NeuroImage 59 (2012) 3641–3651

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Pattern analysis of EEG responses to speech and voice: Influence of feature grouping

Lars Hausfeld [a,b,*], Federico De Martino [a,b,c], Milene Bonte [a,b], Elia Formisano [a,b]

[a] Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands
[b] Maastricht Brain Imaging Center, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands
[c] Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Minneapolis, USA

ARTICLE INFO

Article history:
Received 25 July 2011
Revised 24 October 2011
Accepted 16 November 2011
Available online 30 November 2011

Keywords:
EEG
ERP
Machine learning
Classification
Audition
Speech

ABSTRACT

Pattern recognition algorithms are becoming increasingly used in functional neuroimaging. These algorithms exploit information contained in temporal, spatial, or spatio-temporal patterns of independent variables (features) to detect subtle but reliable differences between brain responses to external stimuli or internal brain states. When applied to the analysis of electroencephalography (EEG) or magnetoencephalography (MEG) data, a choice needs to be made on how the input features to the algorithm are obtained from the signal amplitudes measured at the various channels. In this article, we consider six types of pattern analyses deriving from the combination of three types of feature selection in the temporal domain (*predefined windows*, *shifting window*, *whole trial)* with two approaches to handle the channel dimension (*channel wise*, *multi-channel*). We combined these different types of analyses with a Gaussian Naïve Bayes classifier and analyzed a multi-subject EEG data set from a study aimed at understanding the task dependence of the cortical mechanisms for encoding speaker's identity and speech content (vowels) from short speech utterances (Bonte, Valente, & Formisano, 2009). Outcomes of the analyses showed that different grouping of available features helps highlighting complementary (i.e. temporal, topographic) aspects of information content in the data. A *shifting window/multi-channel* approach proved especially valuable in tracing both the early build up of neural information reflecting speaker or vowel identity and the late and task-dependent maintenance of relevant information reflecting the performance of a working memory task. Because it exploits the high temporal resolution of EEG (and MEG), such a shifting window approach with sequential multi-channel classifications seems the most appropriate choice for tracing the temporal profile of neural information processing.

Introduction

Electroencephalography (EEG) and magnetoencephalography (MEG) are commonly used to study the time course of neural information processing in the human brain with high temporal resolution. In most cases, EEG/MEG studies rely on the comparison of averaged responses to repeated presentations of experimental conditions either in the temporal domain (event-related potentials [ERPs] or fields [ERFs], respectively) and/or in the frequency domain (event-related desynchronization and synchronization) (Pfurtscheller and Lopes Da Silva, 1999). Often, the statistical analyses (and related inferences on neural processing) are limited to a-priori specified (spectro-) temporal windows of interest – at channel or estimated source level – and therefore only a small subset of the measured signal is actually utilized.

This article illustrates several approaches to EEG data analysis based on *pattern recognition* (e.g. Bishop, 2007; Duda et al., 2001). In contrast to the conventional approach where a single dependent variable is examined (univariate statistics), these techniques exploit the information content in patterns of dependent variables (features), which are extracted from the measured signals. Pattern recognition allows analyzing EEG data in a more exploratory and data-driven manner and – similar to the recent developments in fMRI (e.g. Haynes and Rees, 2006) – promises to complement conventional approaches for EEG/MEG analysis.

A typical application of pattern recognition methods includes three steps, (1) extracting and selecting features (i.e. dependent variables), (2) learning a model with a machine-learning algorithm, and (3) determining the generalization ability of the learnt model using an independent evaluation dataset. In EEG/MEG, various *types* of features can be considered, ranging from signal amplitude in the temporal domain (e.g. Rieger et al., 2008) to power or phase information in the frequency domain (Kerlin et al., 2010; Luo and Poeppel, 2007; Rieger et al., 2008). Specific transformations, such as wavelet coefficients (Åberg and Wessberg, 2007; Rieger et al., 2008), and coherence measures (Besserve et al., 2007) can also be used. Furthermore,

* Corresponding author at: Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Fax: +31 43 3884125.
E-mail addresses: lars.hausfeld@maastrichtuniversity.nl (L. Hausfeld), f.demartino@maastrichtuniversity.nl (F. De Martino), m.bonte@maastrichtuniversity.nl (M. Bonte), e.formisano@maastrichtuniversity.nl (E. Formisano).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.11.056

features can be differently *grouped* in the (spectral-) temporal and spatial domain. For example, limiting the information to pre-defined temporal windows of interest is essential to many realizations of EEG-based *brain-computer interface* (BCI) systems (e.g. Birbaumer, 2006; Blankertz et al., 2011; Wolpaw et al., 2002). Alternatively, the information contained in a sliding time interval of EEG data can be used, e.g. to detect the occurrence of seizures in epileptic subjects (Schad et al., 2008). Concerning the spatial (channel) domain, many BCI systems employed spatial filters (i.e. linear combinations of channels; see Blankertz et al., 2011) to enhance performances. For the same reason sophisticated feature selection or reduction methods were applied in BCI systems (see Bashashati et al., 2007).

Several machine-learning algorithms have been used to learn the relation between selected features of the EEG/MEG data and experimental labels. These algorithms include simple correlation (e.g. Luo and Poeppel, 2007), support vector machines (SVMs) (Vapnik, 1995), linear discriminant analysis (LDA) (e.g. Duda et al., 2001), and neural networks or Bayesian approaches (Bishop, 2007). Most frequently, learning algorithms are based upon linear models (e.g. Lotte et al., 2007; Rieger et al., 2008; van Gerven et al., 2009) due to their fast computation, robustness and simplicity of results interpretation.

To determine the generalization ability of the computed model, an independent set of test data is required. This can be done at single-subject level, splitting the measured data into training and testing sets (e.g. Luo and Poeppel, 2007) or across subjects, using a subset of subjects for training and the other for evaluating the generalization performance (e.g. Kerlin et al., 2010).

In this study, we consider and evaluate the effects of differently combining and grouping the features in the temporal (*predefined windows*, *shifting window*, *whole trial*) and channel domain (*single channel*, *multichannel*) in the context of a neuro-cognitive EEG paradigm. Using *Gaussian Naïve Bayes* (GNB; Mitchell, 1997) classification, we analyze data from an auditory EEG study aimed at understanding the task dependence of the cortical mechanisms underlying the processing of voice and speech identification (Bonte et al., 2009) and illustrate the results of each possible feature combination in the temporal and channel domain.

Materials and methods

Machine-learning approaches for the analysis of neuroimaging data require single trials to be described by an n-dimensional vector of features. In our approach, basic features are defined as EEG voltages and include time (samples) and measurement channels (electrodes). In particular, we consider six types of classification analyses derived from combining three types of features grouping in the temporal domain (*predefined windows*, *shifting windows*, *whole trial)* with two approaches to handle the channel dimensions (*single channel*, *multichannel*, see Fig. 1). These different types of analyses can be combined with any classification algorithm (e.g. LDA classifier or SVMs). Here, we use a modified *Gaussian Naïve Bayes classification,* because of its simplicity which implies lower computational costs (e.g. compared to SVM classification) and interpretability of model parameters. We examine the case of pairwise classifications of EEG responses to simple vowels (/a/, /i/, /u/) spoken by three speakers (sp1, sp2, sp3) (see EEG experiment and data section).

Predefined windows

In the first approach, we use prior hypotheses (e.g. typical ERP windows) to select the temporal windows entering the analysis. As depicted in Fig. 1.a, the temporal samples within a specific interval are used as features to classify single trials either for each of the K channels (right upper panel) or for all channels (right lower panel). In the latter case, the feature set is defined by concatenating sampling points of multiple channels. In the case of a channel-by-channel analysis accuracy values are obtained for each electrode. This allows creating a topographic map of classification performance for the predefined intervals. Classifying based on features from multiple channels results in one classification accuracy value. In this case, a topographic map is created from the weights estimated during model training (see Eq. (4)) that indicate the relevance of each electrode contribution to the classification.

Shifting windows

In the second approach (Fig. 1b), the analyses are not restricted to specific latencies and are based upon features from *shifting windows* either on a channel-by-channel basis (right upper panel) or by concatenating features from multiple channels (right lower panel). Results of the single-channel approach can be depicted as a time series of topographic plots indicating classification performance.

The multi-channel classification allows retrieving the information content over time (information time-course). A weight vector – indicating the relevance of individual channels – is obtained for each time window.

Whole trial period

In the third temporal approach (Fig. 1.c) all temporal samples within a trial period are used. Classifications are performed either using the channel-wise (right upper panel) or multi-channel (right lower panel) approach. Results for the channel-wise approach may be used to create a topographic map of the information content within the entire trial period. For the multichannel approach, the analysis returns an overall accuracy value. Weights are defined for each sampling point and channel and thus indicate the temporal and topographical variation of the information content.

Gaussian Naïve Bayes classification

We report below a short description of GNB classification with reference to EEG data; see Mitchell (1997), for a more complete and general formulation of this algorithm.

Let us consider a supervised learning problem in which we wish to approximate the function $f:X\rightarrow C$ or equivalently $P(C|X)$, where C is a Boolean random variable representing the categories in our classification problem and $X=\langle x_1,\ldots,x_n\rangle$ is a n-dimensional feature vector obtained from the EEG data. Using Bayes rule we can write:

$$P(C=c_m|X)=\frac{P(X|C=c_m)P(C=c_m)}{\sum_j P(X|C=c_j)P\left(C=c_j\right)} \tag{1}$$

where c_m represents the mth category. One way to learn $P(C|X)$ is to use the training data to estimate $P(X|C)$ and $P(C)$ and then use Eq. (1) to classify any new instance of X.

The Naïve term is introduced when in the estimation of $P(X|C)$ the n features are assumed to be conditionally independent and Eq. (1) can be written as:

$$P(C=c_m|X)=\frac{\prod_{i=1}^{n} P(x_i|C=c_m)P(C=c_m)}{\sum_j \prod_{i=1}^{n} P(x_i|C=c_j)P\left(C=c_j\right)}. \tag{2}$$

Following Eq. (2) and having estimated $P(x_i|C)$ and $P(C)$ from the training data, any new EEG trial $Y_{new}=\langle y_1,\ldots,y_n\rangle$ can be classified following:

$$C\leftarrow\arg\max_{C_m} P(C=c_m)\prod_{i=1}^{n} P(y_i|C=c_m), \tag{3}$$