Distinct dynamical patterns that distinguish willed and forced actions

Luis Garcia Dominguez · Wojciech Kostelecki · Richard Wennberg · Jose L. Perez Velazquez

Abstract The neural pathways for generating willed actions have been increasingly investigated since the famous pioneering work by Benjamin Libet on the nature of free will. To better understand what differentiates the brain states underlying willed and forced behaviours, we performed a study of chosen and forced actions over a binary choice scenario. Magnetoencephalography recordings were obtained from six subjects during a simple task in which the subject presses a button with the left or right finger in response to a cue that either (1) specifies the finger with which the button should be pressed or (2) instructs the subject to press a button with a finger of their own choosing. Three independent analyses were performed to investigate the dynamical patterns of neural activity supporting willed and forced behaviours during the preparatory period preceding a button press. Each analysis offered similar findings in the temporal and spatial domains and in particular, a high accuracy in the classification of single trials was obtained around 200 ms after cue presentation with an overall average of 82%. During this period, the majority of the discriminatory power comes from differential neural processes observed bilaterally in the parietal lobes, as well as some differences in occipital and temporal lobes, suggesting a contribution of these regions to willed and forced behaviours.

Keywords Magnetoencephalography · Single trial analysis · Free will · Fischer criterion · Common spatial patterns

L. Garcia Dominguez (✉) · W. Kostelecki
Neuroscience and Mental Health Programme,
Brain and Behaviour Centre, Division of Neurology,
Hospital for Sick Children, Neurology, Room 6535 Hill Wing,
555 University Avenue, Toronto, ON M5G1X8, Canada
e-mail: l.garcia.d@gmail.com

R. Wennberg
Division of Neurology, Toronto Western Hospital, Toronto,
ON, Canada

J. L. Perez Velazquez
Department of Paediatrics and Institute of Medical Science,
University of Toronto, Toronto M5G1X8, Canada

Introduction

The experience of conscious free will is universal amongst people. By drawing causal inferences between thoughts and actions, ownership is attributed to one's actions. A fundamental question in the psychophysics of consciousness is how mental entities are derived from neurophysiological processes (Damasio 1999). In general, defining free will is quite complicated and its investigation shares similar conceptual ambiguities. However, as expressed by J. D. Schall: "If we ask whether we are free, the kind of answer we want may not be possible. A better question to ask is: do we make choices? The answer is certainly yes […] Are our choices constrained? Yes" (Schall 2001). While there may be uncertainties and ambiguities regarding concepts of free will, nobody can deny the validity of the questions posed above and how the process of making choices relates to an understanding of an individual's free will.

Since the celebrated and much debated experiments of Benjamin Libet and colleagues, which showed a slow negativity in the electroencephalogram (EEG) preceding a self-initiated movement (Libet 1985), the idea of measuring brain signals that may indicate mechanisms in the performance of willed actions has been a subject of debate. This seminal study indicated that the slow negativity or "Libet potential" precedes not only the willed motor act

but also the awareness of deciding to perform the act. This suggests that some part of an individual's brain already "knows" about the execution of a motor act before that information is consciously accessible and this observation has motivated increased interest in the study of free will and its physiological basis. Recent primate studies have suggested that the interaction between frontal and parietal cortices is fundamental for free choice and decision making (Pesaran et al. 2008; Soon et al. 2008). Parietal cortex has been implicated in movement intention and awareness (Desmurget and Sirigu 2009), and frontal cortices have been thought to play an important role in decision making and willful acts (Frith et al. 1991; Ingvar 1994; Manes et al. 2002; Pesaran et al. 2008; Rushworth 2008). Auditory decision-making was found to be associated with enhanced gamma band activity using magnetoencephalography (MEG) recordings in humans (Kaiser et al. 2008). Neuroimaging studies indicated an increase in blood flow in dorsolateral prefrontal cortex associated with willed acts (Frith et al. 1991). These observations suggest a chain of cortical neural circuits involved in the free generation of a motor act—it is now believed that decisions are made, unconsciously, in frontal regions and stored until executed some time later. The stored decision in the frontal cortex is sent to parietal regions where it is made consciously accessible (Haggard 2008). This decision in frontal cortex also initiates the motor component of movement through the supplementary motor area. However, a clear distinction between the brain activity recorded preceding forced and willed acts is needed to better understand the functional dynamics of cortical activity that determines willed actions. To address these issues, this study uses a chosen (free) versus forced action paradigm, in which a subject either (1) performs a task when given specific instruction to do so or (2) freely chooses to perform that same action from other possible actions. The analysis performed is exploratory in nature and it is guided by a simple hypothesis; brain recordings preceding forced and free actions can be distinguished from one another due to distinct spatiotemporal cortical dynamics supporting each behaviour. We can also inquire further: In what sense are these dynamics unique? In these regards we considered a direct and general approach to data analysis using no

specific model of brain function. We forwent traditional tools like source localization, dipole analysis and synchronization analysis, which presume a specific physical scenario for uncovering neural mechanisms, and addressed the analysis in a less constrained setting focused on direct differences in spatiotemporal patterns at the sensor level.

Materials and methods

MEG recordings

MEG data were obtained from a whole-head 151-channel CTF System (Port Coquitlam, Canada) at a sampling rate of 625 Hz. The experiment was undertaken with the understanding and written consent of each subject to the protocol approved by the Hospital for Sick Children Review Ethics Board. Recordings were performed with subjects in a supine position and looking at a screen positioned approximately 1 m in front of their face. Subjects also had one button placed on each side of their body and were asked to be ready to press these buttons with the index finger of either hand in response to visual cues presented on the screen. They were also reminded to remain as still as possible for the duration of the study. For each trial, a symbol defining an instruction was presented from a set of three different symbols (Fig. 1a–c) which include the symbol that instructs the subject to press the right button (Fig. 1a), the left button (Fig. 1b) or to select and press one of either the right or left buttons (Fig. 1c). After a button press, a fourth symbol appears (Fig. 1d) instructing the subject to press both buttons at the same time to initiate the next trial after a fixed delay of 0.5 s (symbol D is also the starting symbol for each recording session). By having the subject press both buttons at the same time we expect to reduce any bias in subsequent trials that is introduced by the last button pressed. Symbols A–C are presented in a random sequence with the frequency of presentation per symbol set to $F_A = ¼$, $F_B = ¼$ and $F_C = ½$. This way, the number of times the subject faces the free choice is roughly the same as the number of times facing the forced condition. Seven right-handed participants (1 female) were tested, with subject ages ranging from 23 to 49 with a median

Fig. 1 *Symbols*/instructions used in button pressing task. *Symbols* instruct the subject to press the *right button* (**a**), *left button* (**b**), choose the button to press (**c**), and press both buttons to initiate the next trial (**d**)
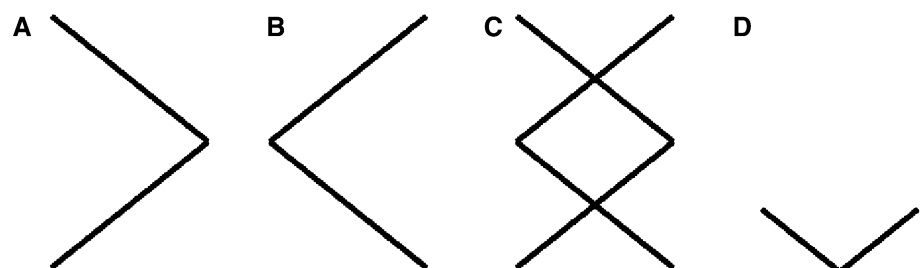
age of 27. Recordings from one male subject were discarded due to poor head positioning in the MEG apparatus and substantial movement artifacts in the recordings. Subjects were instructed to respond to the free choice symbol with their chosen button but to also keep responses random. Free responses had no detectable patterns and were evenly distributed between left and right button presses for five of the six subjects. Only one subject had a right hand bias (66–34%) but this appeared to not have significant impact on the results. Subjects completed the experiment at their own pace in three epochs of 5 min each with 2–5 min of relaxation between epochs. The average number of trials in each condition across subjects, after deleting erroneous responses, was 265 for $F_A + F_B$ and 268 for F_C.

Data analysis

For each subject, trials with incorrect responses were discarded although no additional rejection of trials based on artifacts was performed. All trials belonging to the forced condition (left and right) were grouped together (group **X**) and a second group was formed containing all the trials (both left and right responses) belonging to the free condition (group **Y**). Data for a single trial consists of an $N \times T$ matrix where N is the number of recording channels and T is the number of samples in time. The data was band-pass filtered using a finite impulse response (FIR) filter between 10 and 80 Hz. Other band-pass filters implemented in this study resulted in a similar performance as long as they were relatively wide (e.g. 3, 8, 10–50, 60, 80 Hz).

Three different approaches for analysis were performed on the data sets: the Fisher criterion, a correlation analysis over groups of trials and a linear classification of single trials.

1) Fisher criterion

The Fisher criterion (FC) was used to quantify differences between the mean MEG signal of trials from groups **X** and **Y** over time (Müller et al. 2004). The FC is defined in terms of the optimal separation of $\mu^{\mathbf{X}}(t)$ and $\mu^{\mathbf{Y}}(t)$ which are $N \times 1$ vectors with coefficients $\mu_i^{\mathbf{X}}(t)$ and $\mu_i^{\mathbf{Y}}(t)$ that denote the mean MEG signals from the population of trials in each condition at time t. These mean vectors have covariances $\Sigma^{\mathbf{X}}(t)$ and $\Sigma^{\mathbf{Y}}(t)$ with variances (coefficients in the diagonal) denoted by $\Sigma_i^{\mathbf{X}}(t)$ and $\Sigma_i^{\mathbf{Y}}(t)$, respectively. The separation between groups **X** and **Y**, in Fisher's sense, is defined as the ratio of the variance between the groups to the variance within the groups. For a single channel, i, this separation is expressed by the equation

$$F_i(t) = \frac{\left(\mu_i^{\mathbf{X}}(t) - \mu_i^{\mathbf{y}}(t)\right)^2}{\Sigma_i^{\mathbf{X}}(t) + \Sigma_i^{\mathbf{Y}}(t)}. \tag{1}$$

A given $N \times 1$ vector, w, applied to each trial produces a projection of the N-dimensional data into a 1-dimensional space (with mean $w^{\mathbf{T}}\mu$, and variance $w^{\mathbf{T}}\Sigma w$) where the FC, $F(w, t)$, is evaluated using the equation

$$F(w,t) = \frac{(w^{\mathbf{T}}(\mu^{\mathbf{X}}(t) - \mu^{\mathbf{Y}}(t)))^2}{w^{\mathbf{T}}(\Sigma^{\mathbf{X}}(t) - \Sigma^{\mathbf{Y}}(t))w} \tag{2}$$

where the optimal weight vector, $w*$, is obtained by the equation

$$
\begin{aligned}
w^* &= \arg \max_w F(w,t) \\
&= (\Sigma^{\mathbf{X}}(t) - \Sigma^{\mathbf{Y}}(t))^{-1}(\mu^{\mathbf{X}}(t) - \mu^{\mathbf{Y}}(t))
\end{aligned} \tag{3}
$$

such that $w*$ maximizes $F(w, t)$ with higher values of $F(w, t)$ indicating a greater difference in MEG signals between groups **X** and **Y**. The measure, $F(w, t)$, was applied over time using the statistic described in VanRullen and Thorpe 2001, where a departure from baseline is significant if at least 15 consecutive t-statistics obtain $P < 0.01$ when compared to a baseline that consist of all data from 160 ms before the cue to 50 ms after.

2) Correlation analysis of trials

A correlation analysis of trials comparing within- and between-group correlations was performed to determine spatial and temporal information about the involvement of cortical regions in the completion of forced and chosen behaviours. This analysis was repeated on a sliding window of 80 ms with a 60% overlap between adjacent windows. Seventeen windows were obtained this way with the first one starting 128 ms before the stimulus and the last one ending 465 ms after the stimulus.

For each window, correlations between trials from different conditions were evaluated and these correlations were averaged according to the equation,

$$r_i(\mathbf{X}, \mathbf{Y}) = \frac{1}{N_X N_Y} \sum_{j=1}^{N_X} \sum_{k=1}^{N_Y} \mathrm{corr}\left(X_i^{(j)}, Y_i^{(k)}\right) \tag{4}$$

where the subscript, i, denotes the channel being considered, the superscripts, (j) and (k), refer to the trial number, N_X and N_Y are the number of trials in **X** and **Y**, and $\mathrm{corr}\left(X_i^{(j)}, Y_i^{(k)}\right)$ refers to the correlation coefficient between the recordings of one channel during two different trials, $X_i^{(j)}$ and $Y_i^{(k)}$. This average is then compared with the one that is obtained from the same analysis over two other groups of the same size as the original formed with trials randomly selected from the whole set of trials ($\mathbf{X} \cup \mathbf{Y}$); we identify these new sets as $\tilde{\mathbf{X}}$

and $\tilde{\mathbf{Y}}$. If trials between conditions $\mathbf{X}$ and $\mathbf{Y}$ are significantly different from each other but fairly consistent within the same condition, the mean correlation that results from the randomized groups should be higher, on average, than the original correlation. That is, $r_i(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) > r_i(\mathbf{X}, \mathbf{Y})$ should hold for a majority of cases if there is a notable within-group correlation that does not exist between trials of groups $\mathbf{X}$ and $\mathbf{Y}$.

To test this hypothesis the distribution of $r_i(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$ is estimated with $N = 500$ random assignments of trials into $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ and this distribution is compared to the original averages, $r_i(\mathbf{X}, \mathbf{Y})$. A statistic, P, is calculated as the number of times the above inequality holds divided by the total number of randomized comparisons in the estimated distribution. If $r_i(\mathbf{X}, \mathbf{Y})$ is lower than 95% of the average correlations from the shuffled groups the resulting P is considered to be statistically significant ($\alpha = 0.05$). In other words, P_i is calculated using the equation

$$P_i = \frac{1}{\tilde{N}} \sum_{k=1}^{\tilde{N}} \theta\left(r_i^k(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) - r_i(\mathbf{X}, \mathbf{Y})\right) \quad (5)$$

where θ is a step function ($\theta(x) = 1$ if $x > 0$ and 0 otherwise) and $P_i > 1 - \alpha$ denotes significance. This analysis is computed for each channel, window, and subject separately. For a more technical explanation of the statistical methodology used, see (Golland and Fischl 2002).

3) Single trial classification

Classification of trials was performed using two different procedures of feature extraction on which subsequent Linear Discriminant Analysis (LDA) was applied.

The first was the Common Spatial Patterns (CSP) method (Müller-Gerking et al. 1999; Ramoser et al. 2000) which works by producing spatial filters (linear combination of channels) that optimally separate two populations of trials based on the variance of the filtered signals. Each spatial filter is a vector of coefficients (one per channel) that maximizes the variance in one condition while minimizing the variance of the signal in the other condition. More than one filter can be used producing multidimensional feature spaces for classification. In our case a total of six linear transformations—corresponding to a 6-dimensional feature space, or $m = 3$ in Ramoser et al. (2000)—were found to be enough for a good discrimination of conditions. With this method, the classification was performed over a single window starting 112 ms after the presentation of the stimulus and ending 176 ms later (central point located 200 ms after the cue)—some other locations in time were also attempted with inferior results in terms of classification accuracy. An illustration of what the CSP method accomplishes, specifically in this study, is provided in Fig. 2. Figure 2a depicts the average trial of groups $\mathbf{X}$ and $\mathbf{Y}$ for the channel where the greatest difference in variance was observed between the two conditions. Figure 2b and c depicts two new trials obtained after the CSP filter is applied when the method focuses on differentiating the window delimited by vertical dotted lines. In Fig. 2b, the filter maximizes the ratio of the variance of the free condition to the variance of the forced condition, and in Fig. 2c it does the opposite. In both cases the discrimination between the two conditions based on the variance of the trials improves over the original, unfiltered condition.

The second procedure applied for feature extraction was the already mentioned FC. As in the case of the correlation analysis the single trial classification using FC was performed on a sliding window of $T = 50$ (80 ms wide) with 90% overlap covering the period from 158 ms before the cue to 465 ms after. On each window the data was transformed by applying the optimal w^*-vector at each time point (Eq. 3), resulting in a one-dimensional time series for each trial. The time-average $F(w, t)$ of these time series was then taken as the only feature. An early attempt to reduce the dimensionality of this time series using again the FC did not show a better cross validation error than the simple average.

LDA was applied to single trials using—independently—the variances of the signals transformed by the selected CSP filters and the time-averaged projection of each trial using the optimal weight vectors obtained from the training sets. For estimation of the classification accuracy we used a repeated random sub-sampling validation approach; the data set for each subject was divided into a training set (80% of trials) and a testing set (20%). The training set was used to obtain the optimal CSP filters and Fisher projection vectors that were then applied to the testing set of trials. Once these linear combinations were applied on the testing set the resulting features obtained were fed into the LDA to subsequently evaluate the classification accuracy on the testing set. This procedure was repeated 20 times with different random partitions for the training and testing sets. The total classification accuracy that is reported is the mean classification accuracy from these 20 attempts. Certain channels were non-functional in some subjects and so the channels for which there was insufficient data (3 or fewer subjects) were discarded.

Results

The FC is the only measure in our study where trial populations can be compared instantaneously, at every point in time, using no temporal averages over running windows. This allows us to pinpoint instantaneous changes in the dynamics that differentiate neural activity arising from free
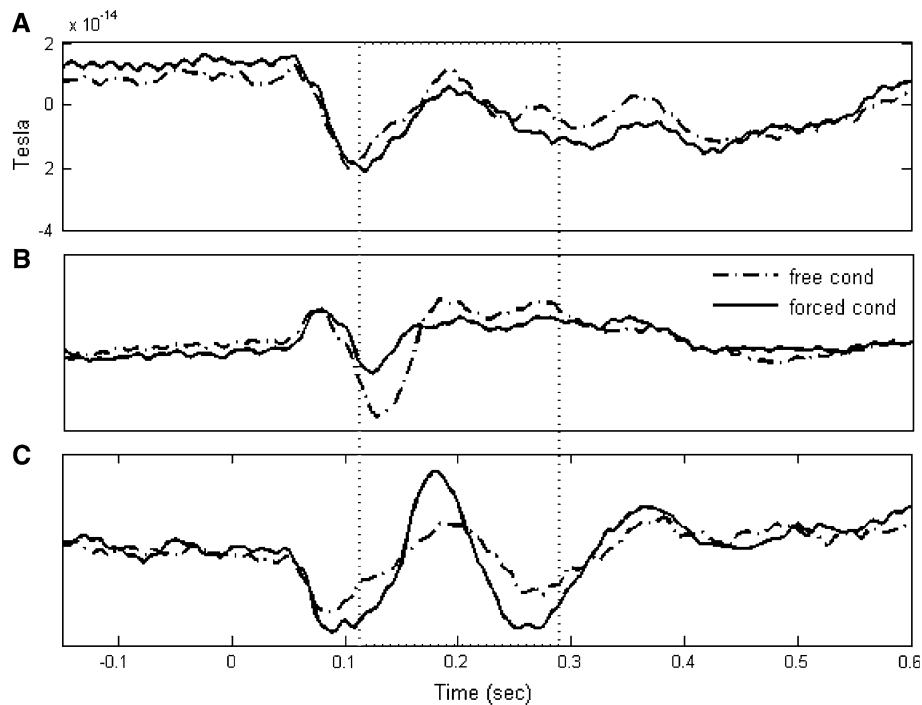
Fig. 2 a Average trials in one subject for a single channel in each condition. The channel (in the occipital region) was selected as the one that shows the greatest difference in variance between the two conditions. The cue was presented at 0 s. **b** and **c** New average trials resulting from the application of CSP filters (described in part 3 of the "Data analysis" section). The filter was optimized for the period of time between the two *dotted lines*. Two different projections (*first* and *last row* (filters) of the projection matrix, Ramoser et al. 2000) have been applied to each average trial. In **b** the filter (first projection) maximizes the separation in variances between conditions by increasing the variance of the free condition and decreasing the variance of the forced condition, in **c** (last projection) it increases the variance of the forced and decreases the variance of the free condition. Note that in **b** and **c**, trials are much more easily discriminated based on variance differences than they were originally in **a**

and forced behaviours. Figure 3 (bottom panel), shows the temporal course of $F(w, t)$ and this illustrates the general differences between the event related potentials (ERPs) associated with the free and forced conditions. Following the same criterion to that used in (VanRullen and Thorpe 2001; see "Data Analysis: Fisher's Criterion") it can be found that as early as 64 ms after the cue is presented (the arrow in Fig. 3, bottom) there is a significant jump in $F(w, t)$. The curve increases from here and peaks at approximately 200 ms. Four points are selected in this trace and labelled A, B, C and D at 96, 123, 200 and 326 ms, respectively, and for each point, a diagrammatic head of F_i values (Eq. 1) is plotted showing the specific spatial patterns of the channels that best separate the two conditions. The cortical activity that distinguishes free and forced behaviours is initiated over occipital channels (A) and progressively moves anteriorly towards parietal and temporal regions (B, C). At the moment of average reaction time (when the button is pressed) the activity on the left side may suggest motor and sensory feedback areas (D).

The number of channels showing statistical significance ($P > 0.95$) on the correlation analysis using shuffled datasets shows an unambiguous pattern over time that is depicted in Fig. 4 (top panel). The maximum proportion of channels with a significant P is reached at around 200 ms after the presentation of the cue (note that the button press, averaged across all subjects, occurs 320 ms after the cue). The percentages shown are the grand mean over the average number of channels with $P > 0.95$ for each subject. The dotted lines show the standard deviations of these six averages. These results indicate that a considerable number of channels show significant differences between the two conditions. Furthermore, this curve resembles the one obtained from the single trial analysis using the FC (Fig. 4, bottom panel); both peak at 0.2 s and show similar temporal patterns suggesting that the two methods are describing a common phenomenon. Better classification accuracies were achieved when window sizes were increased from 80 to 150 ms but at the expense of temporal discrimination. The overall (across subjects) classification accuracy at the peak was 82% and individual accuracies were 84.3, 83.5, 78.8, 78.8, 81.6 and 84.7%.

The spatial distribution of each channel's P_i averaged over subjects is shown in Fig. 5 and presented over seventeen time-windows. The elapsed time from the cue presentation to the central point of each window is displayed on the top of each head. The pattern observed suggests a spreading of cortical involvement from the occipital lobe to
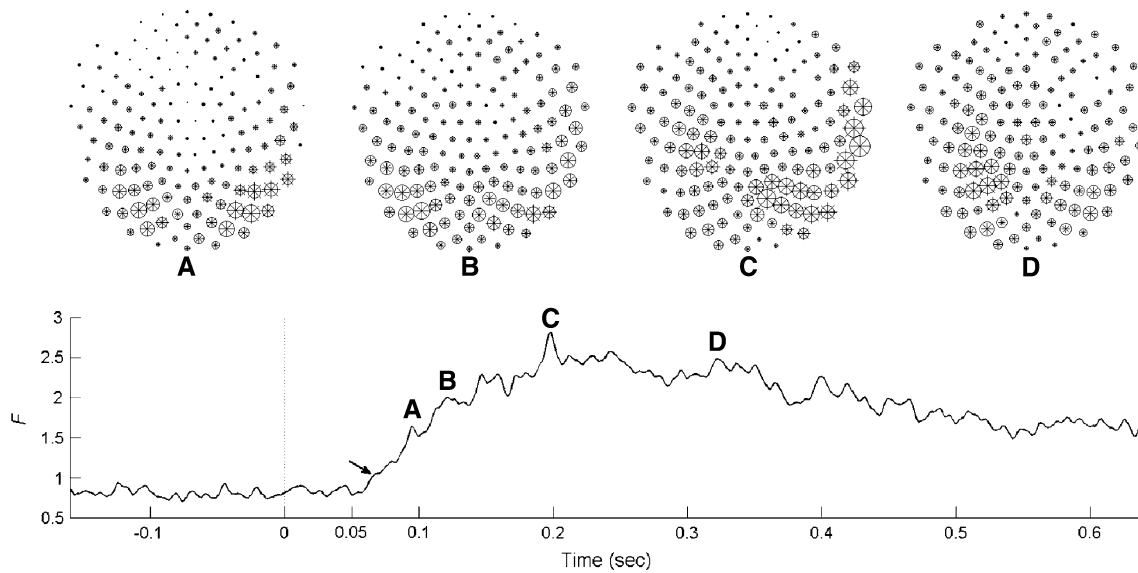
Fig. 3 *Bottom*: Fisher's parameter $F(w, t)$ averaged across subjects (described in part 1 of the "Data analysis" section). For each subject each trace is the result of applying the ratio of variances F (Eq. 2) to the best linear separation for each time point. Each subject is independently processed and the results are averaged. The times labelled **a**, **b**, **c** and **d** correspond to the diagrammatic heads in the *top panel*. *Top*: the size of each *circle* is a function of the Fisher parameter F_i for every sensor i. These are also averages across subjects. The *stars* inside the *circle* are the standard deviations of these averages

Fig. 4 *Top*: Grand mean of the percentage of channels with $P > 0.95$ (defined by Eq. 5) across subjects over 17 time windows. The cue is presented at 0 s. *Bottom*: Overall mean of classification accuracy from the single trial analysis using FC (described in part 1 and 3 of the "Data analysis" section). For both *panels*, *dotted lines* denote inter-subject standard deviations
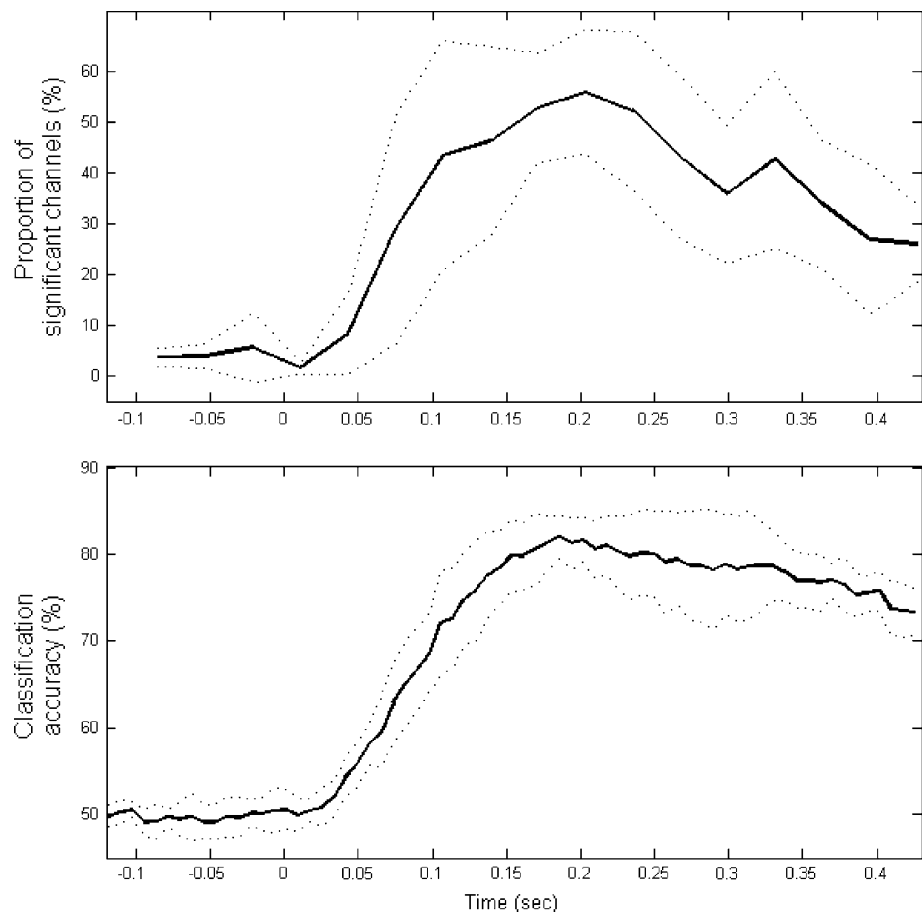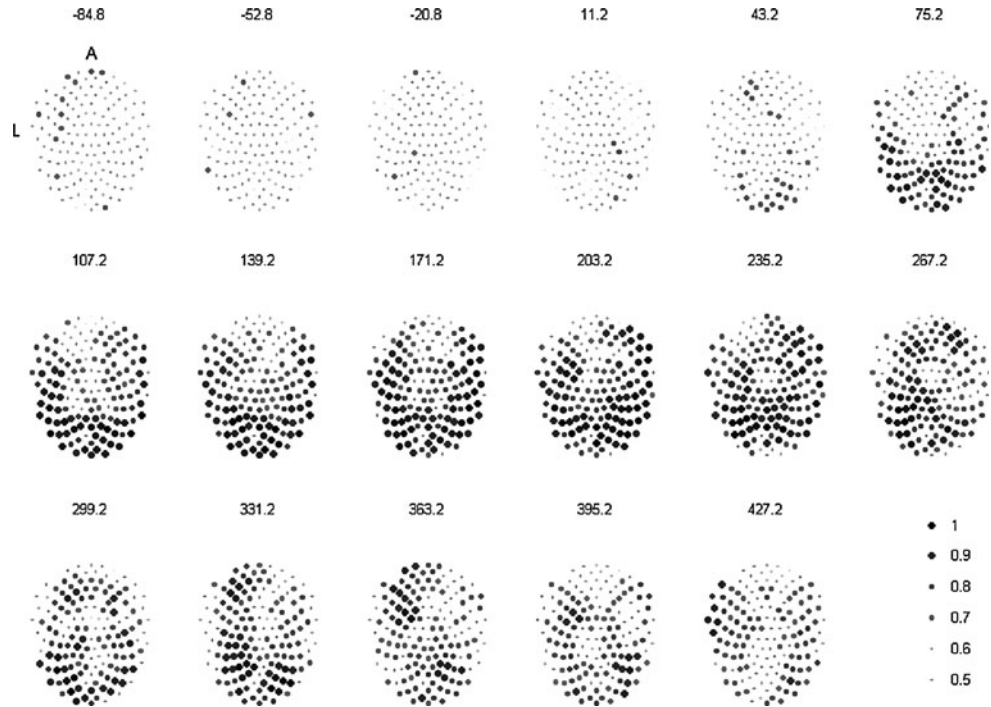
Fig. 5 Spatial distribution of P_i ($i = 1,…,N$) (defined by Eq. 5) over 17 time windows. The *number* above each image denotes the time (ms) at the middle of the respective 80 ms windows with reference to the presentation of the cue. The area of each *dot* represents the value of P. A and L indicate *anterior* and *left sides* of the head in the plots

the parietal and temporal lobes and finally, an increase in frontal lobe sensors which occurs at about the same time that the button is pressed and afterwards.

The best overall classification accuracies obtained by applying the CSP method for a fixed time window of 176 ms wide were achieved when the window was centered at 200 ms after the cue, although some variation existed across subjects. Figure 6 shows the relative contribution of each MEG channel to linear classification using this method averaged across all 6 features and all subjects. Since CSP coefficients can be either negative or positive the graph depicts the average absolute value of the coefficients. The overall mean classification accuracy was 80.6% with classification accuracies for each subject being 80.3, 73.2, 77.3, 83.2, 85.1 and 84.8% with all standard errors less than 1%. A non-parametric test using classification of trials with randomized labels (Golland and Fischl 2002) showed that, in 10^6 attempts, no classification accuracy over 55% should be expected. Better classification accuracy can be achieved when the window is further increased in size, as with the FC. This may be explained by (Huang et al. 2010) where it was found that when working with a large set of channels—in our case, 143–151—longer windows improved accuracy by preventing overfitting.

Discussion

Although many studies describing the neural correlates of decisions, choices and actions have been produced over the last years, to our knowledge no specific comparison of
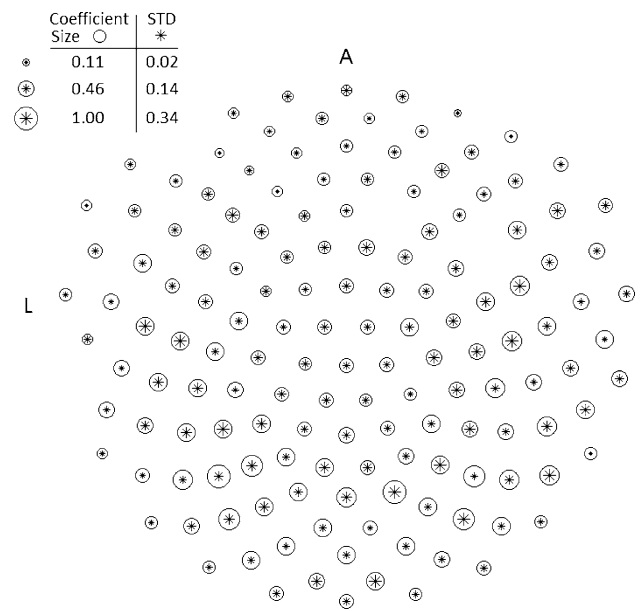

Fig. 6 Relative weights of CSP coefficients (described in part 3 of the "Data analysis" section). For each subject these weights were calculated as the normalized sums of absolute values of the six coefficients per channel for a window 176 ms wide centered at 200 ms. The diameter of *each circle* is a linear function of the weights and it is a measure of their relative contribution to the discrimination of the conditions. The *stars* represent the standard deviation of these sums across subjects. In the legend, the *stars* and the *circles* values should be interpreted separately. A and L indicate *anterior* and *left sides* of the *head* in the plot

brain states preceding free and forced actions, in relation to the selection of alternatives, has been described. Our initial result using the FC provides some evidence for the

differences between conditions in a way analogous to traditional ERP analysis. This time, instead of using the raw ERPs from different channels, we used a linear combination that produces the best separation of conditions, at point in time, according to FC. The finding of a significant deviation from baseline in $F(w, t)$ at 64 ms should be cautiously interpreted. A study on visual processing (VanRullen and Thorpe 2001) shows that these early differences (in their case 75–80 ms, also in MEG) may not be task related, only due to early categorizations of the visual input not yet related to behaviour. Notice, also, that the maximal separation was reached at around 0.2 s, in agreement with the other two measures in this study.

Although we found considerable differences in mean MEG signals during forced and free button presses, we were most interested in studying these differences in single trials. That is, taking a broader approach and using potentially more sources of information, as for example, dynamical patterns of brain activity. Since trials are not averaged when performing single trial analysis, there is a risk that forced actions might not be differentiable from willed ones in extracranial recordings as the difference may only involve the activity of a small number of neurons (Schall 2001), which may remain buried in noise. Our observations, however, indicate that this is not the case. Using the correlation analysis of trials we explored whether temporal patterns can also be used as a criterion to identify the experimental condition to which a trial belongs. This analysis revealed that, in some cases, more than 55% of MEG channels show a pattern of activation in time that is significantly different between the two conditions even in short 80 ms windows. While the exact dynamical pattern that produces the best separation is left unidentified and no actual trial classification was attempted in this case, the simple correlation function was sufficient to identify similarities between signals from a common experimental condition. The spatial distinction observed here involved mostly MEG sensors located above occipital areas and spreading to parietal and temporal areas, with little involvement of the frontal sensors until after a decision was made.

The classification of single trials is commonly used for practical applications, as in the case of brain computer interfaces (BCIs), but is increasingly being used for statistical analyses. In our case the classification approach was designed to complement the previous two analyses. The task of finding the appropriate classification features that contain enough information to distinguish one condition from the other is complex, due to the unlimited possibilities for defining a feature and the considerable amounts of uncorrelated noise polluting individual trials. In practice, a few methods are available based on physiological motivations and we used two of these: FC because it can detect divergence in MEG signals from different conditions with high temporal resolution, and CSP because it is one of the most successful algorithms for feature invention and has a straightforward spatial interpretation. The results of the FC classification indicate similarities to the temporal and spatial aspects of the correlation analysis (Fig. 4). The CSP method (Fig. 6) identifies a similar spatial pattern to that observed with previous analysis (Figs. 3, 5) while obtaining a high degree of classification accuracy (80.6%). Although the original CSP method has been extended and improved in a number of directions (Lemm et al. 2005; Tomioka et al. 2006; Dornhege et al. 2006; Blankertz et al. 2008), the present study utilizes the original implementation since we were not interested in specific practical applications, where the highest classifications rates are desirable. Although it is possible that using one of the many extensions of the method could improve our classification accuracy, the values already attained are high considering that no artifacts were removed in the data and no trials were discarded based on artifacts. With respect to spatial findings, since typical inter-subject variability in these types of tasks is relatively high (Blankertz et al. 2006a, b), and the sample is not large, the tools we used in the present study may only be appropriated to produce a rough or coarse-grained picture of the spatial characteristics that separate one condition from the other. Correspondingly it was not our purpose in the present study to precisely locate neural generators but to assess whether the conditions can, in principle, be separated using raw neurophysiological recordings by an approach based on high-resolution temporal information.

The forced or willed motor actions considered in our study are most likely a result of distinct coordinated activity in distributed networks (Baars et al. 2003; Pesaran et al. 2008; Desmurget and Sirigu 2009) as supported by the widespread spatial nature of cortical involvement in distinguishing these actions. Neuroimaging evidence suggests that motor behaviour is initiated in frontal cortices and that the sense of volition occurs after a corollary discharge in multiple brain areas (Hallett 2007), and points to a close relation between the activity in frontal and parietal cortices which can precede the outcome of a free decision by up to 10 s (Soon et al. 2008). We did not find much contribution from frontal areas, which may be attributed to the almost "mechanical" task considered here, where symbols are designed to be very intuitive so that they can be directly interpreted without a large cognitive load. It is conceivable that in more demanding tasks for decision making, frontal areas may play a key role, and we are now proceeding with further experiments to assess this. Another possible interpretation of these findings is that the number of cells in frontal areas in which these differences are realized may be few and insufficient to produce a

measurable field. Furthermore, as reported by (Soon et al. 2008) in an fMRI study, there was no increase in signal strength in frontal regions during a decision but rather, predictive information was encoded in differences in spatial patterns of fMRI measurements, and so the methods discussed in this study may be best suited for detecting differences in signal strength rather than fine spatial changes in neural activity.

Viewing these analyses as a whole, this study cannot determine, with certainty, when the subjects make the decision to press a button or when they become conscious of their decision. However, comparing our results in the temporal domain with previous ones in visual awareness (Lamme 2003) it seems most likely that the peak in discrimination around 200 ms is the single event most directly correlated to the awareness of the current condition, and that the early discrimination at ∼64 ms may only account for early visual categorization of the cue, not yet conscious. The stepped slope in classification accuracy which initiates around 50 ms and reaches its maximum at 200 ms seems to indicate that during the early stages of visual processing, the subject is already initiating the neural processes that support decision making. Although not shown here, similar analyses were performed using the button press as a reference for all trials and this showed that neural activity associated with a decision is detectable as early as 300 ms before a button press (recall that the average time between cue presentation and the response was 320 ms). For the purposes of our analysis, the early categorization of the cue can be considered as an unwanted effect because it may mask the classification that is due to the conscious experience. We are now proceeding to the design of related experiments where no cue categorization is involved. The new task, however, has to make heavy use of short term memory. Consequently, unlike in classical physics, no simple model may be possible when it comes to study consciousness/awareness, and keeping a reductionist approach in trying to isolate conscious experience from early unconscious processes may not even make sense in most cases after all (Perez Velazquez and Frantseva 2010).

In this study, we show that by using single-trial analysis, it is possible to classify MEG recordings that were obtained during the performance of either a forced or chosen behaviour. This is possible because of the differences in the underlying cortical activity related to producing these two behaviours. In particular, occipital, parietal, and temporal sensors contributed most to these differences at various stages of the decision process and as a result, the cortical structures underlying the signals at these sensors can be thought of as important for the cortical processing involved in generating a decision based on an instruction. We also observed, contrary to previous findings, that frontal areas did not contribute significantly to the discrimination of forced and free actions although the nature of the experiment being studied here may have contributed to this finding.

References

Baars BJ, Ramsoy TZ, Laureys S (2003) Brain, conscious experience and the observing self. Trends Neurosci 26:671–675

Blankertz B, Dornhege G, Lemm S, Krauledat M, Curio G, Müller KR (2006a) The Berlin Brain-Computer Interface: machine learning based detection of user specific brain states. J Universal Comput Sci 12:581–607

Blankertz B, Krauledat M, Losch F, Curio G, Müller KR (2006b) Combined optimization of spatial and temporal filters for improving brain-computer interfacing. IEEE Trans Biomed Eng 53:2274–2281

Blankertz B, Kawanabe M, Tomioka R, Hohlefeld F, Nikulin V, Müller KR (2008) Invariant common spatial patterns: alleviating nonstationarities in brain-computer interfacing. Adv Neural Inform Process Syst 20:113–120

Damasio A (1999) The feeling of what happens: body and emotion in the making of consciousness. Harcourt Brace, New York

Desmurget M, Sirigu A (2009) A parietal-premotor network for movement intention and motor awareness. Trends Cogn Sci 13:411–419

Dornhege G, Blankertz B, Krauledat M, Losch F, Curio G, Muller KR (2006) Optimizing spatiotemporal filters for improving brain-computer interfacing. In: Advances in Neural Info Proc Systems (NIPS 05), vol 18. MIT Press, Cambridge, MA, pp 315–322

Frith CD, Friston K, Liddle PF, Frackowiak RS (1991) Willed action and the prefrontal cortex in man: a study with PET. Proc R Soc Lond B 244:241–246

Golland P, Fischl B (2002) Permutation tests for classification: towards statistical significance in image-based studies. Inf Process Med Imag 2003(18):330–341

Haggard P (2008) Human volition: towards a neuroscience of will. Nat Rev Neuro 9:934–946

Hallett M (2007) Volitional control of movement: the physiology of free will. Clin Neurophysiol 118:1179–1192

Huang G, Liu G, Meng M, Zhang D, Zhu X (2010) Model based generalization analysis of common spatial pattern in brain computer interfaces. Cogn Neurodyn 4:217–223

Ingvar DH (1994) The will of the brain: cerebral correlates of willful acts. J Theor Biol 171:7–12

Kaiser J, Heidegger T, Lutzenberger W (2008) Behavioral relevance of gamma-band activity for short-term memory based auditory decision-making. Eur J Neurosci 27:3322–3328

Lamme VAF (2003) Why visual attention and awareness are different. Trends Cogn Neurosci 7:12–18

Lemm S, Blankertz B, Curio G, Muller KR (2005) Spatio-spectral filters for improved classification of single trial EEG. IEEE Trans Biomed Eng 52:1541–1548

Libet B (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action. Behav Brain Sci 8:529–566

Manes F, Sahakian B, Clark L, Rogers R, Antoun N, Aitken M, Robbins T (2002) Decision-making processes following damage to the prefrontal cortex. Brain 125:624–639

Müller KR, Krauledat M, Dornhege G, Curio G, Blankertz B (2004) Machine learning techniques for brain-computer interfaces. Biomed Eng 49:11–22

Müller-Gerking J, Pfurtscheller G, Flyvbjerg H (1999) Designing optimal spatial filters for single-trial EEG classification in a movement task. Clin Neurophysiol 110:787–798

Perez Velazquez JL, Frantseva MV (2010) The brain-behaviour continuum: the subtle transition between sanity and insanity. Imperial College Press (in press)

Pesaran B, Nelson MJ, Andersen RA (2008) Free choice activates a decision circuit between frontal and parietal cortex. Nature 453: 406–409

Ramoser H, Müller-Gerking J, Pfurtscheller G (2000) Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans Rehabil Eng 8:441–446

Rushworth MFS (2008) Intention, choice, and the medial frontal cortex. Ann NY Acad Sci 1124:181–207

Schall JD (2001) Neural basis of deciding, choosing and acting. Nat Rev Neurosci 2:33–42

Soon CS, Brass M, Heinze HJ, Haynes JD (2008) Unconscious determinants of free decisions in the human brain. Nat Neurosci 11:543–545

Tomioka R, Dornhege G, Nolte G, Blankertz B, Aihara K, Müller KR (2006) Spectrally weighted common spatial pattern algorithm for single trial EEG classification. Technical Report 40, Department of Mathematical Engineering, The University of Tokyo, Tokyo

VanRullen R, Thorpe SJ (2001) The time course of visual processing: from early perception to decision-making. J Cogn Neurosci 13:454–461