Journal of Neuroscience Methods 250 (2015) 106–113

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Basic Neuroscience

Decoding auditory EEG responses in healthy and clinical populations: A comparative study

Marzia De Lucia[a,b,c,*], Athina Tzovara[a,b,c]

[a] Laboratoire de recherche en Neuroimagerie (LREN), Department of Clinical Neurosciences, Lausanne University and University Hospital, 1011 Lausanne, Switzerland
[b] Radiology Department, Vaudois University Hospital and University of Lausanne, 1011, Switzerland
[c] Center for Biomedical Imaging (CIBM) of Lausanne and Geneva, 1011, Switzerland

ARTICLE INFO

Article history:
Received 30 May 2014
Received in revised form 21 October 2014
Accepted 22 October 2014
Available online 3 November 2014

Keywords:
Decoding
Single-trial
EEG
Logistic regression
Coma
MMN

ABSTRACT

Background: Analyses of brain responses to external stimuli are typically based on the means computed across conditions. However in many cognitive and clinical applications, taking into account their variability across trials has turned out to be statistically more sensitive than comparing their means.
New method: In this study we present a novel implementation of a single-trial topographic analysis (STTA) for discriminating auditory evoked potentials at predefined time-windows. This analysis has been previously introduced for extracting spatio-temporal features at the level of the whole neural response. Adapting the STTA on specific time windows is an essential step for comparing its performance to other time-window based algorithms.
Results: We analyzed responses to standard vs. deviant sounds and showed that the new implementation of the STTA gives above-chance decoding results in all subjects (in comparison to 7 out of 11 with the original method). In comatose patients, the improvement of the decoding performance was even more pronounced than in healthy controls and doubled the number of significant results.
Comparison with existing method(s): We compared the results obtained with the new STTA to those based on a logistic regression in healthy controls and patients. We showed that the first of these two comparisons provided a better performance of the logistic regression; however only the new STTA provided significant results in comatose patients at group level.
Conclusions: Our results provide quantitative evidence that a systematic investigation of the accuracy of established methods in normal and clinical population is an essential step for optimizing decoding performance.

1. Introduction

An increasing trend in developing powerful and neurophysiologic interpretable multivariate decoding techniques has considerably improved the sensitivity in extracting brain signals of interest in high spatio-temporal resolution neuroimaging data (Pereira et al., 2009; Blankertz et al., 2011; Lemm et al., 2011). Typically, multivariate decoding aims at assessing spatio-temporal patterns of activity which best discriminate single neural responses across experimental conditions of interest. Many neuroimaging studies have shown that a data-driven approach in extracting such spatio-temporal patterns can reveal effects that would remain overlooked after collapsing data to their means and by comparing them based on univariate statistical approaches (Kahnt et al., 2011; Wyart et al., 2012; He and Zempel, 2013).

In electroencephalography (EEG) studies the main challenge is developing multivariate decoding methods with two important features: the first important property is the sensitivity of the method in revealing statistically significant results, that is to say being robust to the variability induced by the noise characteristics of the EEG data recorded in humans. This property is required by any kind of application based on EEG data acquired in humans ranging from Brain Computer Interface (Millan et al., 2002; Wolpaw et al., 2002; Muller et al., 2008) to classification problems in clinical applications and cognitive studies (Blankertz et al., 2011; Lemm et al., 2011). In basic neuroscientific research a second important property of multivariate decoding analysis is the neurophysiologic interpretability of the employed features (van Gerven et al., 2009;

* Corresponding author at: Laboratoire de Recherche en Neuroimagerie, Department of Clinical Neurosciences, Lausanne University Hospital, Chemin de Mont-Paisible 16, 1011 Lausanne, Switzerland. Tel.: +41 213146828.
E-mail address: Marzia.De-Lucia@chuv.ch (M. De Lucia).

http://dx.doi.org/10.1016/j.jneumeth.2014.10.019
0165-0270/

Blankertz et al., 2011; Haufe et al., 2014). This interpretability is essential for gathering insights into the neural process underlying the differential effect uncovered by an above-chance decoding performance. In other words we are interested in determining which frequencies, latencies or brain regions subtend a differential activity that can be quantified by a multivariate decoding analysis.

In EEG research it is possible to derive evidence about the difference in the spatial configuration of neural sources between experimental conditions by comparing the spatial distribution of the corresponding voltage topographies (Lehmann, 1987; Tzovara et al., 2012a). This is simply a consequence of the physics governing the EEG, according to which a change over time or across conditions of the voltage distribution across the electrodes montage is necessarily a consequence of a change in the underlying neural sources.

Previous research in this direction has demonstrated the feasibility of extracting informative patterns of voltage topographies for decoding responses at the level of single-trial EEG and across sensory modalities (Bernasconi et al., 2011; De Lucia et al., 2012; Tzovara et al., 2012d; Cossy et al., 2014). A single-trial topographic analysis (STTA) has been originally designed so as to discover in a data-driven manner those time periods and those prototypical voltage topographies carrying the most accurate discriminative information between conditions of interest (Tzovara et al., 2012c). This method can discover in a data-driven manner the latency or latencies at which the distribution across electrodes of the voltage measurements differ mostly. In contrast, in other popular decoding methods such as those based on logistic regression (Parra et al., 2005; Ratcliff et al., 2009; van Gerven et al., 2009; Philiastides et al., 2010; Brandmeyer et al., 2013; Farquhar and Hill, 2013) or support vector machines (Das et al., 2010; Charles et al., 2014; Taghizadeh-Sarabi et al., 2014), the accuracy of the method is usually assessed along predefined fixed-length temporal windows sliding over the whole response epoch (Philiastides et al., 2010; De Vos et al., 2012; Hausfeld et al., 2012). In this paper we aim at adapting the previously proposed STTA in order to allow a direct comparison of its performance with other commonly used multivariate decoding methods. We specifically focus on the comparison between this adapted version of the STTA and the logistic regression algorithm (Parra et al., 2002, 2005). One of the main reasons of our interest in logistic regression is the simplicity of its implementation and also the possibility to derive the scalp projection of the most discriminative features (Parra et al., 2002). The STTA and the logistic regression analysis both share such properties that it is in principle possible to estimate the neural sources underlying the difference between experimental conditions on the basis of the distribution of voltages across electrodes montage responsible for the above-chance decoding (Parra et al., 2005; Tzovara et al., 2012b).

The comparison between these two decoding algorithms is based on two sets of EEG data; one recorded on healthy subjects during passive listening of standard and deviant sounds in a mismatch negativity protocol; the second recorded at bedside on comatose patients using the same experimental protocol. In addition we perform a further analysis with the logistic regression in order to verify the replicability of previously published results obtained with the STTA and showing that the progression of auditory discrimination over the first two days of coma is highly predictive of patients' chance of awakening (Tzovara et al., 2013).

2. Materials and methods

2.1. Control subjects and coma patients

We included 11 healthy control subjects, 3 males and 8 females with a mean age of 54 ± 1 (mean $\pm$ standard error are indicated here and in the following) years old. All reported normal hearing and none of them had a history of neurological or psychiatric illnesses. The first 5 of these subjects were already included in another study from our group (Tzovara et al., 2013).

In addition, we considered data from 11 post-anoxic comatose patients, 8 males and 3 females, with a mean age of 59.2 ± 4.0 years old randomly selected from those included in a previous study (Tzovara et al., 2013). All comatose patients included in this study were treated with therapeutic hypothermia (TH; Oddo et al., 2006) as previously described. The study was approved by the Ethics Committee of the institution.

On these two groups of 11 healthy controls and 11 patients we carried out an extensive analysis all the analyses using the original implementation of the STTA, the STTA optimized on fixed temporal window and the logistic regression.

Additionally, we re-examined a larger dataset of 30 comatose patients (Tzovara et al., 2013) for studying the reproducibility of our previous results, using the logistic regression algorithm (see Supplemental Material).

2.2. Paradigm

We used a mismatch negativity (MMN) paradigm, including frequently repeated standard sounds and three types of deviants, in terms of duration, pitch or location. All sounds were pure sinusoidal tones and were presented to the healthy controls and coma patients via earphones. They were 16-bit stereo sounds, sampled at 44.1 kHz. At the onset and offset of all the stimuli a 10-ms linear amplitude envelope was applied to avoid clicks.

The inter-stimulus interval was kept constant at 700 ms. Standard sounds were 1000 Hz sinusoidal tones of 100 ms duration. The three types of deviant sounds were presented in a pseudo-randomized way, at 10% of the time each. Duration deviants were identical to the standards but for their duration (1000 Hz tones; 150 ms duration; 0 ms inter-aural difference). Pitch deviants were 1200 Hz tones with 100 ms duration and 0 ms inter-aural difference, while location deviants were 1000 Hz tones, with 100 ms duration and 700 μs inter-aural time difference (left ear leading).

Sounds were presented in three blocks consisting of 500 stimuli and lasting ~7 min each (1500 stimuli per patient/control). The sounds' presentation was pseudo-randomized, so that there were never two deviants occurring one after the other. In all the analyses we report in this study we always consider responses to standard sounds preceding a deviant one and following another standard.

2.3. EEG acquisition

We used the same clinical EEG system for the recordings in healthy controls and comatose patients (Viasys Neurocare, Nicolet). For reasons of consistency between the two groups, recordings in controls took place in a hospital room, where they were instructed to lie on an inclined chair, close their eyes and listen to the sounds. Recordings on coma patients were performed by the patient's bedside, in the intensive care unit without interrupting the clinical routine. EEG was recorded in one session in healthy controls and two sessions in comatose patients, the first during hypothermia and <24 h from coma and the second after re-warming to normal body temperature (normothermia; NT) and within ~24–48 h from coma onset.

EEG was recorded from 19 electrodes, positioned according to the international 10–20 system. Sampling rate was of 1000 Hz, with an online reference to Fpz electrode. All impedances were controlled at the beginning of the recording and were kept below 10 kΩ.

2.4. EEG preprocessing

Data were filtered offline using high- and low-pass butterworth filters, with cutoff frequencies of 0.1 and 40 Hz, respectively and a notch filter at 50 Hz. Peri-stimulus epoch were extracted, with 50 ms of pre-stimulus baseline and 500 ms post-stimulus interval (550 ms epoch length in total). A semi-automatic artifact rejection criterion of ±100 μV was applied on all electrodes for excluding noisy trials. Finally, data from noisy electrodes were interpolated using 3D splines (Perrin et al., 1987). Data were then re-referenced to the common average reference. All pre-processing was performed using Cartool software (Brunet et al., 2011). Prior to the decoding analyses, all voltage topographies were normalized by their instantaneous Global Field Power, computed as the standard deviation across electrodes (Tzovara et al., 2012a).

2.5. Single-trial decoding

We analyzed single-trial EEG responses at the level of each single patient/control subject and recording, using two different decoding approaches, one based on modeling voltage topographies using a mixture of Gaussians statistical distribution to which we refer as STTA (Tzovara et al., 2012c) and the second based on logistic regression (Parra et al., 2005). The general scheme of the single-trial analysis was the same for both approaches and consisted of the following steps:

1. *Splitting the available trials for each experimental condition* in two datasets: a cross-validation (CV) and a validation (V) dataset.
2. *Feature extraction in the CV dataset*: the CV included 70 trials for training and testing the decoding algorithm. The CV datasets were split in 7 partitions; among them, 6 were used for training the algorithm and the remaining one for testing based on a 7-fold cross-validation procedure. During training the decoding algorithm learned the discriminative features between two experimental conditions (here responses to standard vs. deviant sounds). The suitability of the extracted features is tested by decoding the single-trials of the 7 test datasets. The most discriminative features/parameters are selected by maximizing the mean decoding performance across the 7 test datasets.
3. *Validation of the extracted features:* this step consists of decoding V trials as belonging to one of two experimental conditions and examining if decoding performance was above-chance level.

It is worth noting at this point that the results based on the CV dataset provide a first estimation of the decoding performance. However these results might be biased because they are based on optimal discriminative features selection. In order to obtain a more objective evaluation of the decoding performance and to assess how well the CV results generalize to new data, we evaluated the performance of the selected algorithm on a validation dataset (i.e. V dataset). For a direct comparison between the STTA and the logistic regression, we always use the same splits of single-trials for CV and V.

In the following, "$\boldsymbol{m}$" consists of a 19-dimensional vector of voltage values, recorded across the electrode montage at a given point in time. "h" refers to a time-point within a trial, with h ranging from 1 to L, where "L" is the single-trials' length (here, 550 ms).

2.5.1. STTA: single-trial topographic analysis

The analyses we report in the current study are based on a previously introduced decoding algorithm adapted for easing the comparison of its performance to other decoding methods. We briefly describe here the bases of the single-trial topographic analysis (STTA) and how it has been modified for the analysis of single-trial EEG on predefined time-windows.

2.5.1.1. General principles of STTA. The STTA is based on modeling the configuration of voltage topographies with a Mixtures of Gaussians Model (GMM), and then on using these models for decoding the type of response for new, validation trials. Full details of this method have been extensively described elsewhere (Tzovara et al., 2012d) and the suitability of this approach for analyzing EEG data from healthy controls and comatose patients has been demonstrated in several studies (Bernasconi et al., 2011; De Lucia et al., 2012; Tzovara et al., 2012b; Cossy et al., 2014).

Voltage topographies of single trials belonging to the training datasets of the two experimental conditions are used for computing the GMM models, based on an expectation-maximization algorithm (Dempster et al., 1977). Models are computed in an n-dimensional space, where n equals the total number of electrodes, separately for each experimental condition. After the GMM calculation, we use posterior probabilities for computing the degree of confidence according to which we can assign the voltage topographies to each of the Gaussians in the two mixtures:

$$P(c_k|\boldsymbol{m}) = \frac{P(\boldsymbol{m}|c_k)\cdot p_k}{p(\boldsymbol{m})} \quad (1)$$

c_k indexes one Gaussian within the mixture, $p(\boldsymbol{m})$ the unconditional density function and p_k the prior probability of the kth Gaussian.

The analysis described so far does not take into account the temporal order of appearance of the voltage topographies within a trial, but for the following, they are re-arranged to their original time-point and trial. By considering the posterior probabilities across all trials, at a time-point by time-point level, one can then carry out statistics and identify discriminative time-periods H, with $H \leq L$ between the two experimental conditions, and also the Gaussians in the two models which are responsible for each time-point within these periods (see Figure 6 in Tzovara et al., 2012a for a schematic example). All this analysis is done in the training dataset and consists of a feature extraction (i.e. time-periods of difference and template maps for the two experimental conditions during these periods).

The extracted features are then used for decoding the test trials. Specifically, for each test trial t, a discrimination function, DF_t, is computed by subtracting the posterior probabilities for the topographies of this trial to belong to the model of one experimental condition or the other, along the discriminative time-intervals H:

$$DF_t = \sum_{h\varepsilon H}[\mathrm{P}(c_{i_h}|\boldsymbol{m}_{\mathrm{th}}) - \mathrm{P}(c_{j_h}|\boldsymbol{m}_{\mathrm{th}})] \quad (2)$$

where c_{i_h} and c_{j_h} represent the most prominent Gaussians at the time-point h of the two models i and j, respectively. The index h spans along the discriminative interval H, as it has been identified in the training dataset.

Each single-trial t can be assigned to that experimental condition for which the sum of posterior probabilities is higher during the interval H (here, responses to standard or deviant sounds). Decoding performance is quantified by the Area Under the Receiver Operator Characteristic curve (AUC; Green and Swets, 1966). The abovementioned analysis is repeated with a varying range of total Gaussians in the two models from 3 to 10, for each of the two experimental conditions. Through a cross-validation procedure we tested each possible combination of the computed models and selected as optimal the pair of models [q1, q2] that maximized the mean AUC values across the 7 test datasets.

Finally, the selected models are validated in the set of validation trials. The performance of this final validation is tested against

chance level, where chance is computed by randomly permuting the labels of the CV trials 200 times and recomputing the GMM models for random partitions of the data. The significance of the actual performance obtained on the V dataset is tested against chance levels by a Wilcoxon signed rank test ($p < 0.001$) with Matlab's function *signrank*.

2.5.2. STTA on predefined time-window

The algorithm described so far represents the generic implementation of the STTA analysis. In the present study we specified this general framework by optimizing the model selection for short intervals, of 50 ms each (10 time-windows in total). The original GMM models were computed using the whole dataset (i.e. the whole trials' length) but instead of selecting one pair of optimal models for the whole trial [q1, q2], in this case we selected a different pair for each of the short intervals [$q1_\tau$ and $q2_\tau$], where τ is an index for each window. This selection was done using the CV trials, by maximizing the discrimination in the 7 test datasets for each of the windows. Using the selected pairs for each of the windows, we then computed decoding for the validation trials and assessed its significance with a Wilcoxon signed rank test.

2.5.3. Logistic regression

2.5.3.1. General implementation. Logistic regression is a popular machine learning algorithm for binary classification problems (Parra et al., 2002, 2005; Ratcliff et al., 2009). The main assumption of this algorithm is that the posterior probability of a set of N-dimensional data $\boldsymbol{m}$ can be modeled as a logistic function parameterized by the vector $\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N)$:

$$P(k = 1|\boldsymbol{m}, \boldsymbol{w}) = \frac{1}{1 + \exp(-\boldsymbol{m}^T \boldsymbol{w})} \quad (3)$$

where k can assume the two values 0 and 1, according to classes' labels. Based on this assumption it is possible to estimate a linear hyperplane that best separates the two classes:

$$y(t) = \boldsymbol{w}^T \boldsymbol{m}(t) = \sum_{i=1}^{N} w_i m_i(\boldsymbol{t}) \quad (4)$$

This model performs at its best when data follow a Gaussian, Poisson or Bernoulli distribution. The parameter $\boldsymbol{w}$ defining the separating hyperplane $y(t)$ is computed based on a maximum likelihood estimator. In this study we adopted a penalized maximum likelihood methods based on an l2-regularization as proposed in (Conroy and Sajda, 2012). In many previous applications of the logistic regression for the classification of EEG components, $y(t)$ is computed over a predefined temporal window τ sliding over the whole response epoch. In correspondence to each discrimination, this classification algorithm allows extracting the scalp topography describing the electrical coupling of the discriminating component y_τ.

$$\boldsymbol{a}_\tau = \frac{\boldsymbol{X}\boldsymbol{y}_\tau}{\boldsymbol{y}_\tau^T \boldsymbol{y}^T} \quad (5)$$

This scalp topography can be also interpreted as the linear projection of the neural sources responsible for the best separation between the experimental conditions and provides in principle the basis for computing an inverse solution (Parra et al., 2005).

3. Results

In this section, we always contrasted standard vs. pitch deviant sounds (Fig. 1 for event-related potentials in response to the two sounds for an exemplar healthy control and an exemplar comatose patient). We report the results obtained on a validation dataset for each of the subjects separately first by applying the STTA in its standard formulation and second as based on the optimal parameters selection of a predefined sliding window. We refer in the following to this second type of algorithm as STTA-W.

Table 1
Summary of decoding results for controls and coma patients with the STTA-W and the logistic regression algorithms. For every group of subjects we report the ⟨AUC max⟩, which corresponds to the average of the maximum AUC value that we could obtain for every control/patient, irrespective of the latency. Max(⟨AUC⟩) refers to the maximum average of the AUC values across all controls/patients, at the same latency. Max(⟨AUC⟩) corresponds to the peaks of AUC values in Fig. 2. In healthy controls the logistic regression algorithm outperforms the STTA-W, while in comatose patients the STTA-W provides higher decoding results.

a. Healthy controls

	STTA-W	Logistic regression
⟨AUC max⟩	0.59 ± 0.01	0.67 ± 0.01
Max(⟨AUC⟩)	0.54 ± 0.02	0.63 ± 0.02
Latency	50 ms	300 ms

b. Patients TH

	STTA-W	Logistic regression
⟨AUC max⟩	0.66 ± 0.02	0.66 ± 0.02
Max(⟨AUC⟩)	0.55 ± 0.03	0.54 ± 0.02
Latency	400 ms	200 ms

c. Patients NT

	STTA-W	Logistic regression
⟨AUC max⟩	0.63 ± 0.02	0.67 ± 0.03
Max(⟨AUC⟩)	0.55 ± 0.03	0.53 ± 0.04
Latency	250 ms	350 ms

3.1. Decoding results in healthy controls

3.1.1. STTA and STTA-W in healthy controls

Based on the STTA, 7 of the 11 control subjects provided an above-chance decoding performance ranging from 0.52 to 0.60 with an average of 0.56 ± 0.01. Using the same split of cross-validation and validation datasets for each subject, we computed the decoding performance on a sliding window of 50 ms starting from 50 ms before stimulus arrival on the basis of the STTA-W.

Based on the STTA-W, all subjects exhibited at least one latency of significant decoding performance. In particular, for the 4 subjects who did not have an above-chance decoding performance with the STTA, with the new approach we found a significant improvement in the decoding performance with an AUC of 0.63 ± 0.01. The other 7 significant subjects with the standard analysis all confirmed an above-chance level decoding performance with similar performance as obtained with the previous analysis (0.57 ± 0.01 with STTA-W, vs. 0.56 ± 0.01 with STTA). A summary of all the results across all the 11 subjects are shown in Table 1a and displayed in Fig. 2a.

3.1.2. Logistic regression in healthy controls

Decoding single-trial EEG responses to standard vs. pitch deviant sounds using the logistic regression algorithm provided in general better results on all the healthy controls (Fig. 2b). Results were significant on most of the windows during the post-stimulus period with maximal decoding performance for each subject ranging from 0.60 to 0.70 (mean maximal AUC of 0.67 ± 0.01, referred to as ⟨AUC max⟩ in Table 1a). The mean decoding across all subjects revealed two periods of high discrimination between standard and deviant sounds, starting at 150 ms post-stimulus onset and 300 ms post-stimulus onset with average AUC values of 0.59 ± 0.02 and 0.63 ± 0.02 respectively (referred to as max(⟨AUC⟩) in Table 1a). The STTA-W algorithm provided relatively lower values of decoding with maximal decoding ranging from 0.53 to 0.64 (mean maxima

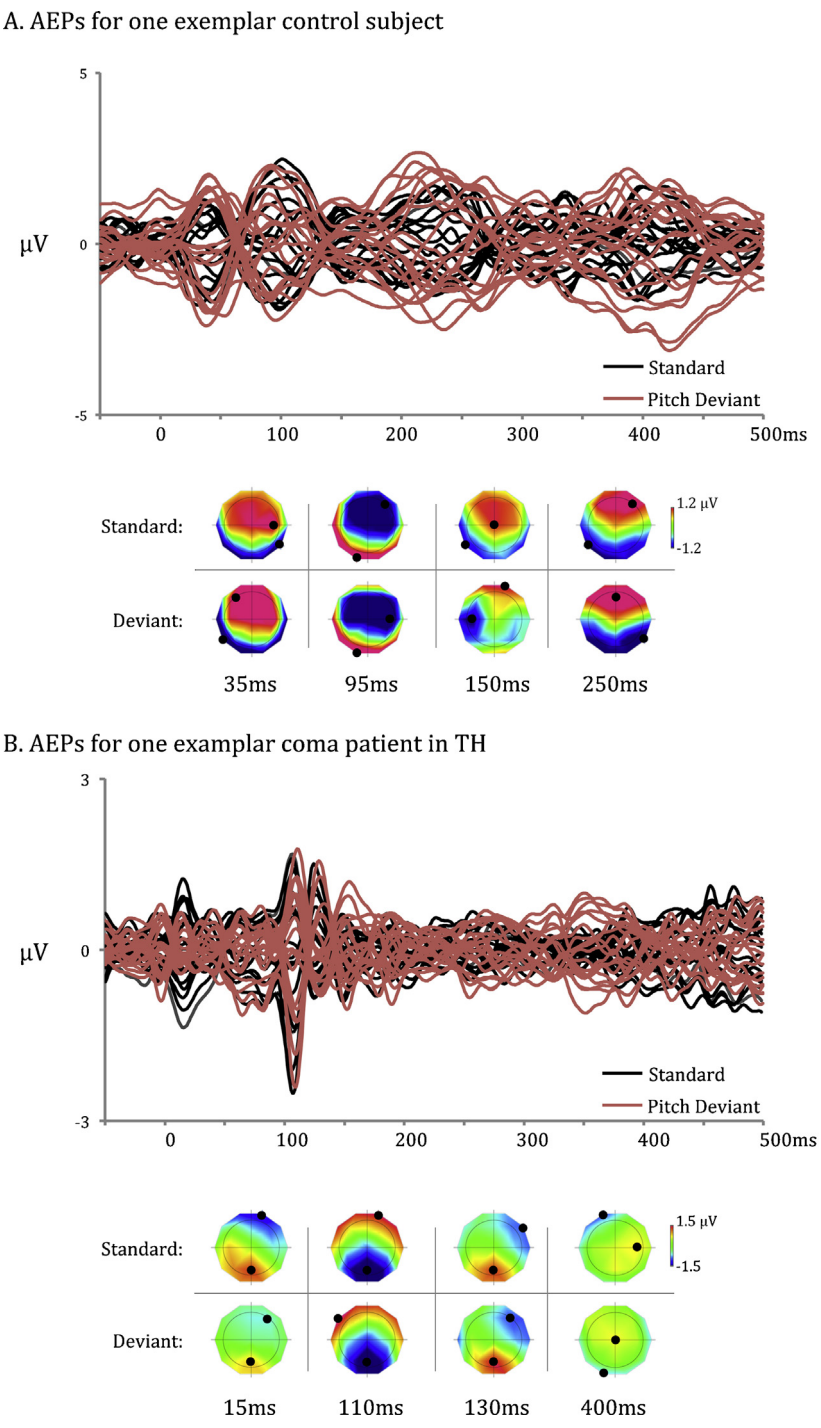

Fig. 1. Average AEPs in response to standard vs. pitch deviant sounds. (a) Average AEPs for one exemplar healthy control, across the whole electrode montage. The displayed maps correspond to average voltage topographies at representative latencies corresponding highest voltages-exhibiting typical configurations in response to auditory stimuli as for example at ~100 ms and at latencies for which the STTA-W identified significant differences (i.e. at ~150 ms, see also Fig. 2). (b) AEPs for one exemplar comatose patient, during TH. The displayed maps correspond to typical latencies of AEP responses and also at latencies for which the STTA-W identified significant differences (i.e. at ~400 ms, see also Fig. 2). The black dots highlight the position of the minimum and maximum values in the electrode montage.

of 0.59 ± 0.02). On average across subjects and for each latency we observed a maximum value of AUC starting at 50 ms post-stimulus onset ($\max(\langle AUC \rangle) = 0.54 \pm 0.02$ in Table 1a and Fig. 2a). At average level across subjects, by comparing the decoding performance for each latency to the chance level obtained for each subject separately we found significant results with the logistic regression at several latencies (*t*-test; $p < 0.05$; Fig. 2b, highlighted latencies with a circle); the STTA-W provided significant results at the same latency as its maximum value (i.e. 50 ms, Fig. 2a, circle).

3.2. Decoding results in comatose patients

3.2.1. STTA and STTA-W in comatose patients

We run the same type of analyses as described for the healthy subjects in 11 post-anoxic comatose patients, each having 2 EEG recordings (22 recordings in total). Specifically, we run the STTA, the STTA-W and the logistic regression for each patient and recording separately, and always considering single trials in response to standard and pitch deviant. The STTA analysis on the 11 patients resulted in 6 significant decoding performances during TH with

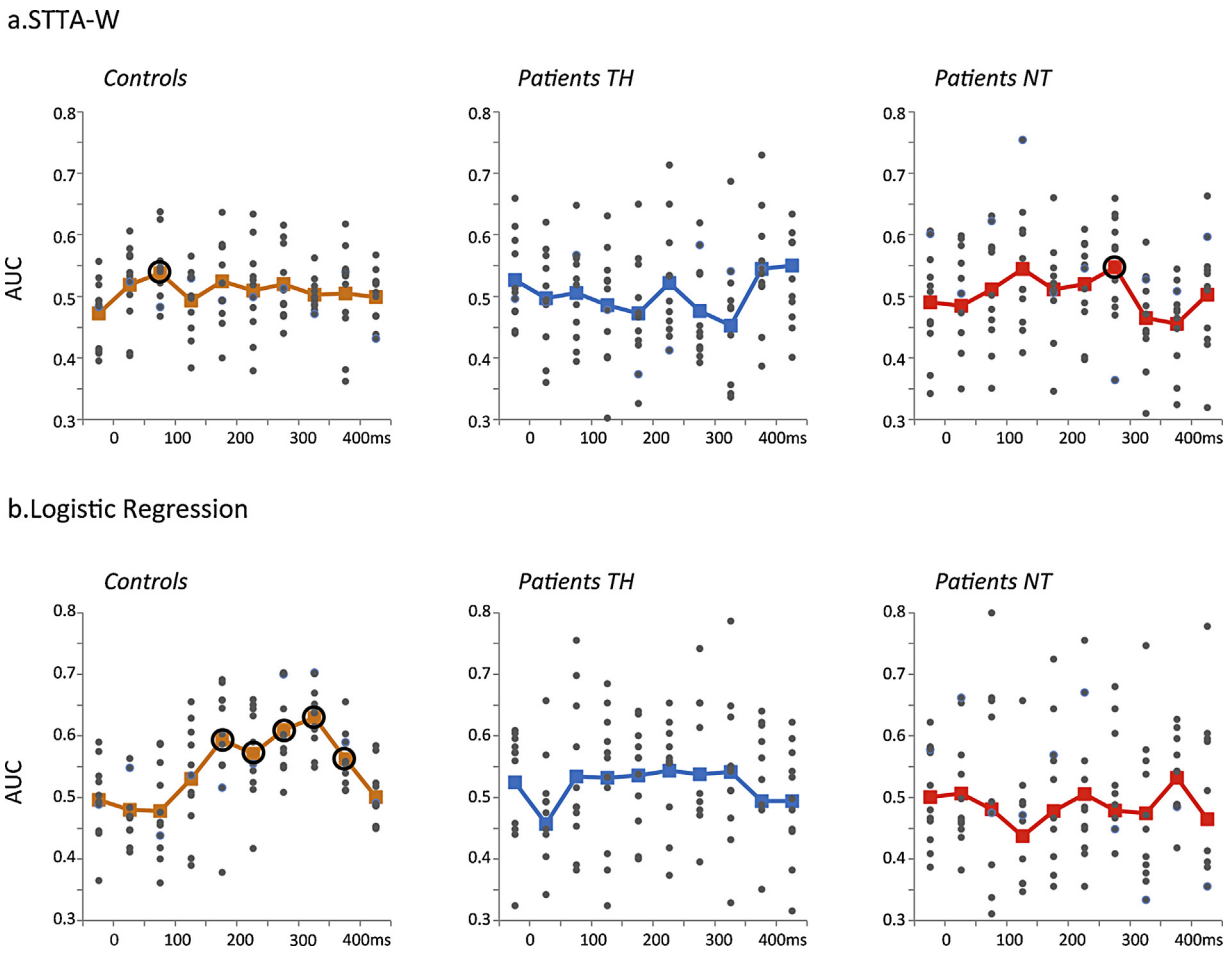

Fig. 2. Comparison of decoding results obtained with (a) STTA-W and (b) logistic regression based on sliding windows of 50 ms duration. Every point in grey corresponds to the AUC value obtained by contrasting standard vs. pitch deviant sounds and refers to results obtained in the validation dataset for single control subjects or coma patients (each point is centered on the middles of each time interval). The colored dots and lines highlight the mean decoding performance across subjects/patients at every time-point. The circled points indicate the highest decoding performance in the average across subjects and for each recording, in the cases when this performance is above-chance at the group level. The chance level for each latency and for both algorithms was 0.50 as resulted by averaging the chance level obtained for each subject and patients separately. The logistic regression provided higher decoding results than the STTA-W in healthy controls and especially at its maximum value (max⟨AUC⟩) of 0.63 ± 0.02, see also Table 1a) at 300 ms post-stimulus onset. By contrast, the SSTA-W outperformed the logistic regression in comatose patients as it resulted in the only significant decoding accuracy for patients at group level during NT (0.55 ± 0.03) at 250 ms post-stimulus onset.

AUC values ranging from 0.55 to 0.80 (average 0.63 ± 0.04) and 4 significant results during NT (range 0.52–0.59 and an average AUC of 0.56 ± 0.02). The optimization of the STTA on short time windows of 50 ms resulted in an increase of the decoding performance both during TH and NT: the 6 patients previously significant with the standard analysis provided an average maximal AUC value of 0.69 ± 0.02 in TH and 0.62 ± 0.03 in NT. The maximum value of the average decoding performance across all 11 patients during TH occurred at 400 ms post-stimulus onset (with max⟨AUC⟩) of 0.55 ± 0.03, Table 1b and Fig. 2a). During NT, the STTA-W analysis resulted on average AUC of 0.55 ± 0.03 at 250 ms post-stimulus onset (max⟨AUC⟩), Table 1c and Fig. 2a). In addition, based on the STTA-W, all patients exhibited at least one period of significant decoding with maximal AUC values for each patient ranging from 0.60 to 0.8 (average AUC 0.66 ± 0.02) in TH and a range of 0.55 to 0.75 (average AUC 0.63 ± 0.02) in NT (Table 1bc).

3.2.2. Logistic regression in comatose patients

Based on the logistic regression, the average decoding performance during TH was maximal at 200 ms post-stimulus onset (max⟨AUC⟩) of 0.54 ± 0.02 in Table 1b and Fig. 2b) and during NT at 350 ms post-stimulus onset (max⟨AUC⟩) of 0.53 ± 0.02 in Table 1c and Fig. 2b). Irrespective of the latency and across patients, the maximum values of decoding performance during TH ranged from 0.57 to 0.81 (⟨AUC max⟩ of 0.66 ± 0.02 in Table 1b); during NT the range was between 0.54 and 0.8 (⟨AUC max⟩ of 0.67 ± 0.03 in Table 1c). All patients had at least one latency at which the decoding performance was significant.

In summary, the STTA-W provided comparable performance to the logistic regression during TH and outperformed the logistic regression during NT, as it provided significant results during NT when comparing the maximum value at group level (at 250 ms post-stimulus) against the average of the decoding performance at chance level across subjects at this latency (*t*-test; $p < 0.05$). No decoding values at group level based on the logistic regression were above chance level.

4. Discussion

We propose a new optimization of a single-trial topographic analysis based on selecting the mixture of Gaussians models on predefined time-windows sliding along the response epoch (STTA-W). The goal was to assess the added value of this optimization step in comparison to the original implementation of the same algorithm for which the most discriminative time periods and topographies were derived in a data-driven manner. Adapting the

STTA on periods of predefined length offers the basis for comparing its performance to other algorithms on the same grounds. Here we reported the comparison between the decoding results obtained with a logistic regression algorithm and the STTA-W.

In healthy controls, the logistic regression provided very accurate decoding and particular on two windows along the post-stimulus interval. At average level across the five controls the results based on the STTA-W provided an earlier latency of optimal performance and with relatively smaller decoding values. The same analyses performed on eleven comatose patients during TH and NT provided significant results at group level only with the STTA-W and during NT.

These results show that the optimal choice of a decoding algorithm can be strongly influenced by the application even with the same experimental setting. We speculate here that the level of noise defined as the level of reproducibility of the response across trials to the same sensory stimuli is potentially much higher in the clinical population than in the control subjects. This observation is based on previous studies using the same datasets (Tzovara et al., 2013) and on other experimental protocols in the auditory domain in the same type of patients (Cossy et al., 2014) where the analysis of the consistency of the evoked response at each single patient level gave significant results only in two third of the patients. By contrast the same analysis revealed a consistent evoked response for all control subjects. These results are in line with previous studies in comatose patients in general, where the data of ∼30–40% of patients do not show a reliable modulation of the evoked response at typical latencies (i.e. 100 ms post-stimulus onset; cfr Fischer et al., 2004) or are generally noisy (Faugeras et al., 2011). The less consistent response across trials in patients compared to controls can explain why the decoding performance can benefit from the implementation of a more complex algorithm i.e. with higher number of parameters than the logistic regression. Because of a high number of parameters to estimate, it is possible that the STTA captures better variability in comatose patients, but is not beneficial in a context where the response is more stable, as is likely the case for control subjects.

The application of the logistic regression for predicting the patients' outcome shows a similar tendency as previously demonstrated (Tzovara et al., 2013), but provides in general less accurate and not significant predictive power for awakening (Supplemental Material). These results show that the progression of auditory discrimination along the first two days of coma carries to some extent critical information about the patients' general condition although only an ad-hoc decoding algorithm choice can reveal the full potential of this progression in predicting patients' outcome. At present it remains unclear which specific features of the STTA make the positive predictive power so accurate in patients' outcome prediction. One possible explanation of these results is the flexibility in estimating the most discriminative latencies and voltage topographies along the whole epoch length which offers a more flexible framework for capturing discriminative effect than the logistic regression, at least in the implementation we presented here.

One main limitation of the STTA and STTA-W with respect to the logistic regression is the computational time. In particular, because the STTA algorithm in its current implementation is based on an iterative expectation-maximization algorithm (Dempster et al., 1977), it requires several minutes of computations for each mixture of Gaussian model. This computational constraint does not apply in the case of logistic regression, making it much more appealing for current applications in cognitive studies based on normal subjects. This is especially the case when the constrain of imposing a fixed length temporal window can be based on a priori knowledge of the typical length of EEG components, as for example in the case of healthy controls. Future studies could improve the computational time of the STTA algorithm by examining different strategies for selecting the model parameters based on complexity criteria. Such as for example the minimum descriptor length (Rissanen, 1978) or Akaike/Bayesian Information Criteria (Akaike, 1973; Schwarz, 1978). Another future direction worth investigating would be the implementation of our approach for EEG decoding directly in the source space instead of the sensor level. Such an implementation would be informative not only of the degree and time-intervals of auditory discrimination, but also of the brain regions which contribute to it. Indeed, other approaches for analysis of neuroimaging data, such as beamforming or general linear models, have been successfully applied in the source space to localize EEG or magnetoencephalography signals, in healthy and pathological populations (Brookes et al., 2004; Custo et al., 2014). The decoding results obtained with the STTA-W can be further improved by training the GMM models on each time-window as we did in the case of the logistic regression. Another promising direction of optimization is the employment of a regularization scheme in the estimation of the covariance matrices of the GMMs (Blankertz et al., 2011) which should be particularly beneficial in the case of high dimensional space, i.e. in the case of higher EEG density montage than what we implemented here.

Improving the accuracy of any multivariate decoding method is certainly a major challenge, but it is nevertheless central to emphasize that the focus of these methods in fundamental research is mainly their neurophysiologic interpretability. In this respect the STTA not only allows reconstructing the voltage topographies that are mostly discriminative but also offers an intuitive representation of the evoked activity along the response window based on posterior probabilities and strongly reflecting typical event-related potential components (De Lucia et al., 2012; Tzovara et al., 2012a,d). This is not only important for data visualization but also for making its use more compelling and interpretable to a large audience, including the cognitive neuroscience community particularly familiar with findings based on ERP components.

Acknowledgments

We thank Dr. Andrea O. Rossetti and Dr. Mauro Oddo for their exceptional clinical support. We also thank Christine for help in EEG recordings and Olivier Jaccard for technical support. This work was supported by the 'Service Projets et Organisation Stratégiques' of the University Hospital of Lausanne, Switzerland (project number 29062-1144 to M.D.L.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jneumeth.2014.10.019.

References

Akaike H. Information theory and an extension of the maximum likelihood principle; 1973. p. 267–81.
Bernasconi F, Lucia DM, Tzovara A, Manuel AL, Murray MM, Spierer L. Noise in brain activity engenders perception and influences discrimination sensitivity. J Neurosci 2011;31:17971–81.
Blankertz B, Lemm S, Treder M, Haufe S, Muller KR. Single-trial analysis and classification of ERP components – a tutorial. Neuroimage 2011;56:814–25.
Brandmeyer A, Sadakata M, Spyrou L, McQueen JM, Desain P. Decoding of single-trial auditory mismatch responses for online perceptual monitoring and neurofeedback. Front Neurosci 2013;7:265.
Brookes MJ, Gibson AM, Hall SD, Furlong PL, Barnes GR, Hillebrand A, et al. A general linear model for MEG beamformer imaging. Neuroimage 2004;23:936–46.
Brunet D, Murray MM, Michel CM. Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput Intell Neurosci 2011;2011, 813870.
Charles L, King JR, Dehaene S. Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. J Neurosci 2014;34:1158–70.
Conroy B, Sajda P. Fast, exact model selection and permutation testing for l-regularized logistic regression. JMLR Workshop Conf Proc 2012;22:246–54.

Cossy N, Tzovara A, Simonin A, Rossetti AO, De Lucia M. Robust discrimination between EEG responses to categories of environmental sounds in early coma. Front Psychol 2014;5:155.
Custo A, Vulliemoz S, Grouiller F, Van De Ville D, Michel C. EEG source imaging of brain states using spatiotemporal regression. Neuroimage 2014;96:106–16.
Das K, Giesbrecht B, Eckstein MP. Predicting variations of perceptual performance across individuals from neural activity using pattern classifiers. Neuroimage 2010;51:1425–37.
De Lucia M, Tzovara A, Bernasconi F, Spierer L, Murray MM. Auditory perceptual decision-making based on semantic categorization of environmental sounds. Neuroimage 2012;60:1704–15.
De Vos M, Thorne JD, Yovel G, Debener S. Let's face it, from trial to trial: comparing procedures for N170 single-trial estimation. Neuroimage 2012;63:1196–202.
Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm (with discussion). J R Stat Soc Ser B 1977;39:1–38.
Farquhar J, Hill NJ. Interactions between pre-processing and classification methods for event-related-potential classification: best-practice guidelines for brain-computer interfacing. Neuroinformatics 2013;11:175–92.
Faugeras F, Rohaut B, Weiss N, Bekinschtein TA, Galanaud D, Puybasset L, et al. Probing consciousness with event-related potentials in the vegetative state. Neurology 2011;77:264–8.
Fischer C, Morlet D, Luaute J. Sensory and cognitive evoked potentials in the prognosis of coma. Suppl Clin Neurophysiol 2004;57:656–61.
Green D, Swets J. Signal detection theory and psychophysics. New York: John Wiley and Sons Inc; 1966.
Haufe S, Meinecke F, Gorgen K, Dahne S, Haynes JD, Blankertz B, et al. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 2014;87:96–110.
Hausfeld L, De Martino F, Bonte M, Formisano E. Pattern analysis of EEG responses to speech and voice: influence of feature grouping. Neuroimage 2012;59:3641–51.
He BJ, Zempel JM. Average is optimal: an inverted-U relationship between trial-to-trial brain activity and behavioral performance. PLoS Comput Biol 2013;9:e1003348.
Kahnt T, Grueschow M, Speck O, Haynes JD. Perceptual learning and decision-making in human medial frontal cortex. Neuron 2011;70:549–59.
Lehmann D. Principles of spatial analysis. Amsterdam: Elsevier; 1987.
Lemm S, Blankertz B, Dickhaus T, Muller KR. Introduction to machine learning for brain imaging. Neuroimage 2011;56:387–99.
Millan J, Franze M, Mourino J, Cincotti F, Babiloni F. Relevant EEG features for the classification of spontaneous motor-related tasks. Biol Cybern 2002;86:89–95.
Muller KR, Tangermann M, Dornhege G, Krauledat M, Curio G, Blankertz B. Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. J Neurosci Methods 2008;167:82–90.
Oddo M, Schaller MD, Feihl F, Ribordy V, Liaudet L. From evidence to clinical practice: effective implementation of therapeutic hypothermia to improve patient outcome after cardiac arrest. Crit Care Med 2006;34(7):1865–73.
Parra L, Alvino C, Tang A, Pearlmutter B, Yeung N, Osman A, et al. Linear spatial integration for single-trial detection in encephalography. Neuroimage 2002;17:223–30.
Parra LC, Spence CD, Gerson AD, Sajda P. Recipes for the linear analysis of EEG. Neuroimage 2005;28:326–41.
Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 2009;45:S199–209.
Perrin F, Pernier J, Bertrand O, Giard M, Echallier J. Mapping of scalp potentials by surface spline interpolation. Electroencephalogr Clin Neurophysiol 1987;66:75–81.
Philiastides MG, Biele G, Vavatzanidis N, Kazzer P, Heekeren HR. Temporal dynamics of prediction error processing during reward-based decision making. Neuroimage 2010;53:221–32.
Ratcliff R, Philiastides MG, Sajda P. Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. Proc Natl Acad USA 2009;106:6539–44.
Rissanen J. Modeling by shortest data description. Automatica 1978;14:465–71.
Schwarz GE. Estimating the dimension of a model. Ann Stat 1978;6:461–4.
Taghizadeh-Sarabi M, Daliri MR, Niksirat KS. Decoding objects of basic categories from electroencephalographic signals using wavelet transform and support vector machines. Brain Topogr 2014.
Tzovara A, Murray MM, Michel CM, De Lucia M. A tutorial review of electrical neuroimaging from group-average to single-trial event-related potentials. Dev Neuropsychol 2012a;37:518–44.
Tzovara A, Murray M, Bourdaud N, Chavarriaga R, Millán JDR, Lucia MD. The timing of exploratory decision-making revealed by single-trial topographic EEG analyses 2012b;60:1959–69.
Tzovara A, Murray M, Plomp G, Herzog M, Michel C, Lucia MD. Decoding stimulus-related information from single-trial EEG responses based on voltage topographies 2012c;45:2109–22.
Tzovara A, Murray MM, Plomp G, Herzog MH, Michel CM, De Lucia M. Decoding stimulus-related information from single-trial EEG responses based on voltage topographies. Pattern Recognit 2012d;45:2109–22.
Tzovara A, Rossetti AO, Spierer L, Grivel J, Murray MM, Oddo M, et al. Progression of auditory discrimination based on neural decoding predicts awakening from coma. Brain 2013;136:81–9.
van Gerven M, Hesse C, Jensen O, Heskes T. Interpreting single trial data using groupwise regularisation. Neuroimage 2009;46:665–76.
Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. Clin Neurophysiol 2002;113:767–91.
Wyart V, de Gardelle V, Scholl J, Summerfield C. Rhythmic fluctuations in evidence accumulation during decision making in the human brain. Neuron 2012;76:847–58.