streams from multiple modalities; whereas more frontal brain regions (such as the left IFG) might be responsible for higher order processing, such as the evaluation/reconstruction of semantic representations (Dick et al. 2012b; He et al. 2015; Willems et al. 2009). This functional differentiation is able to account for a wide range of fMRI data, demonstrating that semantically more complex integration processes of gesture and speech usually recruit the left IFG [e.g., the integration of metaphoric gestures (Straube et al. 2011) or a semantic mismatch between gesture and speech (Willems et al. 2009)], and that less complex integration processes do not [e.g., integration of beat gestures (Hubbard et al. 2009), intrinsically meaningful gestures (He et al. 2015; Hubbard et al. 2009), or iconic gestures (Straube et al. 2011)]. The pSTS/MTG, on the other hand, seems to be involved in all types of GSI, as well as the integration of

audio–visual speech, irrespective of the semantic relationship between gesture and speech (for a detailed discussion, see He et al. 2015). Besides, fMRI studies on meaningful co-speech gestures focusing at the effective connectivity also showed that the pSTS/MTG mediates visual/auditory inputs from respective sensory regions (Dick et al. 2012a), thus suggesting that the region may not only integrate low-level audio–visual feature, but also match higher order semantic information from both gesture and speech. Moreover, while fMRI-only studies are unable to obtain any data with satisfactory temporal resolution in support of this hypothesis, our approach with simultaneous EEG-fMRI can provide the first direct evidence in this regard (see below).

The role of alpha and beta power during GSI

Partially consistent with the previous EEG experiments on GSI, we observed a power decrease in both the alpha and beta bands for the comparison between the bimodal GG and the auditory-only SG conditions (Biau et al. 2015; He et al. 2015). Oscillations in the alpha and beta bands have generally been established as a marker for action execution and observation (Järveläinen et al. 2004; Perry and Bentin 2009; Perry et al. 2011). More specifically, with regard to gesture, it has been suggested that the observation of gestures elicits a power decrease in both the alpha and beta bands (Quandt et al. 2012). Although, in the multisensory integration literature, the relevance of the alpha/beta bands is not as commonly discussed as the gamma band (for a review, see Senkowski et al. 2008), recent investigations on GSI have shown that oscillations in the alpha/beta bands may be relevant to multisensory integration processes (Leske et al. 2014; Romero et al. 2016), and more specifically, the integration of gesture and speech (Biau et al. 2015; He et al. 2015). However, it has to be noted that, as recent studies on GSI mostly obtain the integration effect by subtracting a unimodal speech condition from a bimodal gesture–speech condition, the modulation of oscillatory activity may reflect a mere modality-specific contribution (e.g., the perception of gesture in addition to speech), and may thus not actually be sensitive to the more subtle and complex semantic processes during GSI. In the present study, for instance, as we only observed oscillatory effects for the GG (and GR) vs. SG but not the GG vs. GR comparison, it is more likely that the oscillatory effects originate from the general processing cost of having an additional visual modality. Especially, finding a similar power decrease for GG vs. SG and GR vs. SG—although partially replicating what we have observed in the previous EEG study (He et al. 2015)—does not straightforwardly support our hypotheses that alpha power decrease is a marker for GSI. Overall, when comparing a bimodal and a unimodal condition during EEG investigations of GSI, it is difficult to disentangle the exact functional roles of alpha/beta oscillations due to the sensitivity for modality.

Our present approach of EEG-informed fMRI analysis may provide a better solution. First, in the classical EEG-informed fMRI analyses, we observed only positive correlation between BOLD and the alpha power and no BOLD–power correlation was observed for the beta power. This seems to suggest that only alpha oscillations are specifically related to the GG condition. Moreover, in our time-window analyses, although both alpha and beta power seem to be sensitive to the bimodal-auditory comparison in a time-sensitive manner, we showed that the power values in the two frequency bands correlated with distinct sets of brain regions: alpha power was found to be related to a more extensive set of regions including the left pSTS/MTG, the left IFG, the cingulate cortex, and the cerebellum. This set of regions includes the two most important regions (the left pSTS/MTG and the left IFG) which may underlie GSI, and thus may be an indication of the relevance of alpha oscillations to the general and/or semantic aspects of GSI (Willems et al. 2008, 2009). On the other hand, BOLD–beta power correlations were less likely to be related to GSI: although the conjunction analysis showed that the BOLD–beta power correlation in several right-lateralized regions including the inferior parietal gyrus/supramarginal gyrus and the superior frontal gyrus, which may be relevant to general perception of gesture (Andric and Small 2012; Buccino et al. 2001). These regions were not supported by results which single condition contrasts in the GG condition. This may be an indication that the beta band is not involved in the GSI processes. The disparate neural correlates may thus suggest distinct roles of the two frequency bands during the integration of gesture and speech. Since correlations with beta power were only observed in a small set of brain regions in the right hemisphere in early time windows (0-400 ms), one can, at most, speculate that the beta band oscillations may only be involved in the early stage gesture perception in general. Alpha power, on the other hand, was found to be correlated with the BOLD signal of multisensory integration regions including the left pSTS/MTG and the left IFG. This finding may thus be an indication of a potential functional link between alpha power and multisensory integration. This differentiation between the alpha/beta function is also supported by our additional correlation analyses: for the GG condition, while alpha power was found to be correlated with the naturalness ratings, no correlation was observed between beta power and task performance or stimulus rating scores.

In the current study, the observed positive correlation between alpha power and BOLD is particularly interesting. Across our classical and time-window-based EEG-informed fMRI analyses, the positive alpha-BOLD correlation was observed in a wide set of regions including the anterior

cingulate gyrus, the caudate, the insula, the left IFG, and the left pSTS/MTG, as well as the cerebellum. However, in the conventional fMRI analysis and the EEG analysis, we observed increased BOLD in the left pSTS/MTG and alpha power decease by comparing the bimodal and unimodal conditions, thus contradicting the direction observed in the EEG-informed fMRI analyses. Of note, recent progress in data collection and analysis has allowed a rich inventory of the literature using simultaneous EEG-fMRI (Debener et al. 2006; Eichele et al. 2005). The method is particularly informative concerning the functional characterization and neural correlates of alpha oscillations. Classically, alpha oscillations have been considered as a pivotal signature of an idling brain state (Pfurtscheller et al. 1996), and this is further supported by a range of simultaneous EEG–fMRI studies showing negative correlations between alpha power and BOLD in the lateral frontal and parietal cortices in the resting state (Laufs et al. 2003b, 2006). Moreover, negative correlations between alpha and BOLD response have also been observed in task-relevant networks during cognitively demanding tasks (Liu et al. 2014; Scheeringa et al. 2011, 2009), suggesting a relation to task demand. Interestingly, positively alpha-BOLD correlations in task-based experimental paradigms were also reported (Liu et al. 2014; Sadaghiani et al. 2010). Together, there are heterogeneous findings regarding the direction of alpha power–BOLD correlations, suggesting diverse functional relationships of the alpha oscillations depending on task and context. However, while interpreting our findings on the positive correlation between alpha power and both the left STS/MTG and left IFG, as the direction of this correlation is clearly opposite to our conventional fMRI and EEG results, it is thus unlikely that this discrepancy originates from either task or context.

One viable explanation may consider the fact that the conventional and parametric fMRI analyses reveal distinct perspectives. Recall that, in the trial-based correlation analysis between the alpha/beta power and behavioral/rating measures, we observed that alpha power correlates positively with the naturalness ratings in the GG condition. This data pattern would suggest that, for trials of semantically 'harder' gesture–speech combinations (i.e., less natural), the alpha power decreases. For these trials with lower alpha power, the integration between gesture and speech may be more difficult. As a consequence, if the increased brain activation in both the left pSTS/MTG and the left IFG (as revealed by unmodulated conventional fMRI analysis) indicates a successful combinatorial GSI process, then the parametric EEG-informed fMRI may suggest that, on a trial-by-trial basis, 'harder' trials may lead to modulated brain activity in the two fronto-temporal regions, reflecting the a relatively more difficult GSI process. This line of account, although being speculative, may explain why we observed positive correlation between the alpha power and BOLD signal in the trial-by-trial parametric analysis. Nevertheless, the correlation between both alpha power and the left fronto-temporal gesture network suggests that alpha oscillations may be related to higher order regions (e.g., the left IFG) during a cognitively demanding task.

Another interesting finding from our EEG-informed analyses lies in the positive correlation between alpha power and BOLD within a set of paralimbic regions including the caudate, the cingulate gyrus, the rolandic operculum, and the insula. The functional interpretation of this set of paralimbic regions is diverse. First, this set of regions corresponds to a 'salience network' which evaluates the processed sensory data with visceral and autonomic data, so that the organism can decide the 'salience' of the stimuli (Seeley et al. 2007). The other account suggest that these regions control goal-directed behavioral through the stable maintenance of task control and task goals (Dosenbach et al. 2007). Besides, it is also suggested that this set of regions refers to the 'tonic alertness': a sustained state of the brain that internally prepares to process and to respond (Posner 2008; Sturm et al. 2004). Despite the divergence of functional interpretation, more relevant to our finding, it is reported that this set of regions correlates positively with alpha power during the resting state (Sadaghiani et al. 2010). This link nicely echoes a line of research, which reports increased alpha power for a better task performance in breath counting and continuous auditory detection, when performance depends on tonic alertness (Braboszcz and Delorme 2011; Makeig and Inlow 1993). In our study, the positive correlation between alpha power and the BOLD increase in this set of paralimbic regions may be considered as an evidence of a putative function that is homogenous to either 'salience', 'task maintenance', or 'tonic alertness' while performing a cognitive task such as GSI: as alpha power positively correlates with the naturalness of bimodal stimuli, for these 'more natural' trials, alpha power is higher, and BOLD in these regions is higher, thus indicating the brain's state of being alert to the sensory stimuli, and being better maintained for the experimental task. Thus, we may argue that, while these paralimbic regions are responsible for the maintenance of GSI, the left fronto-temporal network (consisting of both the left IFG and the left pSTS/MTG) stands for the execution of various substages of GSI. Importantly, both functions are related to the alpha power change. This interpretation is in line with the multifaceted functional role of alpha oscillations across a variety of task-free and task-relevant scenarios (Sadaghiani and Kleinschmidt 2016).

Combining perspectives from EEG and fMRI in a time-sensitive manner

A primary goal of the current study was to add a temporal dimension to the relatively static fMRI data acquired during

gesture–speech integration. By correlating alpha/beta power from five continuous time windows, we were able to identify how alpha/beta-BOLD correlation progresses temporally. Importantly, the results can be considered as important addition to the more classical EEG-informed fMRI approach, which averages across the much wider latency range showing a significant difference between conditions. With the classical approach, we only observed alpha-BOLD correlation in the left IFG but not the left pSTS/MTG. This may suggest that although the left pSTS/MTG is crucial for GSI, its role during GSI, namely, the binding of multisensory input, is rather phasic and time-sensitive. However, with time-window analysis, we identified alpha-BOLD correlation in the left pSTS/MTG, and observed that this correlation occurred in an earlier window (200–400 ms) than the correlation with the left IFG (400–600 ms). Importantly, this temporal progression of alpha-related BOLD activation is also supported by a recent MEG study on the comprehension of co-speech gestures (Drijvers et al. 2018). In this study, during comprehension of co-speech iconic gestures, the source of the alpha power was located in the pSTS in an earlier time window after the onset of the critical word; in a much later time window, the alpha source was located in the left IFG. Our results, together with the findings from Drijvers et al. (2018), can be considered further evidence in support of the two-stage model of GSI (Dick et al. 2012b; He et al. 2015; Willems et al. 2009), as the model hypothesizes a temporal order concerning the recruitment of both regions during two sub-stages. Namely, the pSTS/MTG is hypothesized to be responsible for multimodal matching in an early processing stage, while the left IFG is recruited in a later stage during the semantic integration of gesture and speech. This two-stage model is also in accordance with proposals discussing multisensory integration in a broader sense (Tseng et al. 2015; Wildgruber et al. 2009), even though these models do not offer concrete predictions for GSI regarding the semantic complexity of individual gesture types. As discussed in the previous sections, the two-stage model is able to account for a wide range of fMRI data regarding the GSI of various types of gestures differing in the degree of semantic relation to speech.

The model is also in line with event-related potential (ERP) studies on GSI focusing on the temporal aspects of integration: for the GSI of semantically less complex gestures, e.g., beats, ERP studies have reported integration effects as early as 100–200 ms after word onset in the N1 and the P2 components (Biau and Soto-Faraco 2013). For those semantically more complex gestures, e.g., iconic gestures, ERP literature commonly suggests the GSI effect to be represented in the N400 component, which occurs much later (Kelly et al. 2004; Özyürek et al. 2007; Wu and Coulson 2010). These ERP studies, together with the findings from our time-window analysis, provide further evidence for the assumption that different aspects of GSI may occur in a temporally dissociable manner, depending on semantic complexity.

Methodological implications

Clearly, with simultaneous EEG-fMRI, our current study offers several new perspectives when compared to the previous study with identical design but using separate EEG and fMRI (He et al. 2015). First, it is evident that parametric EEG-informed fMRI and conventional, unmodulated fMRI reveal distinct perspectives. In the current study, we calculated the integration effect (the conjunction of GG > GR and GG > SG) for the bimodal condition for both the conventional and EEG-informed fMRI analyses. While we observed increased activation in left pSTS/MTG for the conventional fMRI analysis, with EEG-informed fMRI, besides the commonly reported left fronto-temporal GSI network, we additionally found a paralimbic network that correlates to the alpha power, and this network has never been identified in the GSI literature; however, both set of regions are related to alpha band oscillations, and may support different sub-processes/state that are relevant to GSI. In this regard, EEG-informed fMRI can be considered as an important complement to the conventional approach, especially for the investigation of complex processes such as GSI. Second, EEG-informed fMRI is shown to be able to disentangle mixed oscillatory effects specific to an experimental manipulation. We clearly showed dissociable brain networks that are related to the alpha and beta band oscillations, even if we were unable to disentangle the functional relevance of the two frequency bands with EEG only. Third, being a novel attempt, our time-window analysis suggests that extracting temporal information in addition to oscillation-BOLD correlation can be an informative approach. Of note, the previous simultaneous EEG–fMRI studies exploited either the temporal resolution or oscillatory aspects of EEG to inform fMRI: that is, BOLD are modeled either with the amplitude of one or more ERP components (Debener et al. 2005; Eichele et al. 2005), which lacks frequency-specific information, or with the averaged parameter of a specific frequency band within a fixed time-window (Arnstein et al. 2011; Liu et al. 2014; Sammer et al. 2007; Scheeringa et al. 2009). Clearly, even though both temporal and oscillatory aspects are important for the understanding of a specific cognitive function, they are rarely integrated together in simultaneous EEG–fMRI studies. The current study makes a first attempt: with time-window-based alpha-BOLD correlation, we identified that the left pSTS/MTG and the left IFG may be relevant to GSI in a time-sensitive manner. Although it has to be admitted that the temporal resolution of band-specific (especially for lower frequencies) time-series is coarser than that of ERP, it is, nonetheless, an important augmentation to the relatively

static fMRI data. Future studies are expected to refine the technical specification with regard to this approach, and to apply the approach to a wider range of experimental settings.

Limitations

A number of limitations may need to be considered when interpreting the findings from the present study. First, in our EEG analysis, we did not observe reliable effects when comparing the bimodal GG and the visual GR condition, whereas, in our separately acquired EEG study (He et al. 2015), an effect in the alpha band for this comparison was reported. This variation might potentially be due to differences in data collection and contamination from the BCG artifact rejection. However, importantly, due to the lack of effects in the alpha band for the GG–GR comparison, unlike the previous study, we were unable to unambiguously link the alpha band oscillations to GSI. Further replications with both simultaneous and separate data collections are necessary to determine the replicability of the effect. Second, the observed EEG effects in the alpha/beta bands were not fully identical to the effects reported with separately collected EEG in our previous study (He et al. 2015). However, it has to be admitted that both versions of data collection are carried out in different settings, this is specially a problem when we kept using identical cluster-based permutation tests for both settings, which, in some cases, might not be proper to detect the exact temporal span (Groppe et al. 2011). Therefore, even if there are latency differences between the results from two settings, we still consider the GG–SG results from two collection methods to be similar, due to the observed identical frequency range. Second, the alpha band effect for GG-SG, as observed in the current study, was only identified in a rather restricted frequency range (9–11 Hz) but not covering the full alpha spectrum (7–13 Hz). This sub-range, although partially overlaps with the alpha range reported in comparable studies at 10 Hz (Biau et al. 2015: 8–10 Hz; Sadaghiani et al. 2010: 10–12 Hz; He et al. 2015: 7–13 Hz; Drijvers et al. 2018: 8–12 Hz), may be derived from the fact that the GSI-related alpha effects can potentially subject to individual differences. Thus, further investigations focusing at the individual alpha peak during different GSI scenarios for different gesture types are expected to shed further light on this issue. Finally, it has to be noted that although we observed correlation effects between time-window-based alpha power and BOLD signal, such relationships are merely correlational in nature and should be treated with caution. Our findings suggesting the link between the left fronto-temporal network and GSI-related alpha power should be interpreted in relation to the current models and experimental findings regarding GSI. To fully test the neural dynamics of GSI within the left fronto-temporal network, further comparable experiments with high-density M/EEG data collection allowing source reconstruction and hypotheses-driven effective connectivity analysis would be more than beneficial (Dick et al. 2012a; Drijvers et al. 2018).

Conclusion

Our study is the first simultaneous EEG–fMRI investigation that explores the integration between gesture and speech at semantic level. The results from our EEG-informed fMRI analyses suggest that alpha oscillations may play multiple roles during GSI. In addition, the observed temporal progression within the left fronto-temporal GSI network corroborates that GSI is multi-stage. Our study has successfully demonstrated that simultaneous EEG-fMRI is able to provide both temporal and spatial precision for investigating information processing within neural networks. It thus has the potential to offer new insights into the investigation of gesture–speech integration and multisensory integration in a broader sense.

Acknowledgements This research project is supported by a grant from the 'Von Behring-Röntgen-Stiftung' (Project no. 59-0002; 64-0001) and by the 'Deutsche Forschungsgemeinschaft' (Project no. DFG: STR 1146/11-2, STR 1146/9-1 and SFB/TRR135 project A3). M. S. is supported by the DFG (project no. STR 1146/4-1). B. S. is supported by the DFG (Project no. STR 1146/8-1).

References

Allen PJ, Polizzi G, Krakow K, Fish DR, Lemieux L (1998) Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction. Neuroimage 8(3):229–239

Andric M, Small SL (2012) gesture's neural language. Front Psychol 3:99

Arnstein D, Cui F, Keysers C, Maurits NM, Gazzola V (2011) μ-suppression during action observation and execution correlates with BOLD in dorsal premotor, inferior parietal, and SI cortices. J Neurosci 31(40):14243–14249

Beauchamp MS (2005) See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. Curr Opin Neurobiol 15(2):145–153

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004a) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci 7(11):1190–1192

Beauchamp MS, Lee KE, Argall BD, Martin A (2004b) Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41(5):809–823

Biau E, Soto-Faraco S (2013) Beat gestures modulate auditory integration in speech perception. Brain Lang 124(2):143–152

Biau E, Soto-Faraco S (2015) Synchronization by the hand: the sight of gestures modulates low-frequency activity in brain responses to continuous speech. Front Hum Neurosci 9:527

Biau E, Torralba M, Fuentemilla L, de Diego Balaguer R, Soto-Faraco S (2015) Speaker's hand gestures modulate speech perception through phase resetting of ongoing neural oscillations. Cortex 68:76–85

Braboszcz C, Delorme A (2011) Lost in thoughts: neural markers of low alertness during mind wandering. Neuroimage 54(4):3040–3047

Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Freund HJ (2001) Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. Eur J Neurosci 13(2):400–404

Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. Science 304(5679):1926–1929

Callan DE, Jones JA, Munhall K, Kroos C, Callan AM, Vatikiotis-Bateson E (2004) Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. J Cogn Neurosci 16(5):805–816

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10(11):649–657

Calvert G, Spence C, Stein BE (2004) The handbook of multisensory processes. MIT Press, Cambridge

Debener S, Ullsperger M, Siegel M, Fiehler K, Von Cramon DY, Engel AK (2005) Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J Neurosci 25(50):11730–11737

Debener S, Ullsperger M, Siegel M, Engel AK (2006) Single-trial EEG–fMRI reveals the dynamics of cognitive function. Trends Cogn Sci 10(12):558–563

Dick AS, Goldin-Meadow S, Hasson U, Skipper JI, Small SL (2009) Co-speech gestures influence neural activity in brain regions associated with processing semantic information. Hum Brain Mapp 30(11):3509–3526

Dick AS, Goldin-Meadow S, Solodkin A, Small SL (2012a) Gesture in the developing brain. Dev Sci 15(2):165–180

Dick AS, Mok EH, Beharelle AR, Goldin-Meadow S, Small SL (2012b) Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. Hum Brain Mapp 35:900–917

Dosenbach NU, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RA, Raichle ME (2007) Distinct brain networks for adaptive and stable task control in humans. Proc Natl Acad Sci 104(26):11073–11078

Drijvers L, Özyürek A, Jensen O (2018) Hearing and seeing meaning in noise: alpha, beta, and gamma oscillations predict gestural enhancement of degraded speech comprehension. Hum Brain Mapp 1–13 (**in press**)

Eichele T, Specht K, Moosmann M, Jongsma ML, Quiroga RQ, Nordby H, Hugdahl K (2005) Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. Proc Natl Acad Sci USA 102(49):17798–17803

Foxe JJ, Snyder AC (2011) The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. Front Psychol 2:154

Goldin-Meadow S (2005) Hearing gesture: how our hands help us think. Harvard University Press, Cambridge

Goldman RI, Stern JM, Engel J Jr, Cohen MS (2002) Simultaneous EEG and fMRI of the alpha rhythm. Neuroreport 13(18):2487

Green A, Straube B, Weis S, Jansen A, Willmes K, Konrad K, Kircher T (2009) Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. Hum Brain Mapp 30(10):3309–3324

Groppe DM, Urbach TP, Kutas M (2011) Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. Psychophysiology 48(12):1711–1725

He Y, Gebhardt H, Steines M, Sammer G, Kircher T, Nagels A, Straube B (2015) The EEG and fMRI signatures of neural integration: an investigation of meaningful gestures and corresponding speech. Neuropsychologia 72:27–42

Holle H, Gunter TC, Rüschemeyer S-A, Hennenlotter A, Iacoboni M (2008) Neural correlates of the processing of co-speech gestures. Neuroimage 39(4):2010–2024

Holle H, Obleser J, Rueschemeyer S-A, Gunter TC (2010) Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. Neuroimage 49(1):875–884

Hubbard AL, Wilson SM, Callan DE, Dapretto M (2009) Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. Hum Brain Mapp 30(3):1028–1037

Järveläinen J, Schuermann M, Hari R (2004) Activation of the human primary motor cortex during observation of tool use. Neuroimage 23(1):187–192

Kelly SD, Kravitz C, Hopkins M (2004) Neural correlates of bimodal speech and gesture comprehension. Brain Lang 89(1):253–260

Kircher T, Straube B, Leube D, Weis S, Sachs O, Willmes K, Green A (2009) Neural interaction of speech and gesture: differential activations of metaphoric co-verbal gestures. Neuropsychologia 47(1):169–179

Kita S, Özyürek A (2003) What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: evidence for an interface representation of spatial thinking and speaking. J Mem Lang 48(1):16–32

Laufs H, Kleinschmidt A, Beyerle A, Eger E, Salek-Haddadi A, Preibisch C, Krakow K (2003a) EEG-correlated fMRI of human alpha activity. Neuroimage 19(4):1463–1476

Laufs H, Krakow K, Sterzer P, Eger E, Beyerle A, Salek-Haddadi A, Kleinschmidt A (2003b) Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. Proc Natl Acad Sci 100(19):11053–11058

Laufs H, Holt JL, Elfont R, Krams M, Paul JS, Krakow K, Kleinschmidt A (2006) Where the BOLD signal goes when alpha EEG leaves. Neuroimage 31(4):1408–1418

Leske S, Tse A, Oosterhof NN, Hartmann T, Müller N, Keil J, Weisz N (2014) The strength of alpha and beta oscillations parametrically scale with the strength of an illusory auditory percept. Neuroimage 88:69–78

Liu Y, Bengson J, Huang H, Mangun GR, Ding M (2014) Top-down modulation of neural activity in anticipatory visual attention: control mechanisms revealed by simultaneous EEG–fMRI. Cereb Cortex 26(2):517–529

Luo H, Liu Z, Poeppel D (2010) Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. PLoS Biol 8(8):e1000445

Makeig S, Inlow M (1993) Lapse in alertness: coherence of fluctuations in performance and EEG spectrum. Electroencephalogr Clin Neurophysiol 86(1):23–35

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164(1):177–190

McNeill D (2006) Gesture: a psycholinguistic approach. In: Brown E, Anderson A (eds) The Encyclopedia of Language and Linguistics. Elsevier, Amsterdam, Boston, pp 58–66

McNeill D (2008) Gesture and thought. University of Chicago Press, Chicago

Nagels A, Kircher T, Steines M, Straube B (2015) Feeling addressed! The role of body orientation and co-speech gesture in social communication. Hum Brain Mapp 36(5):1925–1936

Nichols T, Brett M, Andersson J, Wager T, Poline JB (2005) Valid conjunction inference with the minimum statistic. Neuroimage 25(3):653–660

Obleser J, Weisz N (2012) Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. Cereb Cortex 22(11):2466–2477

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9(1):97–113

Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Intell Neurosci 2011:1–9

Özyürek A, Willems RM, Kita S, Hagoort P (2007) On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. J Cogn Neurosci 19(4):605–616

Perry A, Bentin S (2009) Mirror activity in the human brain while observing hand movements: a comparison between EEG desynchronization in the μ-range and previous fMRI results. Brain Res 1282:126–132

Perry A, Stein L, Bentin S (2011) Motor and attentional mechanisms involved in social interaction—evidence from mu and alpha EEG suppression. Neuroimage 58(3):895–904

Pfurtscheller G, Stancak Jr A, Neuper C (1996) Event-related synchronization (ERS) in the alpha band—an electrophysiological correlate of cortical idling: a review. Int J Psychophysiol 24(1):39–46

Posner MI (2008) Measuring alertness. Ann N Y Acad Sci 1129(1):193–199

Quandt LC, Marshall PJ, Shipley TF, Beilock SL, Goldin-Meadow S (2012) Sensitivity of alpha and beta oscillations to sensorimotor characteristics of action: an EEG study of action production and gesture observation. Neuropsychologia 50(12):2745–2751

Romero YR, Keil J, Balz J, Niedeggen M, Gallinat J, Senkowski D (2016) Alpha-band oscillations reflect altered multisensory processing of the McGurk illusion in schizophrenia. Front Hum Neurosci 10:41

Sadaghiani S, Kleinschmidt A (2016) Brain networks and α-oscillations: structural and functional foundations of cognitive control. Trends Cogn Sci 20(11):805–817

Sadaghiani S, Scheeringa R, Lehongre K, Morillon B, Giraud A-L, Kleinschmidt A (2010) Intrinsic connectivity networks, alpha oscillations, and tonic alertness: a simultaneous electroencephalography/functional magnetic resonance imaging study. J Neurosci 30(30):10243–10250

Sammer G, Blecker C, Gebhardt H, Bischoff M, Stark R, Morgen K, Vaitl D (2007) Relationship between regional hemodynamic activity and simultaneously recorded EEG-theta associated with mental arithmetic-induced workload. Hum Brain Mapp 28(8):793–803

Scheeringa R, Petersson KM, Oostenveld R, Norris DG, Hagoort P, Bastiaansen M (2009) Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. Neuroimage 44(3):1224–1238

Scheeringa R, Fries P, Petersson K-M, Oostenveld R, Grothe I, Norris DG, Bastiaansen M (2011) Neuronal dynamics underlying high- and low-frequency EEG oscillations contribute independently to the human BOLD signal. Neuron 69(3):572–583

Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. Trends Cogn Sci 12(3):106–113

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Greicius MD (2007) Dissociable intrinsic connectivity networks for salience processing and executive control. J Neurosci 27(9):2349–2356

Senkowski D, Schneider TR, Foxe JJ, Engel AK (2008) Crossmodal binding through neural coherence: implications for multisensory processing. Trends Neurosci 31(8):401–409

Skipper JI, Goldin-Meadow S, Nusbaum HC, Small SL (2009) Gestures orchestrate brain networks for language understanding. Curr Biol 19(8):661–667

Slotnick SD, Schacter DL (2004) A sensory signature that distinguishes true from false memories. Nat Neurosci 7(6):664–672

Straube B, Green A, Bromberger B, Kircher T (2011) The differentiation of iconic and metaphoric gestures: common and unique integration processes. Hum Brain Mapp 32(4):520–533

Straube B, He Y, Steines M, Gebhardt H, Kircher T, Sammer G, Nagels A (2013) Supramodal neural processing of abstract information conveyed by speech and gesture. Front Behav Neurosci 7:120

Straube B, Green A, Sass K, Kircher T (2014) Superior temporal sulcus disconnectivity during processing of metaphoric gestures in schizophrenia. Schizophr Bull 40(4):936–944

Sturm W, Longoni F, Fimm B, Dietrich T, Weis S, Kemna S, Willmes K (2004) Network for auditory intrinsic alertness: a PET study. Neuropsychologia 42(5):563–568

Thut G, Nietzel A, Brandt SA, Pascual-Leone A (2006) α-Band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. J Neurosci 26(37):9494–9502

Tseng H-H, Bossong MG, Modinos G, Chen K-M, McGuire P, Allen P (2015) A systematic review of multisensory cognitive–affective integration in schizophrenia. Neurosci Biobehav Rev 55:444–452

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15(1):273–289

Uhlhaas PJ, Singer W (2010) Abnormal neural oscillations and synchrony in schizophrenia. Nat Rev Neurosci 11(2):100–113

van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. Proc Natl Acad Sci USA 102(4):1181–1186

Wildgruber D, Ethofer T, Grandjean D, Kreifelts B (2009) A cerebral network model of speech prosody comprehension. Int J Speech Lang Pathol 11(4):277–281

Willems RM, Özyürek A, Hagoort P (2007) When language meets action: the neural integration of gesture and speech. Cereb Cortex 17(10):2322–2333

Willems RM, Oostenveld R, Hagoort P (2008) Early decreases in alpha and gamma band power distinguish linguistic from visual information during spoken sentence comprehension. Brain Res 1219:78–90

Willems RM, Özyürek A, Hagoort P (2009) Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. Neuroimage 47(4):1992–2004

Wu YC, Coulson S (2010) Gestures modulate speech processing early in utterances. Neuroreport 21(7):522

Zhao W, Riggs K, Schindler I, Holle H (2018) Transcranial magnetic stimulation over left inferior frontal and posterior temporal cortex disrupts gesture–speech integration. J Neurosci 38(8):1891–1900