The Journal of Neuroscience

http://jneurosci.msubmit.net

JN-RM-1422-15R2

MEG multivariate analysis reveals early abstract action representations in lateral occipitotemporal cortex

Angelika Lingnau, University of Trento
Raffaele Tucciarelli, University of Trento
Luca Turella, University of Trento
Nikolaas Oosterhof, University of Trento
Nathan Weisz, Università degli Studi di Trento

Commercial Interest:

MEG multivariate analysis reveals early abstract action representations in lateral occipitotemporal cortex

Raffaele Tucciarelli[1], Luca Turella[1], Nikolaas N. Oosterhof[1], Nathan Weisz[1,2,3], Angelika Lingnau[1,2,4*]

1. Center for Mind/Brain Sciences (CIMeC), University of Trento, Rovereto, 38068, Italy

2. Department of Psychology and Cognitive Science, University of Trento, Rovereto, 38068, Italy

3. Centre for Cognitive Neuroscience, University of Salzburg, Salzburg, 5020, Austria

4. Department of Psychology, Royal Holloway University of London, TW20 0EX Egham, Surrey, UK

***Corresponding author:**

Angelika Lingnau, angelika.lingnau@unitn.it; angelika.lingnau@rhul.ac.uk

Conflict of interest: nothing to declare

Acknowledgements: This research was supported by the Provincia Autonoma di Trento and the Fondazione Cassa di Risparmio di Trento e Rovereto.

Abstract

Understanding other people's actions is a fundamental prerequisite for social interactions. Whether action understanding relies on simulating the actions of others in the observers' motor system or on the access to conceptual knowledge stored in non-motor areas is strongly debated. It has been argued earlier that areas that play a crucial role in action understanding should (a) distinguish between different actions, (b) generalize across the ways in which actions are performed (e.g. Dinstein et al., 2008; Oosterhof et al., 2013; Caramazza et al., 2014), and (c) have access to action information around the time of action recognition (Hauk et al., 2008). Whereas previous studies focused on the first two criteria, little is known about the dynamics underlying action understanding. We examined which human brain regions are able to distinguish between pointing and grasping, irrespective of reach direction (left/ right) and effector (left/ right hand), using multivariate pattern analysis (MVPA) of magnetoencephalography (MEG) data. We show that the lateral occipitotemporal cortex (LOTC) has the earliest access to abstract action representations, which coincides with the time point from which there was enough information to allow discriminating between the two actions. By contrast, precentral regions, though recruited early, have access to such abstract representations substantially later. Our results demonstrate that in contrast to LOTC, the early recruitment of precentral regions does not contain the detailed information that is required to recognize an action. We discuss previous theoretical claims of motor theories and how they are incompatible with our data.

Significance Statement

It is debated whether our ability to understand other people's actions relies on the simulation of actions in the observers' motor system, or whether it is based on access to conceptual knowledge stored in non-motor areas. Here we examined where in the brain and at which point in time it is possible to distinguish between pointing and grasping actions irrespective of the way in which they are performed (effector, reach direction) using magnetoencephalography (MEG) in combination with machine learning. We show that, in contrast to the predictions of motor theories of action understanding, the lateral occipitotemporal cortex (LOTC) has access to abstract action representations substantially earlier than precentral regions.

Introduction

How do we assign meaning to actions performed by other people? One of the most dominant views in the literature is the idea that action concepts are grounded in the motor system (Rizzolatti et al., 2001; Kiefer and Pulvermüller, 2012). By contrast, according to classical cognitive theories (Mahon and Caramazza, 2008; Caramazza et al., 2014), the ability to understand the meaning of other people's actions draws on conceptual representations stored outside the motor system, such as posterior temporal regions.

A region involved in action understanding should be able (a) to discriminate between different actions (action specificity), and (b) to generalize across different possible instances of a particular action (Dinstein et al., 2008; Oosterhof et al., 2013; Caramazza et al., 2014). For example, *grasping* has the same meaning for an observer irrespective of whether the movement is performed with the left or right hand, or towards the left or right side of visual space. In other words, a region important for action understanding should represent the action while generalizing across concrete instantiations such as the underlying effector or reach direction. Previous fMRI and TMS studies in humans reported abstract action representations in parietal, frontal and occipital regions (Hamilton and Grafton, 2006, 2008; Cattaneo et al., 2010; Oosterhof et al., 2010, 2013), making it difficult to draw firm conclusions regarding the ongoing debate between motor and cognitive theories. One important factor not well understood so far is the underlying temporal profile of action understanding. Such information is crucial since the two theories lead to opposite predictions: according to motor theories, motor areas should have the earliest access to abstract action representations (Pulvermüller, 2005). By contrast, according to cognitive theories, areas outside the motor system should have the earliest access to such abstract action representations.

Here we use MVPA of MEG data to examine where in the brain and at which point in time it is possible to distinguish between observed pointing and grasping irrespective of reach direction (left, right) or effector (left, right hand). In contrast to motor theories of action understanding, we show that abstract action representations are encoded in lateral occipitotemporal cortex earlier than in precentral regions.

Material and Methods

We carried out two separate experiments with two different groups of participants: one behavioral experiment to identify the time point at which the videos contained enough information to allow participants to discriminate between pointing and grasping, and an MEG experiment. The same stimuli were used for the two experiments.

Participants. Fourteen students (7 females; mean age: 23.13; sd: 2.253; all right-handed) from the University of Trento took part in the behavioral experiment and received a reimbursement of €6,- at the end of the session. A different group of seventeen students (11 females; mean age: 23.3; sd: 2.1; all right-handed) from the University of Trento with normal or corrected-to-normal visual acuity and with no neurological disorders took part in the MEG experiment. All participants received a reimbursement of €25,- at the end of the MEG session. All of them gave informed consent in accordance with the Declaration of Helsinki. The experimental procedures were approved by the Ethics Committee for research involving human participants at the University of Trento.

Stimuli. Stimuli consisted of short video-clips (833 ms) depicting simple center-out hand movements (Figure 1A). Each clip started with the hand of the actor touching the central object (a polystyrene semi-sphere) with the index finger resting in the same position. After a variable amount of time (median: 183 ms; range: [67 – 367 ms]), a center-out movement towards one of the other semi-spheres started. Movement onset was defined as the time point in which the rest position was released and hand preshaping was initiated. The video ended as soon as the hand reached one of the peripheral semi-spheres (see Figure 1A for an example trial sequence). The actions were recorded from four different actors (1 male) using a digital video camera. Only the hands (and part of the forearm) of the actors were visible in the field of view. We instructed the actors to keep the velocity and kinematics of the movements as similar as possible across the two different movements. We discarded, based on our perceptual judgment, videos in which the velocity or kinematics were too dissimilar from the others and videos in which the preshaping of the hands before movement onset could give information regarding the upcoming action, keeping a total of 80 videos (five exemplars for each combination of actor x movement type x direction). We obtained movements performed with the left hand creating specular copies of the right hand movement videos via software (Matlab, Mathworks, Natick, NA), for a total of 160 videos. On each video, we superimposed a small white cross (0.88 x 0.88°) above the central semi-sphere to enable fixation and thus to avoid possible noise in the MEG signal due to eye movements.

<< Figure 1>>

Behavioral experiment

Procedure. To identify the minimum video duration required to be able to distinguish between observed pointing and grasping, we presented participants with videos depicting pointing or grasping movements directed towards the left or right side, performed with the left or right hand. The duration of the videos was parametrically varied (167, 200, 233, or 333 ms). Participants had to classify the type of observed movement by pressing one of two possible buttons while ignoring the other two dimensions (reach direction, effector). A trial started with a fixation period (white cross) of 2 seconds. Then the video appeared for a variable duration. As soon as the video ended, the fixation cross appeared again, and participants had to indicate by button press which movement they had observed. Participants were instructed to respond as accurately as possible. Video duration, type of movement, effector and reach direction were randomized. Each participant completed four experimental runs of approx. 5.5 min, for a total of 512 trials (64 repetitions per conditions). Stimuli were presented on a CRT monitor (ViewSonic Grapic Series G90fB; screen resolution: 1280x1024, refresh rate: 60 Hz) placed approx. 64 cm in front of the participants.

Statistical analysis. The aim of the behavioral experiment was to individuate the point in time in which the two actions started to perceptually diverge. To compute the accuracy for discriminating between the two observed actions as a function of video duration, we divided the number of correct classifications by the total number of trials, separately for each video duration and each participant, collapsing across effector (left, right hand) and reach direction (left, right). We then used a chi-squared test to assess at which video duration the accuracy was higher than chance level (50%).

MEG experiment

Procedure. We presented participants (N = 17) with short videos (833 ms) of reach-to-point and reach-to-grasp movements performed with either the left or right hand towards peripheral targets on the left or right side (Figure 1A) while measuring their brain oscillatory activity. We used a 2x2x2 factorial design (Figure 1B), varying the type of movement (pointing/grasping), the effector (left/right hand), and reach direction (left/right). Each trial consisted of the following events (Figure 1A): a green fixation cross (blink phase: 800 ms), a white fixation cross (fixation phase: randomly jittered within 2000-2500 ms), the video (video phase: 833 ms), and a white fixation cross (resting phase: 1000 ms). Trial duration varied from 4633 ms to 5133 s depending on the duration of the fixation phase. The blink phase at the beginning of each trial allowed participants to blink during a controlled time window and thus reducing the probability of blinking during the fixation phase or during video presentation. Participants were instructed to blink every time they saw the green cross. During the fixation phase, participants had to maintain fixation on the white cross. We jittered the fixation phase to avoid that participants could predict the appearance of the video that could have caused anticipation of the neural response. When the video appeared, participants were asked to keep fixating on the cross and to globally pay attention to the ongoing movement. In particular, in contrast to the task used in the behavioral experiment, we asked them to attend to all three dimensions we manipulated, i.e. movement type, effector, and reach direction. During the resting phase, participants had to keep fixating and to wait for the green cross that indicated the beginning of a new trial.

To ensure that participants were paying attention to the video, we introduced catch trials (10% of all trials) during which we presented a question regarding one of the three dimensions (e.g. 'was the direction to the left?'). Catch trials were presented occasionally with the following constraints: 1) if trial N was a catch trial, trial N+1 could not be a catch trial; 2) no catch trial during the first trial of a run. A catch trial was identical to an experimental trial except for the question that appeared at the end of the catch trial (1 sec after video offset). Since participants did not know when a catch trial would appear, and what the question would be, they had to pay attention to each video and to each of the three dimensions to perform the task correctly. The answer was always binary (yes or no) and participants had MEG-compatible buttons for answering to the questions. The assignment of the response to the two buttons changed randomly for each question to avoid any potential confounds related to motor preparation. Eye movements were

monitored using the OEM system (OEM eye tracker, SMI; 60 Hz sampling rate). After each response, feedback was provided (a cartoon smiling or a sad face).

Each participant performed 10 runs, consisting of 64 trials, plus 6 catch trials, for a total of 640 experimental trials and 60 catch trials. The number of repetitions for each factorial combination (movement type x effector x reach direction) per participant was 80. Before entering the shielded room, participants familiarized with the stimuli and the task. Each run lasted from 4.9 to 5.5 minutes, depending on the duration of the fixation phase, for a total duration of the session of approximately 52 minutes. At the end of each run, participants were allowed to rest for a few minutes before a new acquisition started.

Stimuli were projected on a screen (1280 x 1024 pixels screen resolution, 60 Hz refresh rate) that was placed about 130 cm in front of the participant. The screen was visible as a rectangular aperture of about 21.7 x 13.16°. We controlled visual stimulation during the behavioral and the MEG sessions using ASF (Schwarzbach, 2011), a toolbox for Matlab (Mathworks, Natick, MA) based on the Psychtoolbox (Brainard, 1997).

MEG data acquisition and analysis. At the beginning of the MEG session, the head shape of each participant was digitally acquired using the Polhemus system (Polhemus, Colchester, VT). Moreover, we placed three coils at the participant's forehead and two behind the ears to acquire the head position of each participant within the MEG helmet at the beginning of each run. Prior to entering the shielded room containing the MEG system, participants were asked to remove all magnetic materials that could distort the measurement.

We measured neuromagnetic brain activity using a 306-channels whole head MEG system (Neuromag Elekta Oy, Helsinki Finland) at a sampling rate of 1000 Hz. The system consists of 204 planar gradiometers and 102 magnetometers arranged in a helmet configuration. Here, we are reporting results of the gradiometers only. Triggers were sent at video onset to synchronize stimulus presentation with neural activity. To check for the correct timing of the stimuli, and to take into account possible delays of the stimulus presentation with respect to the triggers, we used a photodiode on the stimulation screen inside the shielded room.

MEG data preprocessing. We analyzed data using the open source Matlab-based Fieldtrip toolbox (Oostenveld et al., 2011). Continuous data were cut into epochs from -1 to 1.3 sec relatively to video onset. Epochs were high-pass filtered at 1 Hz to remove very slow frequencies and Direct Current (DC) offset. Frequencies due to the electrical system were also filtered out using a band-stop filter (Butterworth IIR filter) at 50 Hz and its harmonics (100 and 150 Hz). Trials with blinks or

eye movements during the presentation of the video or during the baseline period were discarded on the basis of the information from the eye tracker. In addition, we visually inspected trials for artifacts, blind to the condition, and rejected trials that were clearly affected by external noise or spike current. On average, we rejected 13% of the trials per participant. If a sensor was very noisy for the entire experimental session, it was rejected. In order to have the same number of sensors for each participant, missing sensors were reconstructed by interpolation of the neighbors.

Time-frequency analysis. To obtain a time-frequency representation of the oscillatory activity associated with movement observation, we applied Fourier transformation to sliding time windows with fixed length of 500 ms. The sliding window moved in steps of 50 ms; power was calculated for frequencies in a range from 2 to 40 Hz in steps of 2 Hz. To avoid spectral leakage and to control for frequency smoothing, a Hanning taper was applied prior to Fourier transformation. Subsequently, for the univariate analysis only, power was averaged across effector and reach direction, and the spectral power was normalized relative to baseline (-0.5 to -0.3 s with respect to the onset of the video, i.e. during a sub-period of the fixation phase).

Source analysis. Neural sources were found using dynamic imaging of coherent sources (DICS), a frequency domain beamforming technique (Gross et al., 2001). We chose the frequencies and times of interest based on the sensor level analysis. Specifically, we considered the sensor with the greatest accuracy of the classifier (multivariate analysis) to distinguish between pointing and grasping, generalizing across effector and reach direction, in those frequency bands that survived the multiple comparison tests. Note that, given the way the sensors were selected, source analysis merely served as a visualization of the sources.

For each participant we used a volume conductor model using the single-shell method (Nolte, 2003). The models were built warping a dipole grid based on a MNI template brain to fit the individual head shape of each participant. We proceeded with DICS for each separate condition using a common spatial filter computed from the combination of the two conditions. In this way, any difference between the two conditions cannot be ascribed to differences between the filters.

MEG Statistical analysis (sensor level). We carried out both uni- and multivariate analyses in sensor space, followed by a beamforming analysis (Gross et al., 2001) to identify sources explaining any observed effects. Univariate analysis was conducted in order to observe the classical decrease in power in alpha and beta bands (Cochin et al., 1999; Pineda, 2005; Hari, 2006). Importantly, to identify at which sensors and at which point in time it is possible to distinguish

between the two movements on the basis of the MEG signal, we applied multivariate analysis on the computed power and the sources adopting an algorithm developed for the analysis of fMRI data (Oosterhof et al., 2012a).

Behavioral analysis (MEG experiment). Participants' accuracy in answering the questions in the catch trials during the MEG experiment was evaluated online by observing the feedback provided after each catch trial. All participants were able to answer the questions and typically made two or three mistakes within the entire session (mostly at the beginning of the experiment). We are thus certain that participants were attending to the videos.

Univariate analysis. Note that in contrast to the multivariate analysis, in which we specifically targeted regions that show movement selectivity *generalizing across effector and reach direction*, the purpose of the univariate analysis was to identify areas with less specific properties. In particular, as a quality control, we examined whether we obtain the typical decrease in the alpha and beta band during action observation (Cochin et al., 1999; Pineda, 2005; Hari, 2006). Furthermore, we aimed to determine which frequency bands and which sensors are modulated differently during pointing and grasping when *collapsing across effector and reach direction*.
All the experimental conditions were baseline corrected by subtracting the fixation period (from -0.5 to -0.3) from the post-stimulus period (from 0 to 1.3 s). To assess the difference between pointing and grasping, we used a non-parametric method (permutation test), with a cluster method for multiple comparison correction (Maris and Oostenveld, 2007) with participants as units of observation. In brief, we computed *t* scores between the two movements for each sensor-frequency-time bin. The *observed* cluster-level statistic was obtained by summing the *t* scores of neighboring bins (in time, frequency and sensors) exceeding an a priori defined critical value ($p<0.05$). We repeated the procedure 1000 times by swapping the condition labels and we obtained the distribution of *permuted* cluster-level statistics. At each iteration, the maximum cluster-level statistic was considered to control for type I error. The p value was the proportion of permuted cluster-level statistics that exceeded the *observed* cluster-level statistic. If the p-value was less than 0.05, the cluster was taken as significant.

Multivariate analysis. The assumption behind multivariate analysis in MEG is that the processing of each stimulus category is associated with a specific neural activity that induces an oscillatory signal (or neural pattern) consisting of a unique combination of sensor, time, and/ or frequency. Multivariate analyses exploit differences in terms of these patterns of activations. By contrast,

univariate analyses do not consider such patterns, but address whether two conditions differ in terms of the average response of a single variable (e.g. averaged frequency over time). This is why multivariate analyses are more *sensitive* than univariate analyses (Haxby et al., 2001; Kriegeskorte et al., 2006). Importantly, multivariate analysis allows analyzing whether the *representational content* of an area – examined via the underlying neural pattern - generalizes across low-level features. In our case, we aimed to identify regions in which the unique neural patterns associated with pointing and grasping generalized across effector (left or right hand) and reach direction (left or right; for a schematic overview, see Figure 2). We trained a classifier to discriminate between the two types of movements using the spatio-spectral-temporal MEG signal (for details, see next paragraph) related to movements performed with one of the two effectors and towards one of the two directions. We then tested on the opposite combination of effector and direction. For example, we trained a classifier to distinguish between observed grasping and pointing actions performed with the left hand towards the left, and tested the classifier to distinguish between observed grasping and pointing performed with the right hand towards the right. In this way, above chance classification could only be due to information related to the type of movement, and not to low-level perceptual features.

Analyses were performed using CoSMoMVPA, an MVPA toolbox in Matlab (Oosterhof et al., in preparation) The toolbox provides an adapted version of the multivariate searchlight approach (Kriegeskorte et al., 2006), an information-based algorithm that allows analyzing the neural contents adopting a multivariate approach at each location in the brain. In this analysis, we used local 'neighborhoods' of features in channel-time-frequency space. We used a sensor radius of 1, a time radius of 100 ms, and a frequency radius of 8Hz. For a given 'center' feature [a (sensor, time, frequency)-triple], its neighbors consisted of features for which its sensor, time, and frequency where all within the corresponding radii.

The main steps used in the multivariate analysis (for a schematic illustration, see Figure 2) were: 1) compute the time-frequency representation separately for each sensor and each trial (Figure 2A); 2) select the 'central' feature and its neighbors in time-frequency-sensor space (dashed rectangles in the insets in Figure 2A; for an enlarged view, see Figure 2B); 3) create a feature vector for each trial by selecting all features in its neighborhood (Figure 2C) and normalize (z transform) the data; 4) create independent partitions for training and testing the classifier (see Table 1); 5) train the classifier; 6) test the classifier. We repeated the steps from 2 to 5 for each sensor and for each time and frequency bin, and the classification result for each center feature was assigned to its

corresponding location in time-frequency-sensor space. For classification, we used a Support Vector Machine (SVM) algorithm, a type of classifier that looks for linear combinations of features to create a decision boundary to discriminate between two classes or stimuli (e.g. Mur et al., 2009; Pereira et al., 2009).

<< Figure 2 >>

To create subsets of trials to feed the classifier with the aim of differentiating between neural responses related with the observation of grasping and pointing actions irrespective of effector and reach direction, for each subject we divided the dataset in two independent halves, each containing only movements with a complementary combination of effector and reach direction. The first half contained left hand movements to the right and right hand movements to the left, and the second half left hand movements to the left and right hand movements to the right. We further divided the data in independent *chunks*, each of which contained at least 136 trials (depending on the number of trials remaining after artifact rejection) of a specific condition of interest. Then, for each half, we adopted a leave-one-chunk-out cross-validation method. We used 3 chunks associated to a specific condition for training, and a corresponding chunk with the complementary effector and direction for testing (cross-condition classification). This procedure was repeated for all chunks. Note that within a chunk the only dimension that differed across trials was the type of movement: grasping vs pointing. Thus, we assumed that the classifier learnt to discriminate between these two classes of stimuli. For example, if the training dataset contained the conditions grasping to the right with the right hand and pointing to the right with the right hand, the testing dataset contained the conditions grasping to the left with the left hand and pointing to the left with the left hand. For this type of classification, the classifier had to rely on differences between the two types of movements. If the model was able to discriminate between the two movements in the independent subset, this indicates that it had learnt the difference between the two types of movements using the previous training subset, generalizing across effector and reach direction. We adopted this approach for each possible factorial combination (for a complete list, see Table 1).

<< Table 1 >>

The testing phase provided accuracy maps for each participant reflecting the classifiers' performance in discriminating between the two observed movements irrespective of effector and

reach directions [in a similar way as traditional fMRI searchlights (Kriegeskorte et al., 2006), except that the features consist of (sensor, time, frequency)-triples rather than voxels]. We thus had information regarding where, when, and at which frequency band it was possible to distinguish between the abstract neural representations of the two movements.

In order to assess the reliability of the performance of the classifier, we used a non-parametric method (permutation test, similar to the procedure described above for the univariate analysis; see Maris and Oostenveld, 2007). In this case, we used the difference between the obtained classification accuracy and chance level accuracy (the accuracy expected under the null-hypothesis of no difference between the two conditions, meaning 50%) to compute the test statistic used in the permutation steps (see univariate method).

Any effect observed at sensor level has to be generated by neural sources. To visualize the sources underlying the cross-decoding effects for the frequency bands and time windows observed at the sensor level, we conducted a multivariate analysis at the source level, adopting the same searchlight approach as before (Kriegeskorte et al., 2006). Note that multivariate analysis was necessary here to identify which regions of the brain represented actions at an abstract level (i.e., generalizing across effector and reach direction). We reconstructed the source activity for the frequency bands and time windows that were significant at sensor level and expected to identify which regions of the brain were able to decode between grasping and pointing across effector and reach direction. We obtained estimates of frequency power at each grid point using a beamformer algorithm (see previous section) on a single trial basis. A searchlight was defined taking the power values at each grid point with its neighbors in a circle of 2 cm radius. For each participant, we found the accuracy maps indicating the performance of the classifier in discriminating between the two observed movements (irrespective of effector and reach direction). For descriptive purposes, we are reporting the clusters showing the greatest classification accuracy.

Results

Behavioral experiment

We computed a chi-square test to evaluate at which time point participants' performance was significantly higher than chance level (50%). We found that performance of the participants was not different from chance level at 167 ms (X-squared = 11.7307, df = 13, p = 0.5498) and at 200 ms (X-squared = 21.4835, df = 13, p = 0.0639). Performance was significantly higher than chance level from 233 ms onwards (X-squared = 58.0318, df = 13, p = 1.178e-07). This means that participants were unable to distinguish the two actions if videos were shorter than 233 ms. Since mean movement onset in the videos (defined as the time point at which the rest position was released and hand preshaping was initiated; see *Stimuli*) was 191 ms (SD: 90 ms; median: 183 ms), this indicates that the two actions were perceptually indistinguishable before movement onset.

<< Figure 3 >>

MEG experiment

Univariate analysis. We first analyzed the MEG signal using classical univariate methods to assess whether the stimuli induced a modulatory activity in the ongoing oscillations relative to rest. Low frequency bands such as alpha- and beta-bands are typically characterized by a decrease in power presumably due to neuronal activity synchronization in specific brain regions (Pfurtscheller and Lopes da Silva, 1999), indicating neural processing of the stimulus. Univariate analyses comparing the activation period (after video onset) with the baseline (before video onset) demonstrated that passive observation of pointing and grasping modulates alpha- (8 – 12 Hz) and beta- (15 – 25 Hz) band power over posterior, parietal and frontal sensors, as well as the theta-band (4 - 7 Hz). Figure 4A shows one central sensor for illustrative purposes. In the depicted sensor, the alpha and beta rebounds related to post-observation processes are evident. Dotted lines approximately indicate the different stages of the movement (see Figure captions for details).

The decrease in power that we observed in the alpha and beta bands is in line with previous studies (Pineda, 2005; Hari, 2006) and has been suggested to reflect sensorimotor system activity. Further, the increase in power in the theta- and low alpha- (4– 8 Hz) band has been observed during memory tasks (Jensen and Tesche, 2002). In addition, these low frequencies have been

reported to be modulated during action observation, both in humans (Frenkel-Toledo et al., 2013; Pavlidou et al., 2014a, 2014b) and monkeys (Caggiano et al., 2014; Kilner et al., 2014).

<<Figure 4>>

A direct comparison of grasping and pointing movements (collapsing over effector and reach direction; see Materials and Methods, Univariate Analysis) showed a significant differential modulatory activity in beta (central frequency: 24 Hz) and alpha (central frequency: 16 Hz) band power over sensorimotor sensors at a late latency only (from around 750 to 1100 ms, and from around 500 to 750 ms, respectively). Figure 4B illustrates this effect for the same representative significant sensor as in Figure 4A over central regions. Bluish colors indicate that the power decrease is greater for grasping than for pointing; reddish colors indicate the opposite. Figure 4C-D shows the topography representations of the significant sensors in two selected subsets of frequency bands and time windows that were all located over central and right central sensors. These results show that (a) the brain processes the two actions as being different, and that (b) sensorimotor areas might be involved. The fact that *grasping* induces a greater decrease than pointing could be due to the higher complexity of this movement, which in turn is likely to recruit more neural sources. However, this differential activity seems quite late (at around 600 ms), long after the two movements were perceptually distinguishable. Thus, there must be another, earlier, process that allows discriminating the two movements, which the univariate analysis did not reveal.

Multivariate analysis. Figure 5A-C shows the results of the multivariate analysis at sensor level. Two types of representations are provided: 1) a time-frequency representation, to show the dynamics of all the considered frequencies at each time point in a specific subset of sensors (panel A); 2) a topographical representation, to show the spatial information at specific time points and frequency bands (panel B, C). The inset of Figure 5A shows the two time-frequency clusters that survived the multiple comparisons correction. The lateral plots show the averaged *t* values over the sensors highlighted on the two topoplots in the middle. We observed that the classifier was able to significantly ($p<0.05$; corrected for multiple comparisons using a cluster-based method; maximum accuracy: 53.46%) discriminate between the two observed movements, generalizing across effector (left and right hand) and reach direction (left and right) over posterior sensors as early as 150 ms and lasting until 550 ms in the low alpha/theta range (see Figure 5A, left panel; for a direct comparison with univariate analysis, see Figure 7). By contrast, significant discrimination

over more anterior sensors was possible only within a window of 550 – 1200 ms, i.e. at a late stage of the video, when the hand interacts with the object (see Figure 5A, right panel). Figure 5B-C show the topographies at different times and frequencies, selected according to the following criteria:

1) As time of interest, we selected the central point of the time windows selected based on the significant clusters that survived the significance test, i.e. 400 ms [200 - 600ms] for the cluster obtained in the earlier time window, and 900ms [600 - 1200ms] for the cluster obtained in the later time window.

2) Frequency bands were chosen based on previous studies showing a modulation of the low alpha (8-10 Hz) and high-theta (6-8 Hz) bands (Frenkel-Toledo et al., 2013) and the high alpha- (8-14 Hz) and beta- (15-25 Hz) bands during action observation (e.g. Pineda, 2005). For each time of interest (400ms and 900ms), we selected the peak frequency within each considered frequency band (i.e., 6, 8, 10, and 18Hz).

To examine the cortical sources of the effects shown in Figure 5A-C, we carried out another multivariate analysis at source level, adopting the same cross-comparisons as we did for the sensor analysis (see Materials and Methods for details). To find the sources at 400 ms for the frequencies 6Hz and 8Hz, we used temporal smoothing of 4Hz and time windows of 150ms to 650ms and 212ms to 587ms, respectively. Figure 5D-E shows the decoding accuracies of all the sources projected on surface template MNI brains, thresholded to retain only those voxels with the 10% of the highest accuracies (for the corresponding mean and individual decoding accuracies, see Figure 6). For the 6 Hz signal, the highest decoding accuracies were found bilaterally in the LOTC, extending into the inferior temporal gyrus and the superior temporal gyrus in the right hemisphere, and the left superior parietal cortex, extending into the inferior parietal cortex (Figure 5D, left panel; see Table 2 for MNI coordinates of the peak voxel in each cluster). The highest decoding accuracies for the 8Hz signal were located in the left LOTC (Figure 5D, right panel), slightly anterior to the source identified at 6Hz.

Regarding the sources related to the decoding obtained in the late time window, we chose 900 ms as time of interest for the frequencies 10Hz and 18Hz (time windows: 600-1200ms and 678-112ms, respectively; smoothing: 3Hz). For the 10 Hz signal, we obtained the highest decoding accuracies in right precentral gyrus (Figure 5E, left panel). For the 18Hz signal, we obtained highest decoding accuracies in the right inferior frontal gyrus (Figure 5E, right panel).

<< Figure 5 >>

<< Figure 6 >>

<< Table 2 >>

<< Figure 7 >>

To show a complete overview of the temporal dynamics of the neural decoding at sensor space, we plotted the decoding accuracy (expressed in *t* values) for separate time bins (50-150, 150-250, 250-350, 350-450, 450-550, 550-650 ms for the early observed decoding, Figure 8A; 350-450, 450-550, 550-650, 650-750, 750-850, 850-950, 950-110 ms for the late observed decoding, Figure 8B), averaged across frequency bands (theta: 2-6 Hz; low alpha: 7-9; alpha: 9-11; beta: 17-19). This figure shows how the effect over posterior sensors evolves over time, and that anterior sensors do not show up before around 700 ms.

<< Figure 8 >>

To further evaluate the reliability of the classifier, we also used a simulation approach. Specifically, we ran a Monte Carlo simulation to estimate the probability of finding an accuracy of 53.46% under the null hypothesis of chance accuracy. The cross-validation partitioning scheme divided the data into two independent halves (see Table 1 and Methods), with the first half containing left hand rightwards and right hand leftwards trials, and the second half containing right hand rightwards and left hand leftwards trials. In each independent half, there were two folds, with a minimum of 136 trials (across participants and halves) after rejecting trials with artifacts and balancing the partitions so that each of the two actions occurred equally often. For each participant separately, we found that the correlation of classification accuracies for the test sets in two folds to be r=0.3289 (median across participants and the two independent halves). Thus, in our simulation we used the same value as follows. For each permutation, uniformly distributed (on the interval [0, 1]) random data was generated for two independent halves, two folds, 136 samples, 17 participants. To assess the effect of dependency we used 3 sets of independently normally distributed data i_1, i_2 and i_{common}. To match the correlation between accuracies, for each independent half of the data, data was made dependent through $d_1 = i_1 * \gamma + i_{common} * (1 - \gamma)$ and $d_2 = i_2 * \gamma + i_{common} * (1 - \gamma)$, with $\gamma = 0.415$ found through binary search to match the correlation (r=0.3289) across dependent folds as observed in the original data. For each iteration, classification accuracy

468 was simulated by dividing the number of samples that exceeded 0.5 in d_1 and d_2 by the number of
469 samples. 0.5 was subtracted to obtain classification accuracies relative to chance.

470 To assess the effect of independence, we also ran the same analysis setting γ=0 (corresponding to
471 r=0, i.e. full independence between folds), and γ=1 (corresponding to r=1, i.e. full dependence
472 between folds).

473 We used 100,000 iterations and found that the maximum classification accuracies found in the
474 data (using r=0.3289 for fold correlation) was significant, $P_{MC,sensor,\ r=0.3289}<0.00001$; for the latter,
475 no iteration showed a higher mean than that observed in the data (Figure 9). We obtained similar
476 results for the additionally simulated cases of fully independent folds (r=0), $p_{sensor,\ r=0.00}<0.00001$,
477 and dependent folds (r=1), $p_{sensor,\ r=1.00}<0.00001$.

478 << Figure 9 >>

479

480

Discussion

Using MVPA of MEG data, we found that LOTC has the earliest access to abstract action representations. By contrast, precentral regions, though recruited relatively early, have access to abstract action representations substantially later than LOTC. Behavioral data indicated that participants were not able to distinguish between the two actions before 233 ms, and this latency is comparable with the one observed in LOTC.

Early abstract action representations in occipito-temporal and parietal regions

Although MEG has a lower spatial resolution than fMRI, we can confidently say, based on the topographical results and source analysis, that the source that accounted for the decoding effect we found at the early stage was located within the left and right LOTC. LOTC hosts regions sensitive to body parts, kinematics, body postures, manipulable objects, and observed movements (Valyear and Culham, 2010; Downing and Peelen, 2011; Buxbaum et al., 2014; Pavlidou et al., 2014a, 2014b; Lingnau and Downing, 2015). LOTC has been shown to be modulated when participants are required to process the meaning, in comparison to the effector, involved in an action (Lingnau and Petris, 2013). Moreover, LOTC is recruited during the semantic processing of verbs (e.g. Papeo et al., 2014), and lesions to this region are associated with impairments in action recognition (Kalénine et al., 2010; Urgesi et al., 2014). In line with this view, a recent lesion study demonstrated that lesions to primary motor, somatosensory and inferior parietal lobule were accompanied by impaired action performance. By contrast, lesions to posterior LOTC were associated with impaired action recognition, whereas lesions to anterior LOTC were accompanied by impairments in both tasks (Tarhan et al., 2015). Taken together, these studies suggest that LOTC is well suited to integrate various sources of information that are crucial for action understanding.

Neuroimaging studies using MPVA of fMRI data have recently shown that LOTC also contains abstract representations of observed actions, e.g. action representations that generalize from action execution to action observation and vice versa (Oosterhof et al., 2010), that generalize across viewpoint (first person, third person; see Oosterhof et al., 2012a), kinematics (Wurm & Lingnau, 2015), and the object involved in the action (Wurm & Lingnau, 2015; Wurm, Ariani, Greenlee, & Lingnau, 2015). Importantly, our study shows that such abstract representations are available before observing this kind of representation in precentral regions, around the time when there is enough information in the stimuli to distinguish between the two types of actions. Our findings are compatible with cognitive theories of action understanding that predict the earliest

encoding of the meaning of an action outside the motor system. By contrast, our results are not compatible with motor theories of action understanding that would predict the earliest access to abstract action representations in precentral regions.

The fact that we observed abstract action representations in LOTC earlier than in precentral regions is compatible with a framework suggested by Kilner (2011). According to this view, the middle temporal gyrus (MTG) in the LOTC and the anterior portion of the IFG (aIFG) encode the most likely goal or intention of an action (e.g. grasping an object), which is communicated to the posterior portion of the IFG, where the most likely action is selected. In this framework, the role of the posterior IFG would be to generate a concrete instance of the action (e.g. grasping an object on the left using the right hand) through motor simulation. In contrast to motor theories of action understanding, the role of this motor simulation would not be to provide access to the meaning of the action, but rather to contribute to the generation of the predicted sensory consequences of the most likely action.

We observed abstract action representations at around 400 ms in the left SPL as well, extending into the inferior parietal lobule (IPL). This result is in line with previous monkey (Fogassi et al., 2005; Rizzolatti et al., 2014) and human fMRI studies (Grafton and Hamilton, 2007; Oosterhof et al., 2010, 2012b; Leshinskaya and Caramazza, 2015; Wurm and Lingnau, 2015; Wurm et al., 2015) suggesting that, similar to LOTC, this region contains abstract action representations. The observation that IPL/SPL has access to abstract action earlier than precentral regions, raises the possibility that this region might play an intermediate role between LOTC and precentral regions (see also Wurm et al., 2015). In line with this view, Pavlidou et al (2014b) demonstrated that the difference between plausible and implausible actions is first obtained over left temporal sensors, followed by parieto-occipital and sensorimotor sensors.

Late abstract action representations in precentral regions

The contrast *observation vs baseline* showed a modulation of the high alpha and beta frequency bands over central sensors during passive action observation (Figure 4E), an effect that has been suggested to be related to sensorimotor processing in motor and premotor regions (Pineda, 2005). Although we observed an early modulation of high alpha and beta frequencies in precentral regions for observation versus baseline, these regions had access to abstract representations of the observed actions substantially later than the time at which the actions were distinguishable. This finding makes a determinant role of precentral regions in action understanding implausible. In

line with this view, damage to precentral regions does not necessarily impair the ability to understand actions (Negri et al., 2007; Kalénine et al., 2010; but see Pazzaglia et al., 2008). If precentral regions do not play a determinant role in action understanding, what could be the alternative role of the late abstract action representations we obtained in these regions? Since LOTC and precentral regions are functionally interconnected (Kilner, 2011; Nelissen et al., 2011; Turken and Dronkers, 2011; Engel et al., 2013; Papeo et al., 2014), higher-level representations in precentral regions have been suggested to be a result of information spreading throughout the network (Mahon and Caramazza, 2008). Instead of providing access to the meaning of an action, precentral regions thus might be recruited to plan an appropriate movement in response to the observed action as a consequence or in parallel to the process of action understanding.

Potential caveats

One potential limitation regarding the interpretation of our results is related with the fact that one of the main distinctions between pointing and grasping, next to the pre-shaping of the hand, is the number of fingers involved. It is therefore difficult to disentangle whether our classification is based on the number of fingers involved in the movement, the pre-shaping of the hand while approaching the target, or a combination of the two. Note that pointing and grasping movements are defined both by the number of fingers involved and by the hand configuration; in other words, understanding actions could rely on the number of used fingers as well as on an examination of the pre-shaping of the hand.

Another possible criticism could be that we were able to distinguish between the two movements based on the MEG signal as early as 150 ms, which seems counterintuitive given that the mean movement onset in the videos was around 191 ms. There are several not mutually exclusive explanations for this observation. First, movements started before 150 ms in 43.8% of the videos (see Material and Methods). By contrast, the peak of decoding from the MEG signal was obtained at around 300 ms. Second, we had to apply a certain amount of temporal smoothing during time-frequency computation and during the searchlight analysis (see Materials and Methods). Consequently, when the algorithm analyzes the time bin at 150 ms, it also considers information present at 200 and 250 ms, which contained more information about movement type. This means that the absolute latency at which the two actions can be distinguished based on the MEG signal has to be interpreted with a grain of salt. Importantly, we do not aim to draw strong conclusions regarding the *absolute* onset at which movements can be decoded in the different regions, but rather about the *relative* difference between putative regions involved in action understanding.

Thus, our conclusion still holds: LOTC encodes abstract representation of actions earlier than precentral regions.

One might argue that although we observed the strongest source in the early time window within LOTC, the source analysis also revealed a small left frontal region. This frontal source is very likely generated by a single temporal source, in line with the observation that no frontal sensors showed significant decoding in this early time window (Figure 8). Note that the absence of a frontal source in the early time window does not proof that such a source does not exist. What we can state with a certain confidence, though, is that the same analysis that revealed a strong and reliable source in LOTC did not reveal any frontal source in the early time window.

Conclusion

Our results demonstrate that LOTC has access to abstract action representations substantially earlier than precentral regions, in line with the idea that action understanding occurs outside the motor system, with subsequent activation of precentral regions due to information provided from LOTC. Our results therefore provide important constraints for biologically plausible models of action understanding.

References

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Buxbaum LJ, Shapiro AD, Coslett HB (2014) Critical brain regions for tool-related and imitative actions: a componential analysis. Brain 137:1971–1985.

Caggiano V, Giese M, Thier P, Casile A (2014) Encoding of point of view during action observation in the local field potentials of macaque area F5. Eur J Neurosci:1–11.

Caramazza A, Anzellotti S, Strnad L, Lingnau A (2014) Embodied Cognition and Mirror Neurons: A Critical Assessment. Annu Rev Neurosci 37:1–15.

Cattaneo L, Sandrini M, Schwarzbach J (2010) State-dependent TMS reveals a hierarchical representation of observed acts in the temporal, parietal, and premotor cortices. Cereb Cortex 20:2252–2258.

Cochin S, Barthelemy C, Roux S, Martineau J (1999) Observation and execution of movement: similarities demonstrated by quantified electroencephalography. Eur J Neurosci 11:1839–1842.

Dinstein I, Thomas C, Behrmann M, Heeger DJ (2008) A mirror up to nature. Curr Biol 18:R13–R18.

Downing PE, Peelen M V (2011) The role of occipitotemporal body-selective regions in person perception. Cogn Neurosci 2:186–203.

Engel AK, Maye A, Kurthen M, König P (2013) Where's the action? The pragmatic turn in cognitive science. Trends Cogn Sci 17.

Fogassi L, Ferrari PF, Gesierich B, Rozzi S, Chersi F, Rizzolatti G (2005) Parietal lobe: from action organization to intention understanding. Science 308:662–667.

Frenkel-Toledo S, Bentin S, Perry A, Liebermann DG, Soroker N (2013) Dynamics of the EEG power in the frequency and spatial domains during observation and execution of manual movements. Brain Res 1509:43–57.

Grafton ST, Hamilton AFDC (2007) Evidence for a distributed hierarchy of action representation in the brain. Hum Mov Sci 26:590–616.

Gross J, Kujala J, Hamalainen M, Timmermann L, Schnitzler A, Salmelin R (2001) Dynamic imaging of coherent sources: Studying neural interactions in the human brain. PNAS 98:694–699.

Hamilton AFDC, Grafton ST (2006) Goal representation in human anterior intraparietal sulcus. J Neurosci 26:1133–1137.

Hamilton AFDC, Grafton ST (2008) Action outcomes are represented in human inferior frontoparietal cortex. Cereb Cortex 18:1160–1168.

Hari R (2006) Action-perception connection and the cortical mu rhythm. Prog Brain Res 159:253–260.

626 Hauk O, Shtyrov Y, Pulvermüller F (2008) The time course of action and action-word
627 comprehension in the human brain as revealed by neurophysiology. J Physiol Paris 102:50–
628 58.

629 Haxby J V, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping
630 representations of faces and objects in ventral temporal cortex. Science 293:2425–2430.

631 Jensen O, Tesche CD (2002) Frontal theta activity in humans increases with memory load in a
632 working memory task. Eur J Neurosci 15:1395–1399.

633 Kalénine S, Buxbaum LJ, Coslett HB (2010) Critical brain regions for action recognition: lesion
634 symptom mapping in left hemisphere stroke. Brain 133:3269–3280.

635 Kiefer M, Pulvermüller F (2012) Conceptual representations in mind and brain: theoretical
636 developments, current evidence and future directions. Cortex 48:805–825.

637 Kilner JM (2011) More than one pathway to action understanding. Trends Cogn Sci 15:352–357.

638 Kilner JM, Kraskov A, Lemon RN (2014) Do monkey F5 mirror neurons show changes in firing rate
639 during repeated observation of natural actions ? J Neurophysiol 111:1214–1226.

640 Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. PNAS
641 103:3863–3868.

642 Leshinskaya A, Caramazza A (2015) Abstract categories of functions in anterior parietal lobe.
643 Neuropsychologia:1–13.

644 Lingnau A, Downing PE (Bangor U/ S of P (2015) The lateral occipitotemporal cortex in action.
645 Trends Cogn Sci.

646 Lingnau A, Petris S (2013) Action understanding within and outside the motor system: the role of
647 task difficulty. Cereb Cortex 23:1342–1350.

648 Mahon BZ, Caramazza A (2008) A critical look at the embodied cognition hypothesis and a new
649 proposal for grounding conceptual content. J Physiol Paris 102:59–70.

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

652 Mur M, Bandettini P a, Kriegeskorte N (2009) Revealing representational content with pattern-
653 information fMRI--an introductory guide. Soc Cogn Affect Neurosci 4:101–109.

654 Negri GAL, Rumiati RI, Zadini A, Ukmar M, Mahon BZ, Caramazza A (2007) What is the role of
655 motor simulation in action and object recognition? Evidence from apraxia. Cogn
656 Neuropsychol 24:795–816.

657 Nelissen K, Borra E, Gerbella M, Rozzi S, Luppino G, Vanduffel W, Rizzolatti G, Orban G a (2011)
658 Action observation circuits in the macaque monkey cortex. J Neurosci 31:3743–3756.

659 Nolte G (2003) The magnetic lead field theorem in the quasi-static approximation and its use for

660 magnetoencephalography forward calculation in realistic volume conductors. Phys Med Biol
661 48:3637–3652.

662 Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: Open source software for advanced
663 analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci
664 2011:156869.

665 Oosterhof NN, Connolly AC, Haxby J V. (in preparation) CoSMoMVPA: multi-modal multivariate
666 pattern analysis of neuroimaging data in Matlab / GNU Octave. Toolbox available from
667 http://cosmomvpa.org.

668 Oosterhof NN, Tipper SP, Downing PE (2012a) Viewpoint (in)dependence of action
669 representations: an MVPA study. J Cogn Neurosci 24:975–989.

670 Oosterhof NN, Tipper SP, Downing PE (2012b) Visuo-motor imagery of specific manual actions: a
671 multi-variate pattern analysis fMRI study. Neuroimage 63:262–271.

672 Oosterhof NN, Tipper SP, Downing PE (2013) Crossmodal and action-specific: neuroimaging the
673 human mirror neuron system. Trends Cogn Sci 17:311–318.

674 Oosterhof NN, Wiggett AJ, Diedrichsen J, Tipper SP, Downing PE (2010) Surface-based information
675 mapping reveals crossmodal vision-action representations in human parietal and
676 occipitotemporal cortex. J Neurophysiol 104:1077–1089.

677 Papeo L, Lingnau A, Agosta S, Pascual-Leone A, Battelli L, Caramazza A (2014) The Origin of Word-
678 related Motor Activity. Cereb Cortex:1–8.

679 Pavlidou A, Schnitzler A, Lange J (2014a) Interactions between visual and motor areas during the
680 recognition of plausible actions as revealed by magnetoencephalography. Hum Brain Mapp
681 35:581–592.

682 Pavlidou A, Schnitzler A, Lange J (2014b) Distinct spatio-temporal profiles of beta-oscillations
683 within visual and sensorimotor areas during action recognition as revealed by MEG. Cortex
684 54:106–116.

685 Pazzaglia M, Smania N, Corato E, Aglioti SM (2008) Neural underpinnings of gesture discrimination
686 in patients with limb apraxia. J Neurosci 28:3030–3041.

687 Pereira F, Mitchell TM, Botvinick MM (2009) Machine learning classifiers and fMRI: a tutorial
688 overview. Neuroimage 45:S199–S209.

689 Pfurtscheller G, Lopes da Silva F (1999) Event-related EEG/MEG synchronization and
690 desynchronization: basic principles. Clin Neurophysiol 110:1842–1857.

691 Pineda J a (2005) The functional significance of mu rhythms: translating “seeing” and “hearing”
692 into “doing”. Brain Res Brain Res Rev 50:57–68.

693 Pulvermüller F (2005) Brain mechanisms linking language and action. Nat Rev Neurosci 6:576–582.

694 Rizzolatti G, Cattaneo L, Fabbri-Destro M, Rozzi S (2014) Cortical mechanisms underlying the

695 organization of goal-directed actions and mirror neuron-based action understanding. Physiol
696 Rev 94:655–706.

697 Rizzolatti G, Fogassi L, Gallese V (2001) Neurophysiological mechanisms underlying the
698 understanding and imitation of action. Nat Rev Neurosci 2:661–670.

699 Schwarzbach J (2011) A simple framework (ASF) for behavioral and neuroimaging experiments
700 based on the psychophysics toolbox for MATLAB. Behav Res Methods 43:1194–1201.

701 Tarhan LY, Watson CE, Buxbaum LJ (2015) Shared and Distinct Neuroanatomic Regions Critical for
702 Tool-related Action Production and Recognition: Evidence from 131 Left-hemisphere Stroke
703 Patients. J Cogn Neurosci:1–21.

704 Turken AU, Dronkers NF (2011) The neural architecture of the language comprehension network:
705 converging evidence from lesion and connectivity analyses. Front Syst Neurosci 5:1.

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

709 Urgesi C, Candidi M, Avenanti A (2014) Neuroanatomical substrates of action perception and
710 understanding: an anatomic likelihood estimation meta-analysis of lesion-symptom mapping
711 studies in brain injured patients. Front Hum Neurosci 8:344.

712 Valyear KF, Culham JC (2010) Observing learned object-specific functional grasps preferentially
713 activates the ventral stream. J Cogn Neurosci 22:970–984.

714 Van Essen DC (2005) A Population-Average, Landmark- and Surface-based (PALS) atlas of human
715 cerebral cortex. Neuroimage 28:635–662.

716 Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH (2001) An Integrated
717 Software Suite for Surface-based Analyses of Cerebral Cortex. J Am Med Informatics Assoc
718 8:443–459.

719 Wurm MF, Ariani G, Greenlee MW, Lingnau A (2015) Decoding Concrete and Abstract Action
720 Representations During Explicit and Implicit Conceptual Processing. Cereb Cortex:bhv169.

721 Wurm MF, Lingnau A (2015) Decoding Actions at Different Levels of Abstraction. J Neurosci
722 35:7727–7735.

723

Figure Legends

Figure 1. Example of a trial sequence and experimental design. **A:** During MEG recording, N = 17 participants watched video clips of simple 'reach-to-point' or 'reach-to-grasp' movements (duration: 833 ms). Participants were instructed to fixate on a central fixation cross while attentively observing the entire video without performing any movements. To ensure that participants paid attention to the videos, different types of questions were asked during occasional catch trials that were later discarded from the analysis (for details, see Material and Methods). The green fixation cross indicated the period during which participants were allowed to blink. Eye movements were recorded using an MEG-compatible eye-tracker. **B:** We used a 2x2x2 design, manipulating the type of movement (pointing/ grasping), reach direction (left/ right), and effector (left/right hand).

Figure 2. Feature selection. Schematic representation of the method we adopted for selecting the features used for the multivariate analysis. Here we show one specific step of the algorithm with the selected central sensor (black dotted circle) with one neighboring sensor only (gray dotted circle) for illustrative purpose. Panel A shows the time-frequency representations (colors indicate power intensity) in the posterior sensors of the MEG helmet in two conditions of interest (condition A and B). The arrows starting from the circles indicate the corresponding magnified sensors. Panel B shows enlarged views of the two example sensors for condition A and B. The dotted rectangles illustrate an example time-frequency bin (2 neighboring bins per side for the time dimension; 4 neighboring bins per side for the frequency dimension; see the Methods section for details). For feature selection, for each time-frequency bin, we scanned each individual sensor with its 10 neighboring sensors. Panel B shows a matrix representation of the specific sensor/frequency/time bins. We then rearranged the dimensions of the matrix from 3D to 1D to obtain the corresponding feature vectors for condition A and B (Panel C). The feature vectors were used as input for the decoding analysis over sensors, frequency, and time. Specifically, the feature vectors were partitioned in independent chunks and used for training and testing the classifier. In the depicted example, each feature within the matrices was assigned with a number to show the same feature within the feature vectors for visualization purposes.

Figure 3. Behavioral results. Behavioral performance (% correct) for categorizing the two observed movements (grasping, pointing) as a function of video duration, collapsed across

effector and reach direction. As expected, participants responded more accurately with increasing video duration. Statistical analysis confirmed that participants reached above chance performance in classifying the two movements from 233ms onwards (for details, see Material and Methods, Statistical Analysis: Behavioral Experiment). Each dot represents data from a single participant. The continuous line indicates the linear model that best fits the data.

Figure 4. Theta-, alpha- and beta-band activity during action observation and univariate contrast. ***A*****,** Time-frequency representation of the difference (expressed in *t* scores) between grasping and pointing (collapsed across effector and reach direction) for the sensor highlighted in the head model. The four dotted lines indicate the following events, from left to right: (1) video onset, (2) median movement onset, (3) approximate time at which the hand touches the object (around 550 ms), (4) video offset (833 ms). ***B*****,** Same as **A**, but those time-frequency bins that did not survive the permutation test with Monte Carlo and cluster-based method for multiple comparisons correction were set to zero. ***C*****, *D*****,** Topography representation of the two frequency-bands observed in ***B***. ***E,*** Power change during action observation relative to baseline (fixation cross) over a representative sensor. The power change was calculated as (*activation – baseline)/baseline*, such that *1* indicates 100% increase respect to baseline and *-1* indicates 100% decrease respect to baseline. The classical power decrease in alpha and beta bands following observed movement onset (at t=0s) is evident.

Figure 5. Results of the neural spatiotemporal decoding. To identify abstract action representations of the observed actions (e.g. observing "grasping" irrespective of whether it was performed with the left or the right hand), we trained the MVPA classifier to discriminate between pointing and grasping using one effector (e.g. the left hand) and one reach direction (e.g. towards the left), and tested the performance of the classifier using an independent data set, using pointing and grasping movements performed with the other hand towards the opposite reach direction. We decoded the observed movements over time bins, frequency bins and sensors using a time-frequency-channel searchlight analysis. ***A*****,** The lateral plots show the time-frequency representation of the decoding in sensors depicted in the inset topoplots. Reddish colors indicate higher classification. Sensors were selected on the basis of the highest decoding accuracy at the frequency of interest. The central inset shows the two clusters that survived the correction for multiple comparisons (cluster obtained at early time point: 200 to

600ms; cluster obtained at late time point: 600 to 1200ms). ***B,*** Topography of the decoding at 400 ms and low frequencies (6Hz and 8Hz; smoothing: 4Hz). ***C,*** Topography of the decoding at 900 ms and higher frequencies (10Hz and 18Hz; smoothing: 3Hz). ***D-E,*** Sources accounting for the decoding effect found at sensor level, thresholded to retain only those voxels with the 10% highest decoding accuracies. For sensor level analysis only, significant differences were computed using permutation analysis and Monte Carlo methods and results are cluster corrected for multiple comparisons. Maps were projected on the PALS altas (Van Essen, 2005), using Caret software (Van Essen et al., 2001).

Figure 6. Maximum accuracy within each region. Within each identified source, the voxel with the maximum mean accuracy was selected and plotted with individual accuracies (black dots). Left MTG: Middle Temporal Gyrus (MNI: -50 -64 12); Left SPL: Superior Parietal Lobule (MNI: -20 -56 48); Right PCG: Precentral Gyrus (MNI: 28 -6 28); Right IFG: Inferior Frontal Gyrus (MNI: 20 24 28). Refer also to Table 2.

Figure 7. Comparison between univariate and multivariate analyses. Comparison between univariate (top row) and multivariate (bottom row) analyses in two time windows ([200 – 600ms] and [600 – 1200ms]). The upper topoplots show the sensors that survived the permutation test when comparing grasping vs pointing (collapsing across effector and reach direction). The lower topoplots show the sensors that survived the permutation test when comparing the observed accuracy of the classifier to distinguish between pointing and grasping (generalizing across effector and reach direction) against chance level (50%). Multivariate analysis was more sensitive in detecting the subtle differences between the neural signals induced by observation of the two movement types in the earlier time window. All shown clusters are corrected for multiple comparisons ($p<0.05$).

Figure 8. Neural decoding over time. The topoplots show the dynamics of above-chance accuracy (expressed as *t* scores) of the classifier in discriminating observed grasping and pointing (generalizing across effector and reach direction) for specific frequency bands (theta: 5-7 Hz; low alpha: 7-9 Hz; alpha: 9-11 Hz; beta: 17-19 Hz). The earliest significant decoding occurs in the posterior part of the configuration helmet in the lower frequency bands.

Figure 9. Simulation analysis. Illustration how 'low' classification accuracy (53.46% for sensor data; 50% is chance level) can be highly significant, using normal distribution probability plots

of Monte Carlo simulated classification accuracy distribution (relative to chance, 50%). The simulation uses the same parameters as used in the study (17 participants, minimum after trial rejection 544 trials per participant, same cross-validation scheme as used in original data). Dependency across cross-validation folds was set to r=0.3289 (green crosses) to match the value observed in the original data; for comparison, also results are shown for the cases of no dependence (r=0.00; blue) and full dependence (r=1.00; orange). The maximum classification accuracy above chance as observed in the original data is indicated by a black line.

Tables

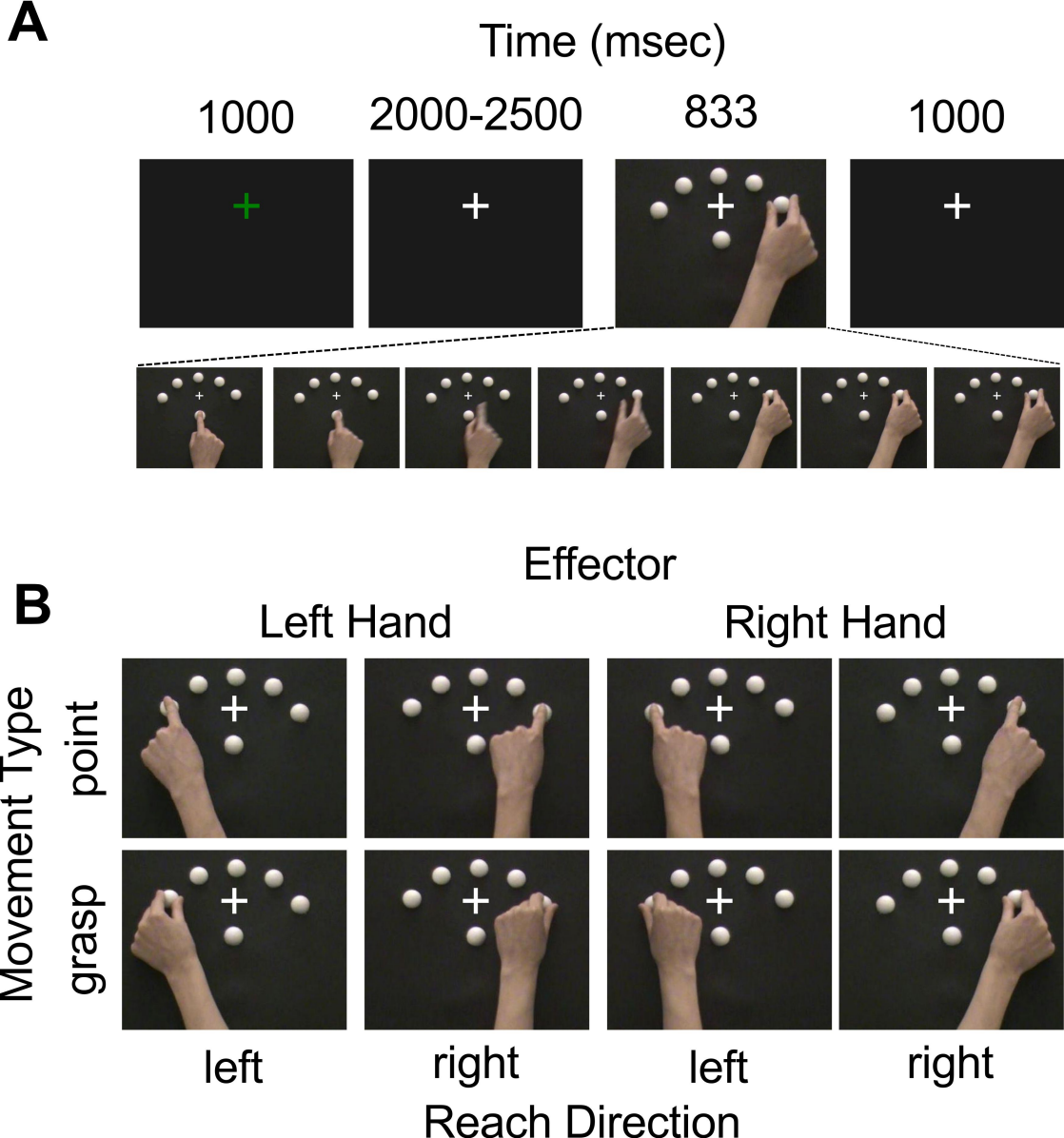
	Training: pointing vs grasping	*Testing: pointing vs grasping*
1	left hand, rightwards	right hand, leftwards
2	right hand, rightwards	left hand, leftwards
3	left hand, leftwards	right hand, rightwards
4	right hand, leftwards	left hand, rightwards

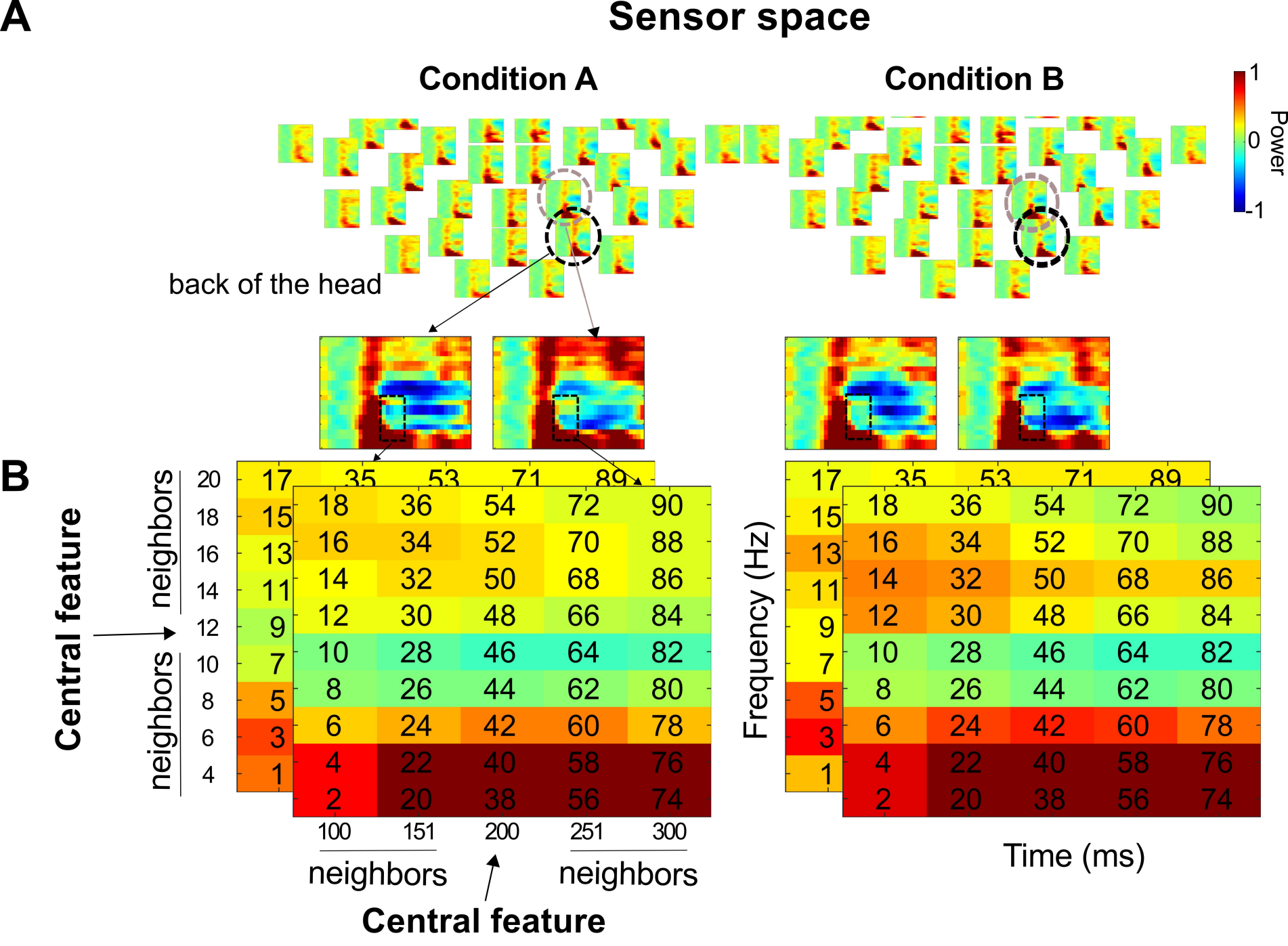
Table 1. Cross-comparisons used for training and testing. Classifiers were trained and tested in the following cross-comparisons: (1) training: distinguish between observed grasping and pointing actions performed with the left hand towards the left; testing: distinguish between observed grasping and pointing actions performed with the right hand towards the right; (2) training: distinguish between observed grasping and pointing actions performed with the right hand towards the right; testing: distinguish between observed grasping and pointing actions performed with the left hand towards the left; (3) training: distinguish between observed grasping and pointing actions performed with the right hand towards the left; testing: distinguish between observed grasping and pointing actions performed with the left hand towards the right; (4) training: distinguish between observed grasping and pointing actions performed with the left hand towards the right; testing: distinguish between observed grasping and pointing actions performed with the right hand towards the left. In this way, the classifiers could use information related to the type of movement only. The four accuracies determined with 1), 2), 3), and 4) were then averaged.

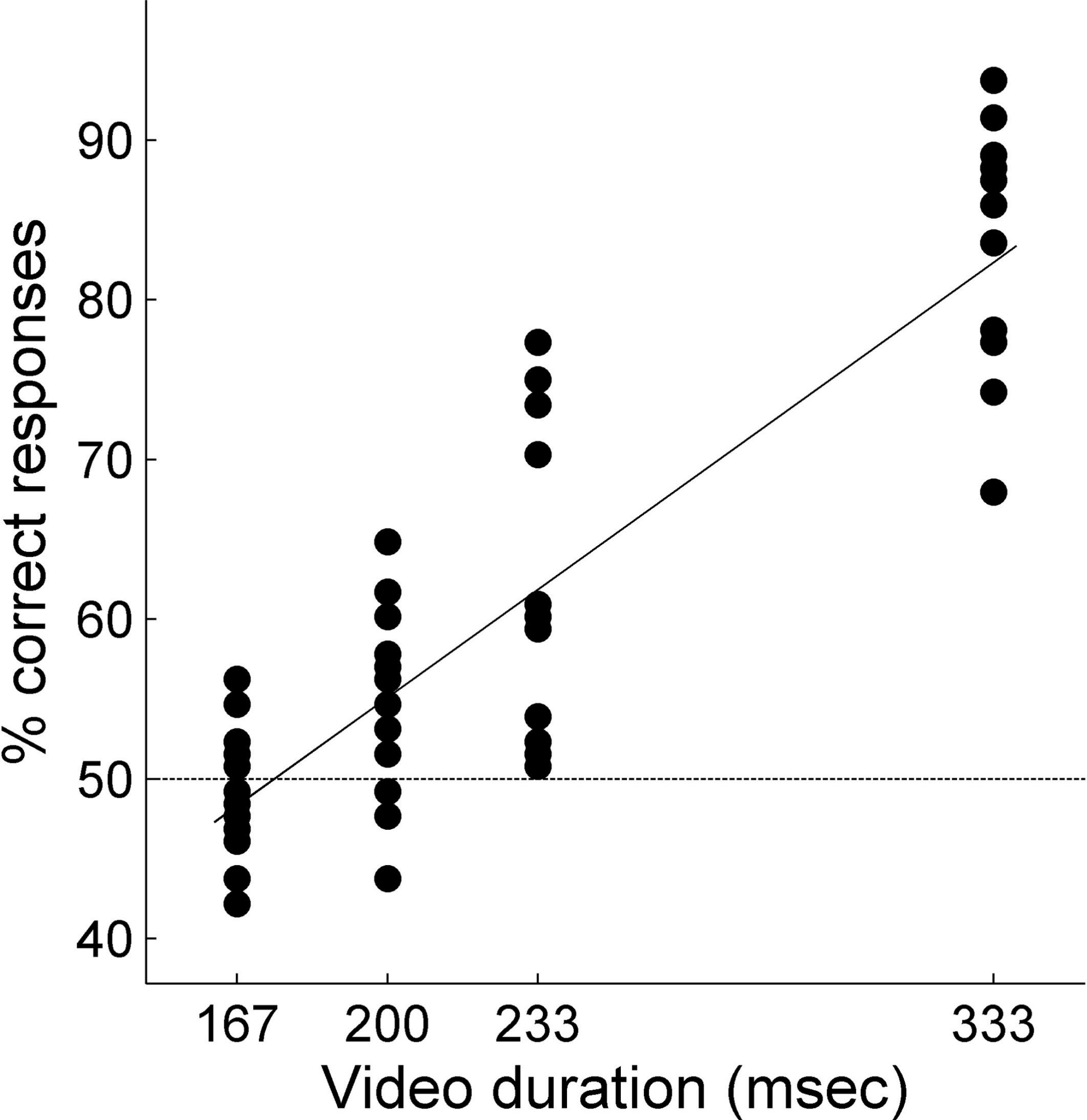
840

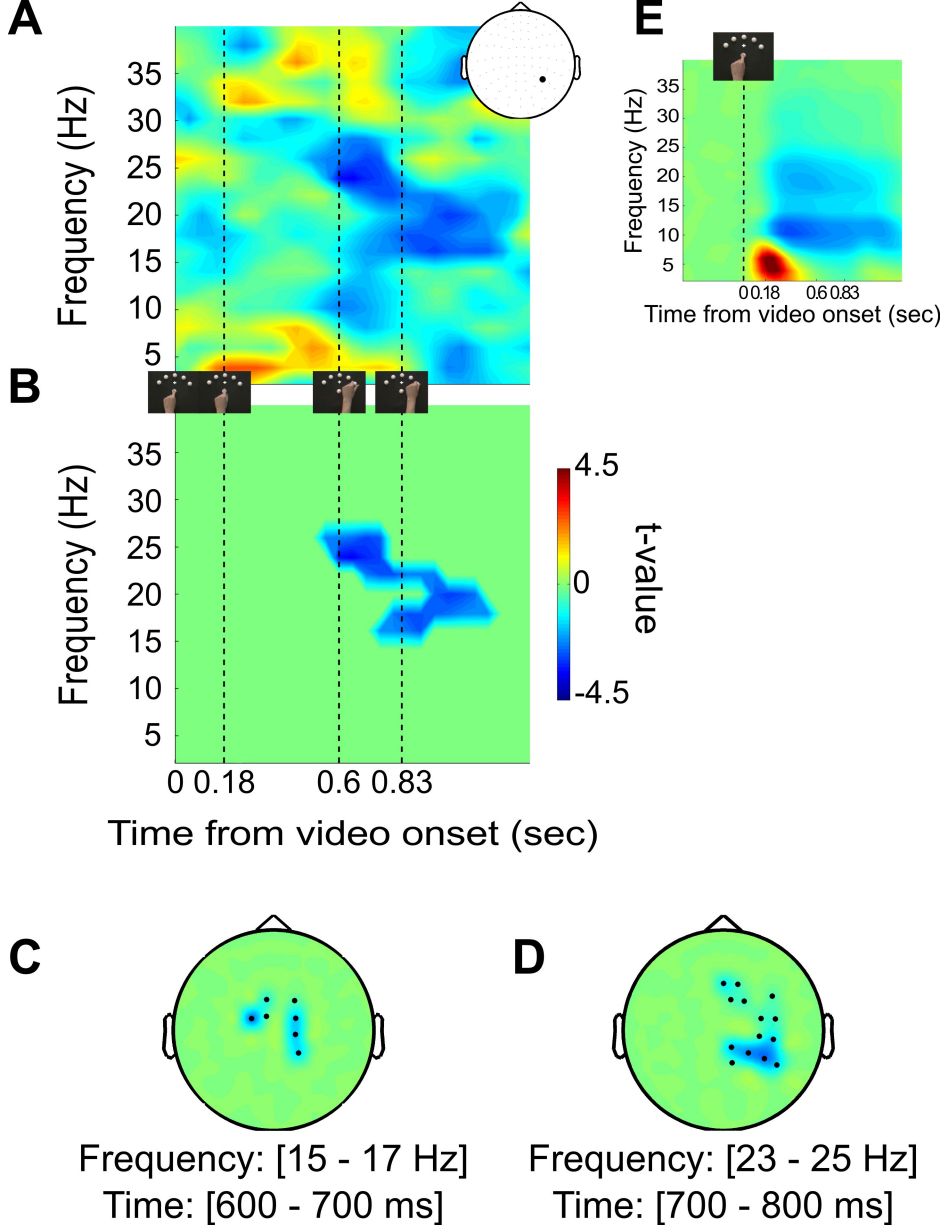
Area	*Frequency (Hz)*	*Time (msec)*	*X*	*Y*	*Z*
Left pITG	6	200	-52	-56	-12
Left SPL	6	400	-20	-56	48
Left pMTG	8	400	-50	-64	12
Right PCG	10	900	28	-6	28
Right IFG	18	900	20	24	28

841 **Table 2. MNI coordinates of the sources.** MNI coordinates of the sources (clusters) found in the
842 two different frequency bands, with the respective labels taken from the Anatomical Automatic
843 Labeling (AAL) database (Tzourio-Mazoyer et al., 2002). **pITG**: posterior portion of the Inferior
844 Temporal Gyrus; **pMTG:** posterior portion of the Middle Temporal Gyrus; **SPL:** Superior Parietal
845 Lobule; **IFG**: Inferior Frontal Gyrus (triangular part); **PCG:** Precentral Gyrus.

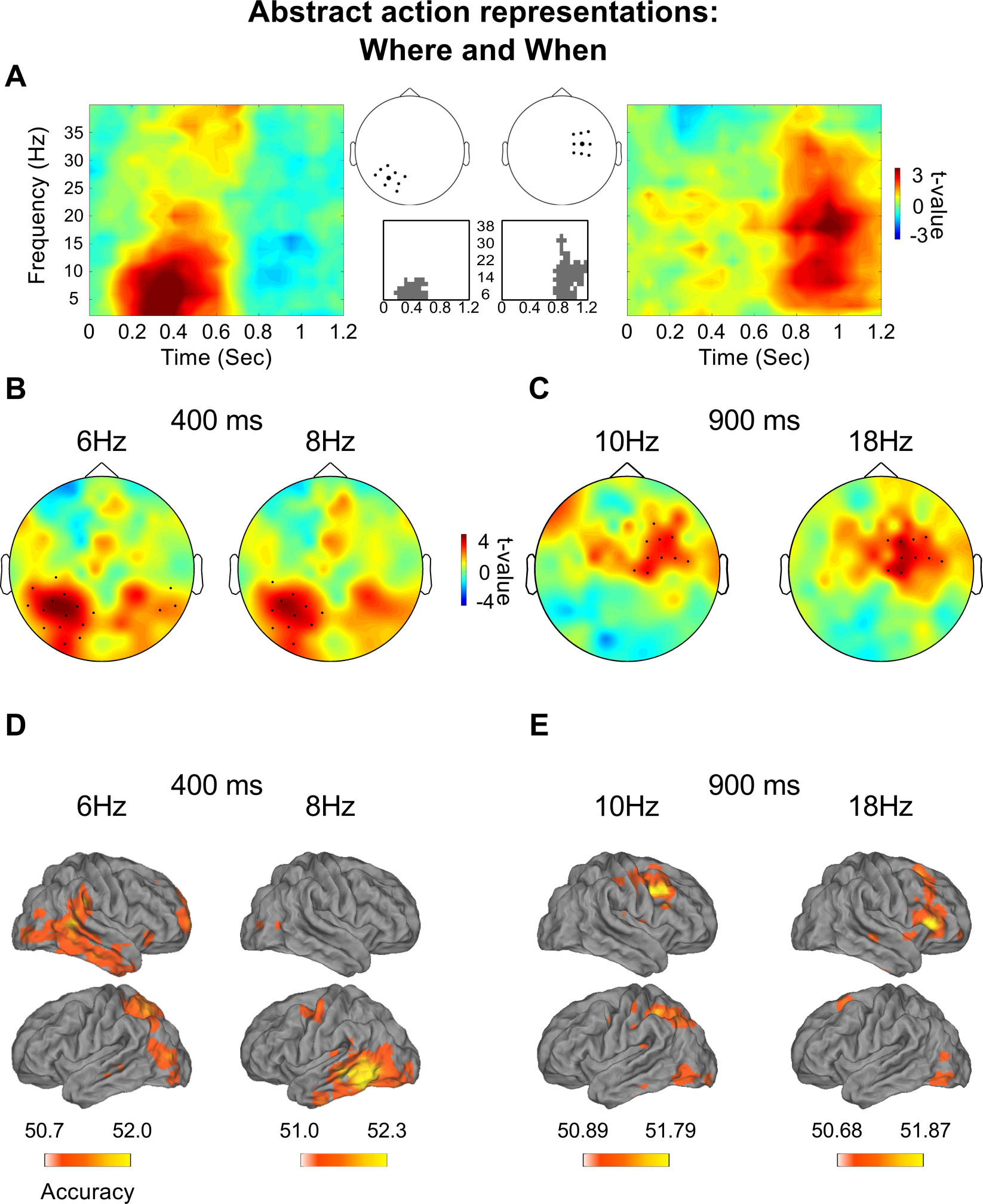

A
Time (msec)
1000
2000-2500
833
1000
B
Effector
Left Hand
Right Hand
Movement Type
point
grasp
left
right
left
right
Reach Direction

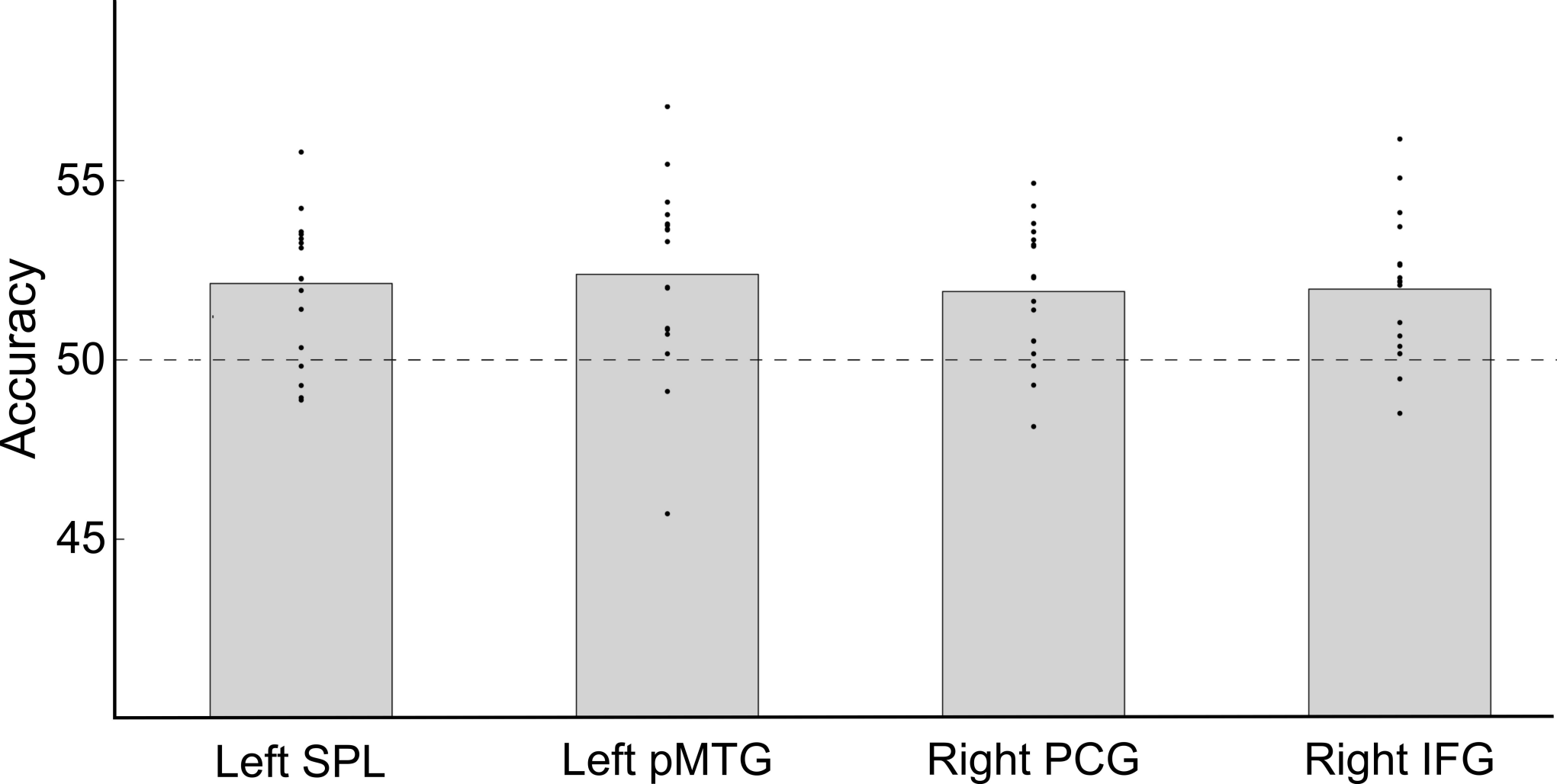

A
Sensor space
Condition A
Condition B
1
0
-1
Power
back of the head
B
Central feature
neighbors
neighbors
20 18 16 14 12 10 8 6 4
17 15 13 11 9 7 5 3 1
17 35 53 71 89
18 36 54 72 90
16 34 52 70 88
14 32 50 68 86
12 30 48 66 84
10 28 46 64 82
8 26 44 62 80
6 24 42 60 78
4 22 40 58 76
2 20 38 56 74
100 151 200 251 300
neighbors
neighbors
Central feature
Frequency (Hz)
Time (ms)
C
Feature vectors
Condition A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ⋯ 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85
Condition B
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ⋯ 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85


90
80
70
60
50
40
% correct responses
167
200
233
333
Video duration (msec)


A
Frequency (Hz)
35
30
25
20
15
10
5
B
Frequency (Hz)
35
30
25
20
15
10
5
4.5
0
-4.5
t-value
0 0.18
0.6 0.83
Time from video onset (sec)
E
Frequency (Hz)
35
30
25
20
15
10
5
0 0.18
0.6 0.83
Time from video onset (sec)
C
Frequency: [15 - 17 Hz]
Time: [600 - 700 ms]
D
Frequency: [23 - 25 Hz]
Time: [700 - 800 ms]

Abstract action representations: Where and When

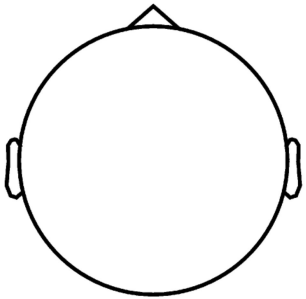


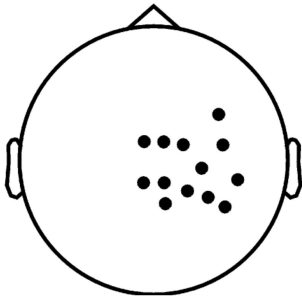

Accuracy
55
50
45
Left SPL
Left pMTG
Right PCG
Right IFG

1 - 10 Hz
400 ms

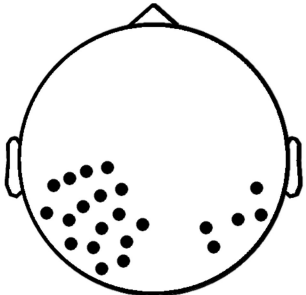
10 - 20 Hz
900 ms

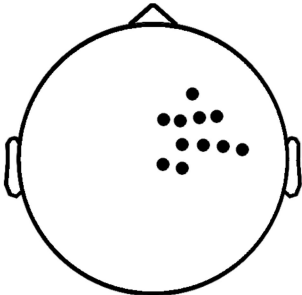
Univariate Analysis
Grasping vs Pointing,
collapsing across
effector and reach direction

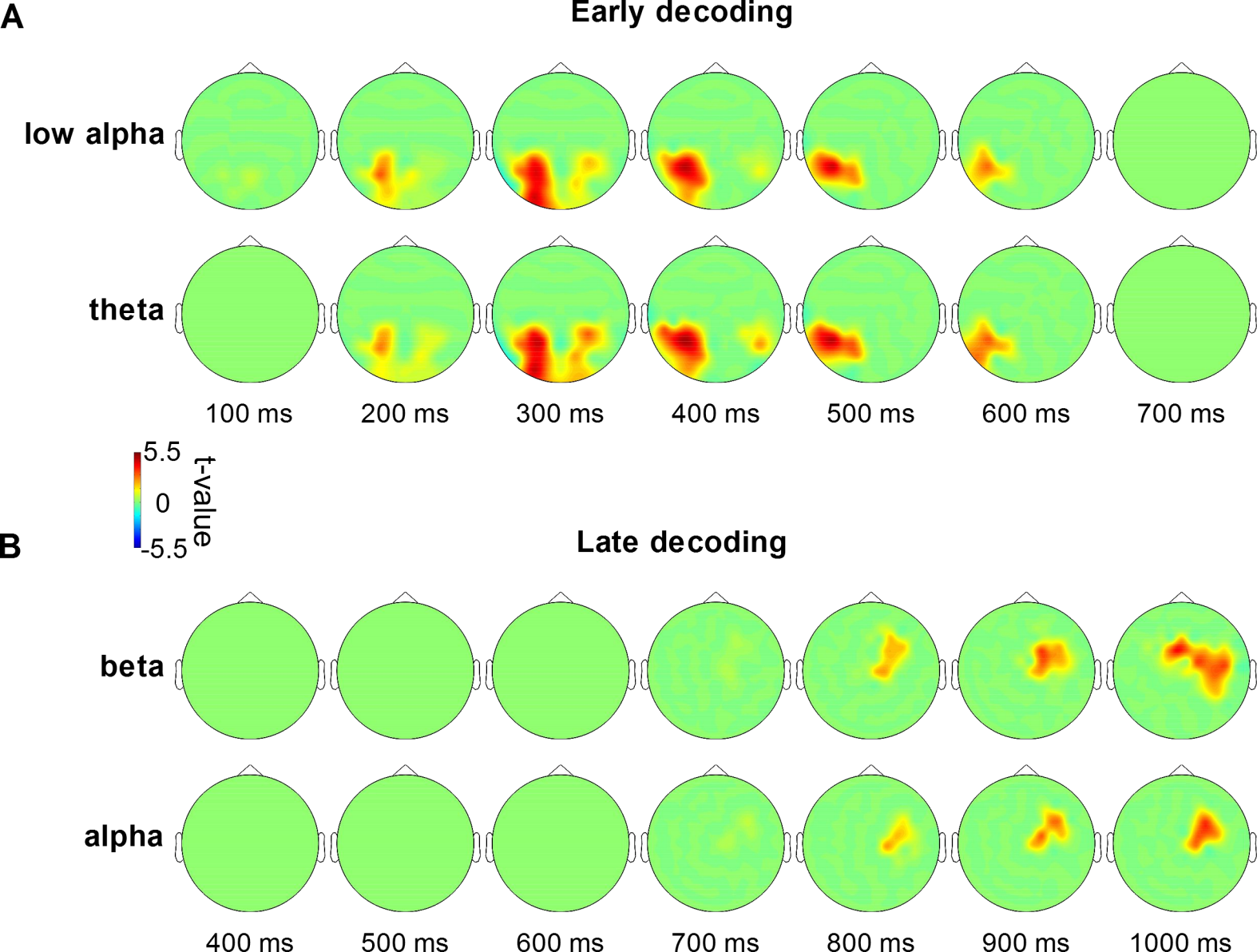


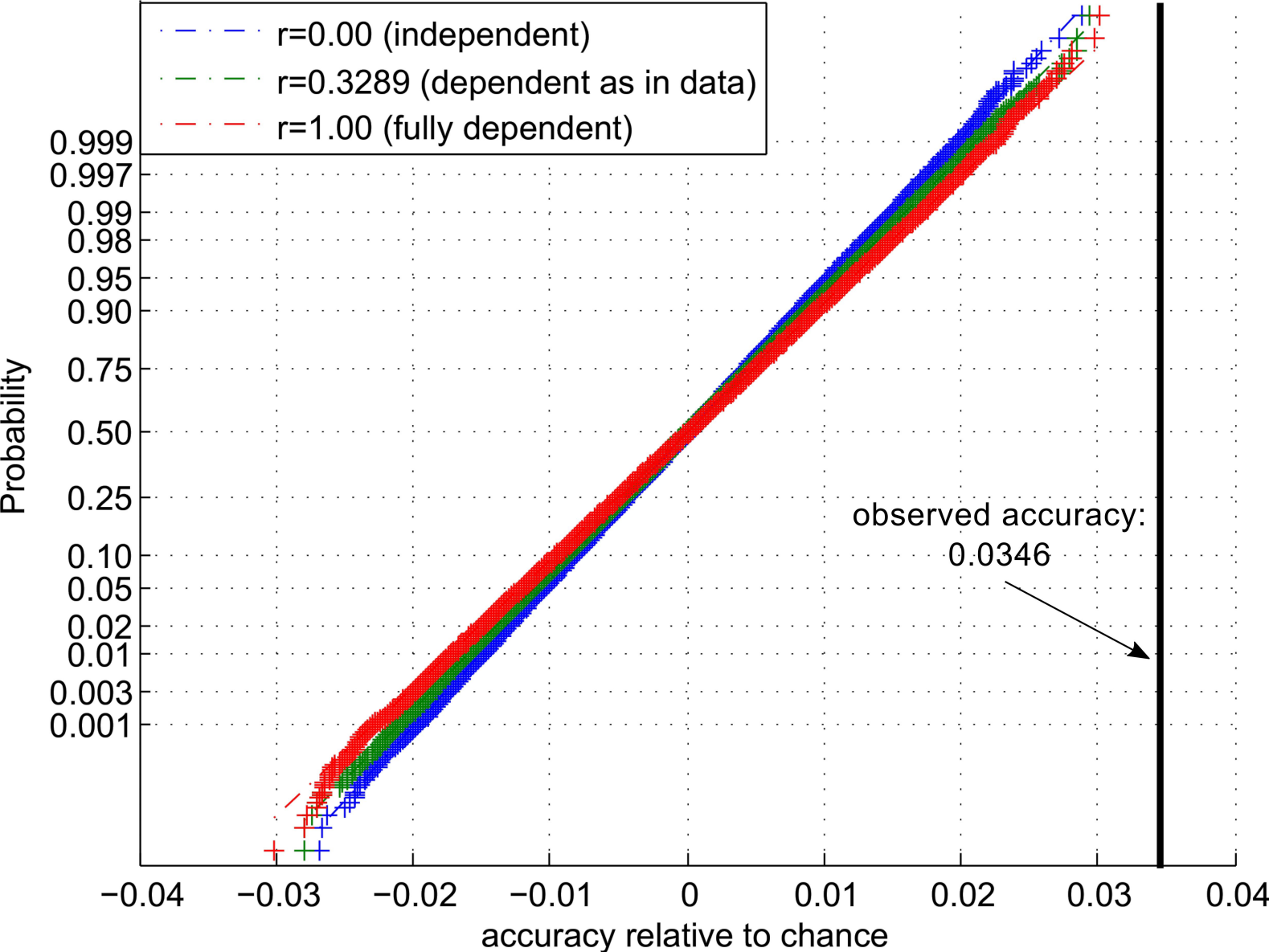


Multivariate Analysis
Grasping vs Pointing,
generalizing across
effector and reach direction






A
Early decoding
low alpha
theta
100 ms
200 ms
300 ms
400 ms
500 ms
600 ms
700 ms
5.5
0
-5.5
t-value
B
Late decoding
beta
alpha
400 ms
500 ms
600 ms
700 ms
800 ms
900 ms
1000 ms


r=0.00 (independent)
r=0.3289 (dependent as in data)
r=1.00 (fully dependent)
0.999
0.997
0.99
0.98
0.95
0.90
0.75
0.50
0.25
0.10
0.05
0.02
0.01
0.003
0.001
Probability
observed accuracy:
0.0346
−0.04
−0.03
−0.02
−0.01
0
0.01
0.02
0.03
0.04
accuracy relative to chance