Handbook of Clinical Neurology, Vol. 168 (3rd series)
Brain-Computer Interfaces
N.F. Ramsey and J. del R. Millán, Editors
https://doi.org/10.1016/B978-0-444-63934-9.00023-8

Chapter 23

General principles of machine learning for brain-computer interfacing

IÑAKI ITURRATE[1†], RICARDO CHAVARRIAGA[1,2*†], AND JOSÉ DEL R. MILLÁN[3,4]

[1]*Center for Neuroprosthetics, École Polytechnique Fédérale de Lausanne, Geneva, Switzerland*

[2]*Institute of Applied Information Technology (InIT), Zurich University of Applied Sciences ZHAW, Winterthur, Switzerland*

[3]*Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, United States*

[4]*Department of Neurology, The University of Texas at Austin, Austin, TX, United States*

Abstract

Brain-computer interfaces (BCIs) are systems that translate brain activity patterns into commands that can be executed by an artificial device. This enables the possibility of controlling devices such as a prosthetic arm or exoskeleton, a wheelchair, typewriting applications, or games directly by modulating our brain activity. For this purpose, BCI systems rely on signal processing and machine learning algorithms to decode the brain activity. This chapter provides an overview of the main steps required to do such a process, including signal preprocessing, feature extraction and selection, and decoding. Given the large amount of possible methods that can be used for these processes, a comprehensive review of them is beyond the scope of this chapter, and it is focused instead on the general principles that should be taken into account, as well as discussing good practices on how these methods should be applied and evaluated for proper design of reliable BCI systems.

INTRODUCTION

Brain-computer interfaces (BCIs) can be seen as systems that translate brain activity patterns into commands that can be executed by an artificial device. This enables the possibility of controlling devices such as a prosthetic arm or exoskeleton (Hochberg et al., 2012; Collinger et al., 2013; Bouton et al., 2016; Soekadar et al., 2016; Lee et al., 2017), a wheelchair (Carlson and Millán, 2013; Leeb et al., 2015; Fernández-Rodríguez et al., 2016), typewriting applications (Spüler et al., 2012a, b; Sellers et al., 2014; Jarosiewicz et al., 2015), or games (Leeb et al., 2013a; Marshall et al., 2013) directly by modulating our brain activity.

For this purpose, BCI systems rely on signal processing and machine learning algorithms to decode the brain activity. In the general sense this process can be seen as applying a mathematical function,

$$\mathbf{y} = f(\mathbf{X}^r, \theta) \tag{23.1}$$

that takes the measured brain activity $\mathbf{X}^r \in \mathbb{R}^{N \times T}$ (with N channels and T samples) and, given a set of parameters θ, produces an output $\mathbf{y}$, corresponding to the inferred command; e.g., opening the hand of a prosthetic hand, steering a wheelchair toward one direction, or selecting a specific character.

Brain activity patterns, irrespective of the recording technology, are generated by multiple, concurrent mental

[†]These authors shared the first authorship

*Correspondence to: Ricardo Chavarriaga, Ph.D., Institute of Applied Information Technology (InIT), Zurich University of Applied Sciences ZHAW, Winterthur, Switzerland. Tel: +41-58-934-69-60, Fax: +41-58-934-69-60, E-mail: r_chavarriaga@ieee.org

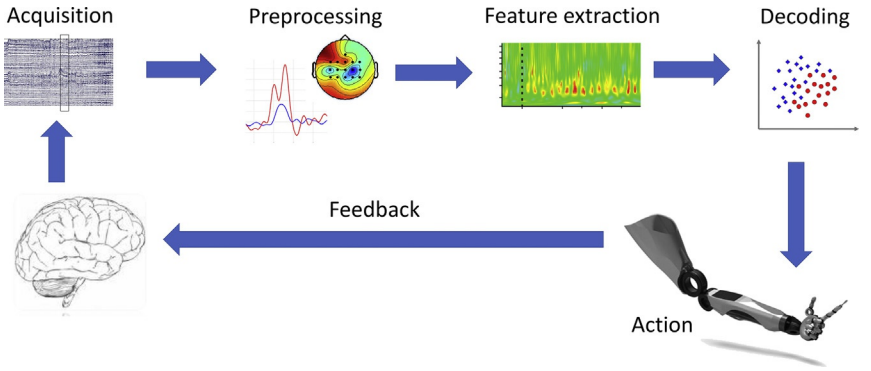

Fig. 23.1. Processing steps of a brain-computer interface. The acquired signal is preprocessed through both spatial and spectral filters. Then discriminant features are extracted and used as inputs to a decoder that yields an output command that is sent to an external device to be executed. BCIs are by definition a closed-loop system as executed actions are perceived by the subjects through their own senses or explicit feedback.

processes. However, not all of these processes are necessarily relevant for the use of the BCI. Furthermore, the recorded signals also contain components that have not directly originated from the brain such as muscular or movement artifacts, as well as electromagnetic noise. Therefore, the signal is typically processed to extract signal components that are correlated to the mental processes of interest. These components are referred to as *features*. The processing steps of a BCI system are illustrated in Fig. 23.1: the raw brain signals ($\mathbf{X}^r$) are preprocessed and a vector of selected features ($\mathbf{x}$) is extracted and used as input to the decoding algorithm. Thus Eq. (23.2) can be rewritten as:

$$\mathbf{y} = f(\mathbf{x}, \theta) \tag{23.2}$$

where $\mathbf{x} = f_p(\mathbf{X}^r)$ are the features extracted from the original signal. This output $\mathbf{y}$ will be sent to the device to be controlled to perform an action. The BCI loop is then closed when the user receives feedback of this action directly through their natural senses or by mediating a feedback actuator (e.g., computer display, auditory or tactile stimuli, neurostimulation).

Depending on the application, the decoded output can be either a continuous value, e.g., hand velocity (Carmena et al., 2003); or a discrete command, e.g., decoding motor imagery of left vs right hand (Leeb et al., 2013b). In the machine learning literature, these approaches are referred to as *regression* and *classification*, respectively. In the case of a classification approach, the function f defines a boundary that segments the set of possible values of $\mathbf{x}$ (i.e., the *feature space*) into subspaces that correspond to different output categories, commonly referred to as *classes*. In other words, $\mathbf{y}$ will take a different value for each of these regions, corresponding to the decoded class. In the case of regression approaches, the value $\mathbf{y}$ will correspond to the estimated BCI output.

This chapter reviews the main steps that need to be taken into account in the choice and evaluation of the feature selection (Section "Feature Extraction and Selection") and decoding methods (Section "Decoding Brain Activity") for BCI applications. Rather than entering into the details of specific machine learning methods, it is aimed at providing the basic concepts and guidelines to consider at different stages of the design process. For illustrative purposes, in this chapter we mainly discuss machine learning approaches used to decode direct measurements of electrical activity of the brain-spiking firing rates of neuronal populations (single/multiunit activity—SUA/MUA), local-field potentials (LFP), electrocorticography (ECoG), and electroencephalography (EEG)—since they are currently the most common approaches used for BCI. Nonetheless, most of the discussed principles are generalizable to other recording techniques.

FEATURE EXTRACTION AND SELECTION

Recording of neural activity, irrespectively of the chosen technique, will yield signals that are composed of a mixture of multiple processes. While some of these components are inherently related to the BCI purpose, others are not of interest and should be discarded. Therefore, from the raw recorded signal we need to extract only these parts that are likely to be informative of neural processes (i.e., features) and filter out other irrelevant components. In this section we describe the most common methods used to define and extract these features.

Feature extraction is understood as the process of translating neural signals into a subspace that represents the correlates to be used to control the device. In both invasive and noninvasive recording modalities, the most common approach to treat the signals is as a temporal series. As such, the feature extraction process follows the same rationale irrespective of the technique used to measure brain activity, comprising electrical activity, e.g.,

SUA/MUA, LFP, ECoG, EEG (Millán and Carmena, 2010; Torres Valderrama et al., 2010; Bundy et al., 2014), as well as magnetic and optical techniques, e.g., magnetoencephalography (Waldert et al., 2008) or functional near-infrared spectroscopy, respectively (Chaudhary et al., 2017). Usually, the feature extraction process encompasses two different phases: (i) signal preprocessing and (ii) feature extraction.

Phase 1: Signal preprocessing

We define preprocessing as the steps that are usually performed on the signal with the attempt to improve its quality and increase its signal-to-noise ratio (SNR).[a] The preprocessing phase can cover several techniques, or can be skipped entirely. In general terms, however, three steps are usually performed: artifact detection and removal, frequency filtering, and spatial filtering.

Artifact detection and removal

Artifact activity is defined as components of the recorded signal that are not of cerebral origin. To a greater or lesser extent, this activity may be present with any recording modality, although their nature may differ. Artifacts can be either generated from external or internal factors. For instance, EEG recordings are prone to contamination due to electromagnetic fields generated by the power line. Internal noise sources can also affect the quality of the recorded signal. Electrical and magnetic fields generated by muscle (electromyographic) and eye movements (electrooculographic) can contaminate EEG signals; similarly, head movements within the scanner and vascular activity affect fMRI recordings (Friston et al., 1996; Diedrichsen and Shadmehr, 2005). The presence of this activity poses an important problem. In the case of EEG, they are typically several orders of magnitude bigger than the neural signals. As a result, artifacts can limit the usability of data used to train a decoder, and also affect the reliability of the BCI in online closed-loop conditions.

Thus, it is of key importance to deal with this spurious activity. The most common approach in the presence of artifacts is rejecting the contaminated portions of data to avoid erroneous decoding. Depending on the artifact source, the rejection method may also vary. However, the two most common ones are rejection via threshold or via z-scores.

Threshold rejection: With this method, the portion of data is removed when it surpasses a specific threshold in the channel of interest, τ_c. This threshold is usually defined empirically.

Z-score artifact rejection: Threshold rejection is a simplified version of removing artifacts based on the z-score of the signal. In this approach, we remove portions of data that are bigger than a specific z value. First, the z-score of the data is computed as,

$$z_c = \frac{x_c - \mu_{x_c}}{\sigma_{x_c}}, \tag{23.3}$$

where x_c is the signal recorded from channel c, and μ_{x_c} and σ_{x_c} the sample mean and standard deviations of the signal. The z-score represents the distance between the data and the mean in units of standard deviation. In practice, data are considered artifactual when z exceeds a certain value, usually $z > 3$.

Other approaches are based on making an explicit model on how the artifact affects the recorded signal. Examples of this approach include models of the contamination induced in EEG by eye movements (Schlögl et al., 2007a), or the effects of deep brain stimulation in recording electrodes (Rossi et al., 2007). Alternatively, artifact removal can be also addressed by filtering as we will see in the next subsection. However, notice that these methods can present limitations, such as the need for training with contaminated data or their inability to remove completely the artifactual activity. A thorough review of methods for artifact filtering or removal is outside of the scope of this chapter. Interested readers are referred to Urigüen and Garcia-Zapirain (2015), Reis et al. (2014), Castermans et al. (2014), Gwin et al. (2010), and Rossi et al. (2007).

Frequency filtering

Neural activity is typically characterized in terms of its oscillatory components. Accordingly, the second step of signal preprocessing usually involves applying a broad bandpass filter to attenuate the power of undesired frequencies. Frequency filters are classified as finite impulse response (FIR) and infinite impulse response (IIR) filters. Although FIR filters provide better attenuation capabilities, IIR filters (such as Butterworth or Chebyshev) are usually preferred due to a less associated computational power. In this preprocessing step, when analyzing oscillatory components of electrical brain activity (EEG, ECoG or LFPs), the bandpass filters applied usually ranges from very low (0.1 or 1 Hz) to relatively high frequencies (40 or 100 Hz for scalp EEG; 100 or 300 Hz for intracranial recordings). Spiking activity, on the other hand, usually requires bandpass filtering at a high frequency range, e.g., from 300 to

[a]The signal-to-noise ratio (SNR) is a term used in signal processing to refer to the level of a wanted signal with respect to the level of background noise observed in a given measurement. In the context of BCI, the wanted signal is the neural correlate of the process we are interested in, and the noise correspond to other elements captured by the sensors used (e.g., muscular activity, electromagnetic noise).

3000 Hz, since the focus is on the fast activity changes of the action potentials (Quian Quiroga, 2007; Quiroga and Panzeri, 2009; Kao et al., 2014).

An important aspect to take into account when designing frequency filters is the signal distortion. In particular, designed filters can provoke undesired phase shifts in the signal, leading to misinterpretations of the results obtained, or even create spurious components on the signal of interest. For more information on this topic readers are referred to Rousselet (2012), Widmann and Schröger (2012), Acunzo et al. (2012), and VanRullen (2011). As a related issue, it is important to note that the choice of zero-phase filters, or noncausal filters, i.e., filters that create no distortions (Yael et al., 2018), is not advised in the case of targeting closed-loop applications, as these filters cannot be applied online (Sani et al., 2016).

Spatial filtering

Spatial filters are defined as a linear transformation in the channel space with the objective of improving the sensitivity of specific channels or brain regions, or the removal of artifacts. The transformation is defined as:

$$X^f = WX^o, \tag{23.4}$$

where $X^o \in \mathbb{R}^{N\times T}$ is the unfiltered recorded data with N channels and T samples, $W \in \mathbb{R}^{M\times N}$ the spatial decomposition matrix, and $X^f \in \mathbb{R}^{M\times T}$ is the filtered data. It is important to remark that the number of dimensions M of the projected data X^f, does not necessarily match the number of original channels N.

Spatial filters play a key role in current BCI applications, thanks to their potential in increasing the SNR of the neural correlate of interest. Depending on the targeted application, spatial filters can be grouped into three different types: rereferencing, data projection, or discriminative filters.

- *Rereferencing*: Recordings of electrical brain activity are in reality measures of the electrical potential between two points in the human body (the electrode of interest and another location chosen as reference). The choice of the reference can thus affect the quality of the recorded signal. In consequence, performing a rereferencing has been shown to increase the SNR and boost BCI performances (McFarland et al., 1997; Cohen, 2015). Common average reference (CAR) (Bertrand et al., 1985), or Laplacian transformations are among the most common rereferencing processes. Examples of the referencing effect in EEG signals are shown in Fig. 23.2.
- *Data projection filters*: The objective of these filters is to project the data into a different dimension (lower or higher). The most common data projection spatial filter is the principal component analysis (PCA) (Lagerlund et al., 1997; Pourtois et al., 2008), that projects the data to a different space based on the variance of the data, under the assumption that the input data follows a normal distribution. A subset of the dimensions in the projected space is then selected for further processing. This specific case of data projection is denominated as "dimensionality reduction," i.e., $M<N$. On the other hand, it is possible to project the data onto a higher dimensional space via spatial filters. Inverse methods, used to estimate intracranial sources from noninvasive recordings, usually follow this approach, where $M\gg N$ (Cincotti et al., 2008; Edelman et al., 2015).

  Another type of projection filter is independent component analysis (ICA). In ICA, the main assumption is that data do not follow a Gaussian distribution, and the projection onto the new space is done by maximizing the independence across dimensions.[b] ICA has been demonstrated to be a successful tool for the removal of electro-oculographic and muscular artifacts, reduce the data dimensionality, as well as a neuroscientific tool to study isolated neurophysiologic processes (Hyvärinen et al., 2004; Makeig et al., 2004; Debener et al., 2005; Delorme et al., 2012; Bigdely-Shamlo et al., 2013).
- *Discriminative filters*: The main objective of these filters is to boost the SNR of the signal by using labeled data. These types of filters can be used with any time-series signals, i.e., EEG, ECoG, LFPs, or firing rates. By exploiting the additional information provided by the labeled data, they try to maximize the differences among classes by linear projections (Parra et al., 2008). The most common discriminative filters are the common spatial patterns, which target maximizing the variance between the two classes (Grosse-Wentrup and Buss, 2008).

Other spatial filters that have been used for BCI applications include canonical correlation analysis (Hardoon et al., 2004; Spüler et al., 2014), xDAWN (Rivet et al., 2009), or optimized spatial filter (Boye et al., 2008) among others. Please refer to the publications for more details.

[b] In statistics, independence is defined as $p(B|A) = p(B)$, i.e., dimension A does not convey any information about dimension B.

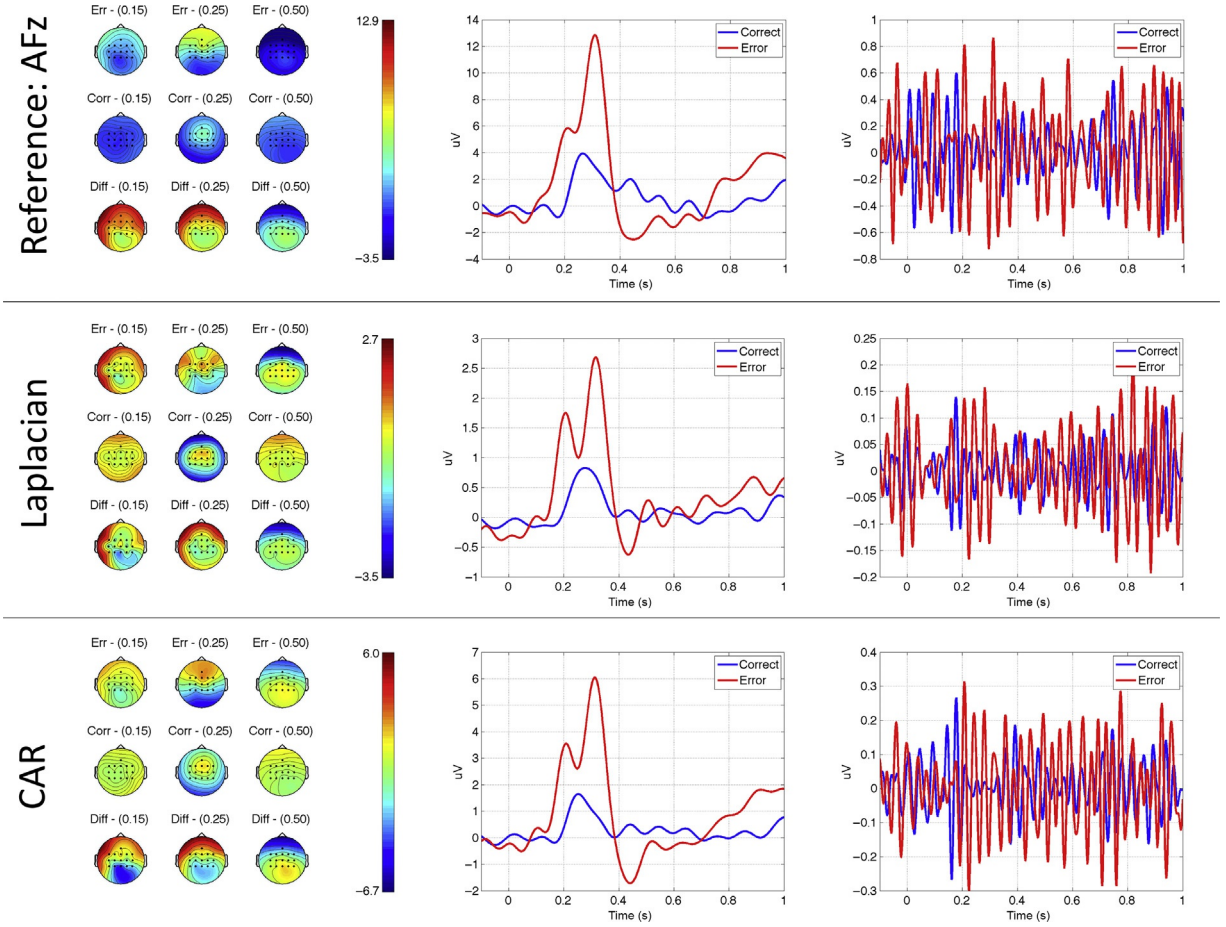

Fig. 23.2. Effect of preprocessing parameters on the error-related EEG potential. The experimental protocol is described in Iturrate et al. (2014). The figure shows the average ERP for one experimental run including 129 trials of correct class (*blue line*) and 33 trials of the error class (*red line*). Three types of rereferencing process are shown. For the left and middle columns the signal is filtered in the [1–10] Hz band, in the rightmost column the signal is filtered in the band [20–40] Hz. This shows how the processing parameter can affect the signal waveform and amplitude, as well as its suitability to provide information for the BCI decoding process. (*Left*) Topographical representation of the recorded patterns at three different time points (0.15, 0.25, and 0.5 s). In each case it shows from top to bottom the average activity for the error and correct classes, as well as their difference (Diff = error − correct). Electrical activity (in uV) at each location is represented by colors from *blue* to *red* (negative to positive values, respectively). Middle Grand average event-related potential (ERP) of the signal at the FCz electrode. $t = 0$ corresponds to the stimulus onset for the signal in the [1–10] Hz band. (*Right*) Grand average ERP in the [20–40] Hz band. The difference between the ERPs in the two frequency bands highlight the impact the preprocessing steps may have in the possibility of reliably decoding a given mental process.

Phase 2: Feature extraction

Once the signal is preprocessed, the next phase involves the extraction of features. The first usual step is the selection of channels that will be used for the feature extraction itself. In the case of a continuous data series (EEG, ECoG, LFP) we mainly distinguish two types of features, temporal and spectral, which will be decided based on the neural process of interest to be decoded. The analysis of spike activity from MUA/SUA, on the other hand, is a particular case since the signal of interest is not inherently a continuous data series, but rather a series of discrete events (i.e., the action potentials).

Spatial features and channel selection

Electrophysiologic recordings are most commonly recorded from several electrodes (namely channels) concurrently. As such, it is important first to choose which of these channels are used for subsequent extraction methods. In this regard, we have two different scenarios. First, we can choose a subset of channels based on prior neurophysiologic knowledge (e.g., motor execution correlates are most likely to be discriminant in channels placed over the motor cortex) or based on data driven, channel scores (see section "Feature selection and normalization"). Alternatively, discriminative spatial

filters (explained in the previous section) provide a spatial pattern that can be directly used as a feature for the BCI.

Temporal features for continuous data series

Temporal features are those that use time points from the filtered neural signals. Their most common use is for processes that are time-locked to an event, e.g., event-related potentials (ERPs). As such, the features are usually selected from specific time windows relative to the event occurrence. Common examples in the case of EEG-based BCIs are the P300 evoked response (Krusienski et al., 2006; Guger et al., 2009) or error-related potentials (ErrP; Chavarriaga et al., 2014), where the features are often time samples of the activity within a time window of 800 ms poststimulus from parieto–occipital or fronto–central regions, respectively. Another example is when decoding motor correlates such as movement initiation (Niazi et al., 2011; Lew et al., 2014; López-Larraz et al., 2014; Schwarz et al., 2017; Pereira et al., 2018) or grasping correlates (Agashe et al., 2015; Randazzo et al., 2015; Jochumsen et al., 2016; Iturrate et al., 2018; Sburlea and Müller-Putz, 2018).

A common process when dealing with temporal features is to apply a bandpass filter that isolates the process of interest (e.g., 1–10 Hz for P300 or ErrP; or 0.1–5 Hz for movement-related cortical potentials), and downsample the resulting signal to reduce the final number of features (e.g., from an original sampling rate of 512 Hz to 64 Hz) (see Fig. 23.2).

Spectral features for continuous data series

Spectral features encompass those based on the power of specific frequency bands and channels. The computation of this power is done by means of the estimation of the power spectral densities (PSDs), with the periodogram and Welch's modified periodogram among the most typical methods used for such estimations. Despite the existence of other methods (e.g., wavelets), periodogram-based PSDs are the most used ones due to their simplicity and low computational cost, which make them usable under closed-loop conditions. Interested readers are referred to Cohen (2014) for a more thorough explanation of these and other alternatives. A classical example where spectral features are exploited is the decoding of voluntary modulation of sensorymotor rhythms, e.g., when the user performs motor imagery (Pfurtscheller and Lopes da Silva, 1999). In this case, decoding is often based on the signal power of μ (8–12 Hz) and/or β (13–32 Hz) frequency bands from electrodes over the motor cortex. The benefits of including spectral features when decoding ERPs (Omedes et al., 2013) is another example. In invasive recordings, γ activity (>40 Hz) is usually exploited thanks to the higher SNR of these techniques, for instance, to decode motor correlates (Kubanek et al., 2009; Miller et al., 2009; Pistohl et al., 2012; Vansteensel et al., 2016; Branco et al., 2017) or cognitive processes such as performed and imagined speech (Pasley et al., 2012; Martin et al., 2016, 2018).

As in the case of temporal features, additional signal preprocessing steps can be applied when using spectral features. In particular, the use of a high-pass filter (around 0.5 Hz) prior to the estimation of PSDs is a common practice to avoid numerical problems when computing the spectral components.

Spike extraction and sorting from intracortical MUA/SUA recordings

Processing of the activity of single or multiple neurons is usually composed of three steps. First, as mentioned in Section "Frequency filtering," a bandpass filter is needed to isolate high frequency components of the recorded signal (i.e., abrupt changes in magnitude) that correspond to individual spikes (Quiroga and Panzeri, 2009; Kao et al., 2014).

Second, since the bandpass-filtered signal is composed of individual spikes generated by different neurons near the electrode, we need to identify which of these spikes are generated by which neurons, a process termed "spike sorting" (Nicolelis et al., 2003; Quian Quiroga, 2007). Following the assumption that spikes from a given neuron will tend to have a particular shape, sorting the extracted spikes by their shape will provide a means to identify activity generated by different neurons. This sorting process is performed by means of clustering methods, unsupervised learning techniques that identify groups that contain patterns that are alike among themselves and different from those of other clusters. For this purpose, the shape of individual spikes is described in terms of characteristics features, typically extracted using PCA decomposition or wavelets, before proceeding to the clustering process. Importantly, this clustering process not only yields a way to identify individual neurons on already recorded data, but also provides a template to identify the source of spikes for subsequent recordings. For online operation, once the new spikes have been observed, the cluster that better corresponds to its shape will indicate the most likely neuron that has generated it.

Finally, the particular activity of a specific neuron is typically represented in terms of its firing rate, corresponding to the number of spikes emitted in a nonoverlapping time window. These windows typically last a

few tens of milliseconds. The use of firing rates yields a time-series signal that is then used as a feature to decode the variables of interest.

DECODING BRAIN ACTIVITY

To decode neural activity we apply a given mathematical model that, given the features as input, provides an output corresponding to the inferred neural process that generated that activity. Model calibration—i.e., finding the parameters that yield the best performance—is done based on examples of how the brain activity looks like when a given mental process is performed. Prior to BCI use, multiple models can be calibrated and evaluated to select the most suitable one to be used in a closed-loop setting to provide online decoding.

As mentioned before, the decoding process can be seen as a mathematical function $\mathbf{y} = f(\mathbf{x},\theta)$ (Eq. 23.2), that maps the features $\mathbf{x}$ onto a BCI output given a set of parameters θ. For both classification and regression, the set of parameters θ should be optimized so as to obtain the desired behavior. In this chapter, we focus on good practices on how these methods should be applied and evaluated in the general sense.

Calibration of the model first requires selection of the features, which are the most informative about the process of interest, then selection of a decoding method, and tuning of its parameters. This process requires evaluating multiple alternatives and selection of the one most likely to succeed based on their performance.

Feature selection and normalization

One of the main characteristics of brain activity is that it is largely subject dependent. On top of that, feature extraction steps usually lead to a large number of features. As such, measures of features discriminability are needed to derive a subset of features that are valid for a specific user. This method, combined with a final manual selection of features by the experimenter, will lead to the set of features that will be used during the closed-loop BCI application. The feature selection procedure varies depending on whether we are on a classification or regression scenario (see next section).

In the most common case of a two-class classification scenario, the classical discriminability measure is the Fisher score (Duda et al., 2000). The Fisher score for a feature f, $FS(f)$ is computed as follows:

$$\mathrm{FS}(f) = \frac{|\mu_1 - \mu_2|}{\sigma_1 + \sigma_2} \in (-\infty, \infty) \tag{23.5}$$

where μ_i and σ_i represent the mean and standard deviation computed over the training set, and $i_{1,2}$ the class estimated from the labels vector $y_{1,2}$. Such a metric can be easily extended to multiclass scenarios by performing the individual scores of pairs of classes.

Alternatively, the squared point biserial correlation, r^2 is also commonly used in the case of two-class classification BCIs, as well as in regression scenarios. The main advantage of r^2 with respect to the Fisher score is that it provides bounded values, and can be computed as follows:

$$r^2(f) = corr\left(x_f, y\right)^2 \in [0, 1], \tag{23.6}$$

where *corr* represents the linear correlation function, x_f represents the feature values over the training set, and $y \in \{1, 2\}$ the labels vector or $y \in \mathbb{R}$ in the case of regression scenarios.

Once the features have been selected, a final step before inputting them into a classifier is that of performing normalization. Feature normalization or scaling is the process of making the features fall within the same range. This process is not only useful in the case of having features with different ranges, but also in case of using concepts such as optimization, regularization, or kernel-based classifiers. The two most common normalization methods are min–max, computed as follows:

$$f_i = \frac{f_i - \min(f)}{\max(f) - \min(f)} \tag{23.7}$$

where $\max(f)$ and $\min(f)$ are the maximum and minimum values of feature f_i computed from the training set; and z-score normalization:

$$f_i = \frac{f_i - \mu_f}{\sigma_f} \tag{23.8}$$

It should be noticed that some decoding algorithms, namely deep learning, rely on the idea of avoiding explicit feature selection and allow the algorithm to find the relation among available features and the intended output values by itself. This has been particularly successful when there are complex relations across features and large amounts of data are available making it difficult to find these relations by mathematical methods.

Decoder calibration

Parameters θ are selected through an optimization process referred to as "training" or "calibration" of the decoder. In the case of classification, this yields a function f that divides the feature space into multiple regions, each corresponding to a different class. Fig. 23.3 represents a two-dimensional feature space where each point corresponds to a sample and its color reflects which class the sample belongs to. The figure shows examples of

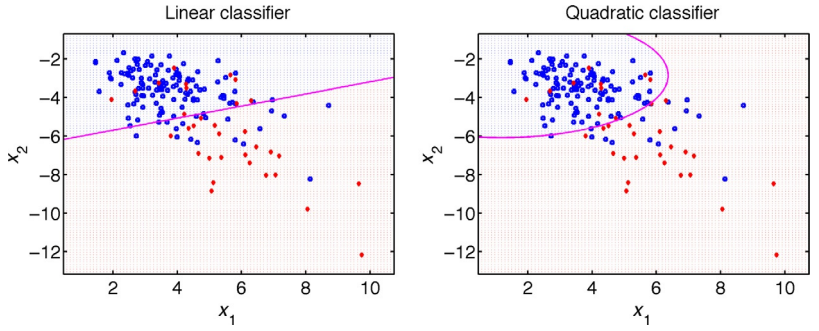

Fig. 23.3. Toy example of the classification process. Both plots show the same set of feature samples of error-related potentials. Signal is filtered using CAR and a bandpass in the [1–10] Hz range (cf. Fig. 23.2). For each trial a sample is composed of a feature vector x_1, x_2, corresponding to the positive and negative peak values in the window [0–800] ms. Samples are color coded *red* and *blue* for error and correct classes, respectively. Linear and quadratic discriminant classifiers are shown in the *left* and *right plot*, respectively. The *magenta line* corresponds to the boundary defined by the classifier function *f*.

quadratic and linear discriminant functions obtained for the same sample distribution (left and right, respectively). It can also be seen that in this case, the functions shown are not able to successfully classify all samples as some red samples fall into the blue subspace and vice versa. The conundrum is thus how to choose the features and decoding method that will yield the best performance (e.g., least decoding errors during online BCI use).

In most BCI applications the decoder calibration follows a supervised approach. Here, it is assumed that a set of N labeled samples, $(\mathbf{X}, \mathbf{Y}^*)$, is available, where $\mathbf{X}=\{\mathbf{x}^1, \mathbf{x}^2, ..\mathbf{x}^N\}'$ is a set of feature vectors corresponding to the measured samples, and $\mathbf{Y}^* = \{\mathbf{y}^{1*}, \mathbf{x}^{2*}, ..\mathbf{y}^{N*}\}'$ is the set of their corresponding desired outputs. In practice, these examples are obtained in a controlled situation where both the input and desired output are observable (i.e., calibration or training phase). For instance, in the case of invasive BCIs decoding upper limb kinematics, SUA is recorded simultaneously with the actual movement of the upper limb (Carmena et al., 2003). Similarly, in the case of motor imagery-based BCIs during the system calibration, subjects are explicitly cued to perform the imaginary movement at specific periods of time; cf. appendix A in Leeb et al. (2013b).

Once the labeled data has been acquired, there are several methods that can be used to calibrate the decoder. As mentioned earlier, depending on the application we may want to decode the neural activity as belonging to one out of multiple discrete categories (classification) or to decode a continuous value (regression).

Among the most common classification methods we can mention linear and quadratic discriminant analysis, which use a probabilistic approach to reduce the likelihood error; support vector machines, which try to maximize the margin of separation between the different classes; and artificial neural networks or random forests based on iterative error minimization algorithms. Following their success in multiple pattern recognition applications, some research groups have explored the use of deep learning algorithms for BCI (Cecotti and Graser, 2011; Sturm et al., 2016; Chai et al., 2017; Schirrmeister et al., 2017). However, most of these studies are restricted to offline analysis results, and are not conclusive on whether these techniques can outperform other types of classifiers. This is presumably due to the fact that the small amount of data available in typical BCI experiments is not enough to exploit the strengths of such methods. Since a full review of them goes beyond the scope of this chapter, interested readers are referred to: Lotte et al. (2007), Parra et al. (2008), Blankertz et al. (2011), Lemm et al. (2011), Haufe et al. (2014), and Krusienski et al. (2006).

Regarding BCI systems based on regression methods, the most common application is the decoding of motor parameters from intracranial recordings. These systems often rely on regularized linear methods (Carmena et al., 2003; Bradberry et al., 2010; Collinger et al., 2013), or Kalman filtering (Hochberg et al., 2012). A review of these and other regression methods can be found in Stulp and Sigaud (2015) and Kao et al. (2014).

In general, all the aforementioned methods aim at estimating the parameters that minimize a measure of error $E(\mathbf{Y}^*,\mathbf{Y})$ between the outputs $\mathbf{Y}$ of the decoder and the labels provided in the dataset $\mathbf{Y}^*$. In other words, the optimal parameters would be those that provide an output as close as possible to the provided examples. Since the parameter optimization process is done by definition in a limited sample, it is based on an empirical estimation of the error. Therefore, it relies on the assumption that the available examples properly reflect the distribution of possible features and their mapping onto the output variable. However, the set of training examples is often small with respect to the dimensionality of the feature space (i.e., the number of selected features). This may lead to inaccurate estimation of the empirical error, and suboptimal decoders. Moreover, samples are

subjected to inherent noise in the measurement processes, as well as artifact contamination. As a result, minimization of the empirical error can lead to a function f that reflects the details of the data samples—including measurement errors—but does not reflect the global properties of the neural correlates of interest (i.e., the real data distribution), i.e., it overfits the training data. These models exhibit a very good performance in the data used for calibration but not in new data samples (i.e., they do not generalize).

Model selection

The existence of multiple machine learning algorithms, and variations of their possible parameters lead to a large number of possible functions (f in Eq. 23.2) that could be used for the decoding. The overarching problem is thus how to choose a proper model (i.e., selected features, type of decoder and its parameters) that performs accurately in the labeled data, while allowing to generalize so as to successfully decode brain activity patterns measured after the calibration phase and during operation of the interface.

As a consequence, to design a BCI it is necessary to compare multiple potential models based on their performance in samples that have not been used in the optimization process. For doing so, one part of the available labeled samples should be left apart for assessing the performance. Importantly, these testing samples should not be used in any step of the decoder design, i.e., feature selection, normalization, or decoder calibration. The need for splitting the labeled data into a training and testing set reduces the amount of available data that can be used to estimate the empirical error, which may lead to overfitting. A means to overcome this is to use the cross-validation method to evaluate the decoder performance on different partitions of the labeled data (Lemm et al., 2011). The most common approach is the n-fold cross-validation where the data is partitioned into n disjoint subsets, termed folds. In each partition, one of these folds is used as testing set and the remaining folds are used for feature selection and calibration.

Of note, samples corresponding to overlapping or nearby windows will be highly correlated. Hence, caution should be taken during calibration of the model to ensure that the training and testing sets are independent. Therefore, random partitions should be avoided and care should be taken to have training and testing sets composed of samples that are not close in time. Otherwise, the testing of the model is prone to yield overly optimistic estimations of its performance.

It is noteworthy, that since the process is performed on each partition, for a given type of model it yields n different sets of parameters (θ_i), each one with its corresponding testing error. The mean and variance of these error measures provide an indication of how suitable that particular type of model is to capture the characteristics of the data; i.e., it is less likely to overfit and generalizes well to unseen data. Once a given method is chosen based on the obtained error across folds, a final decoder to be used during the closed-loop operation can be trained using all the labeled data.

Online classification

During online operation the selected model is used to provide online decoding of the neural measures. The procedures for preprocessing the data and extraction of features as described in the previous sections have to be applied to the incoming stream of data acquired in real-time. BCI implementations can be broadly divided in two categories:

(i). Synchronous interfaces that decode neural signals time-locked to a particular event.
(ii). Asynchronous interfaces that continuously decode the recorded activity.

Synchronous interfaces are typically used to decode ERPs elicited by external stimuli. The most common signals being the aforementioned P300 (Sellers et al., 2014), ErrP (Chavarriaga et al., 2014; Zander et al., 2016), or eye-fixation related potentials (Baccino and Manunta, 2005). In these cases the goal is to decode responses generated by different types of stimuli. Hence, the signals to be analyzed are circumscribed to a specific time window time-locked to the appearance of the stimulus (i.e., often referred to as an epoch). Correspondingly, for model calibration the training dataset will be composed of sample epochs of the different intended classes. During online operation, beside the acquisition of neural signals the system should synchronously provide markers of the relevant events (triggers). The BCI will decode only the signal corresponding to the window of interest time-locked to the corresponding triggers. It is noteworthy that some of the preprocessing steps—most notably the spectral filtering—cannot be limited to the time-locked epochs and should be performed on the whole signal as it is recorded.

In turn, asynchronous interfaces are meant to continuously decode the signal and are typically processed through sliding windows (with or without overlap). The decoding function will be applied on each of these windows producing a corresponding output. This process implies that the frequency at which the BCI produces an output is not necessarily the same at which the signal is recorded. Common examples of this approach are the decoding of sensorimotor rhythms both in invasive and noninvasive BCIs (Leeb et al., 2015); the decoding of

motor kinematics from spike activity (Hochberg et al., 2012; Collinger et al., 2013); the decoding of levels of workload (Brouwer et al., 2012; Borghini et al., 2014); or the decoding of ErrP during a continuous interaction (Milekovic et al., 2012; Omedes et al., 2015; Dias et al., 2018).

PERFORMANCE EVALUATION

Assessing the performance of a BCI system is not trivial, and there is no wide consensus on the best approach to do so. Here we discuss different metrics, inspired from the field of pattern recognition, which are commonly used to identify how capable the system is of correctly decoding mental processes from the brain activity.

As mentioned before, performance of the decoding models is measured in terms of their error, i.e., how close the model's output is to the desired output provided in the training set. There are multiple possible ways of measuring this performance (Thompson et al., 2014).

A common performance measure used in regression problems is the correlation between the targeted values and the output of the decoder. However, this must be taken carefully since this measure is dependent on the number of samples used, and is independent of the scale of the time series (Antelis et al., 2013). An alternative measure is the Fitt's law that computes performance from an index of difficulty of a given movement based on the distance traveled to the size of the target location (Gilja et al., 2012). In the case of classification, the most common metric is the accuracy corresponding to the ratio of misclassified samples with respect to the total number of samples. A system that makes no errors will have an accuracy equal to 1. This metric weighs equally all possible misclassifications irrespective of the output class. In some cases, this may not be appropriate. Let's take the case of classification into one of two classes $\{\mathbf{p}, \mathbf{n}\}$ arbitrarily called positive and negative, respectively. In the case where 80% of the training samples belong to the class **p**, a decoder that always classifies samples in the positive class will yield an accuracy of 0.8. This may be wrongly interpreted as being a good performance, while in reality all samples of the negative class are misclassified.

An alternative is to assess the class-dependent errors in a confusion matrix. Different metrics can be computed with respect to the truly and falsely decoded samples from each of the classes as described in Fig. 23.4 (Left). Based on these measures, besides the overall accuracy, the decoder performance can be analyzed in terms of its true-positive and -negative rates, its sensitivity, or specificity. A common way to analyze a decoder is based on the receiver-operating characteristics (ROC) space, where the true-positive rate (TPR) is plotted vs the decoder's false-positive rate (FPR) (Fig. 23.4, Right; Fawcett, 2006). In this space, a perfect performance will appear in the left-upper corner (TPR = 1; FPR = 0), while random performance will lie on the diagonal curve, $TPR = FPR$. In some cases the classifier function f yields a continuous value as an output reflecting how likely it is that the input sample belongs to a given class. A stereotypical case is a probabilistic classifier where $\mathbf{y}$ corresponds to the posterior probability of a sample $\mathbf{x}$ to belong to a given class C. In this case, the sample will be assigned to this class whenever y exceeds a threshold. In these cases, the performance of the classification method for different threshold values can be plotted as a line in the ROC space, and characterized by the area under that curve (AUC). Given the characteristics of the ROC space, an AUC = 1 will correspond to an optimal classifier while random performance will yield an AUC = 0.5. As it was the case with the cross-validation process, the AUC will inform about the overall performance of a given classification method on the current data and is useful to select the parameters of such a method or to compare across different approaches.

It must be noticed that class imbalance is not the only factor that may lead to misinterpretations. Indeed, the decoder evaluation yields an empirical estimation of the error on the samples of the testing set, and random performance levels are drastically influenced by the number of samples (Winkler et al., 2016; Varoquaux, 2018). For instance, although the theoretical chance of getting heads after tossing a coin is 50% we cannot expect that every time we do it 10 times we will have exactly 5 heads and 5 tails. The same will happen when evaluating the decoder. In a two-class decoding problem the chance level increases from a theoretical value of 50%–70% when only 20 samples are available for each class (confidence interval $\alpha = 0.01$; Müller-Putz et al., 2008). For this purpose, it is important to always evaluate the actual chance level given the particular application and the available testing data samples. This can be done using a theoretical estimate based on binomial or multinomial distributions (Müller-Putz et al., 2008), or empirically through a test where multiple decoders are trained after permutation of the target labels (Winkler et al., 2016). A similar approach can be used to evaluate chance level in regression decoders (Antelis et al., 2013).

Other performance measures follow an information theoretical approach: The BCI decoder can be seen as a communication channel and its performance measured in terms of the amount of information transmitted through this channel. This measure takes into account both the classification accuracy and its speed (Wolpaw et al., 2000). Accordingly, we can estimate the information transfer rate (ITR) of the decoder over a period T as:

	True class	
	p	**n**
p′	*True positives* *TP*	*False positives* *FP*
n′	*False negatives* *FN*	*True negatives* *TN*

$$Accuracy = \frac{TP+TN}{N}$$

$$FPR = \frac{FP}{N} \qquad TPR = \frac{TP}{N}$$

$$Sensitivity = \frac{TP}{P}$$

$$Specificity = 1 - FPR$$

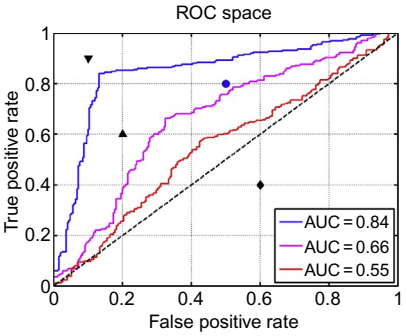

Fig. 23.4. (*Left, Top*) Confusion matrix. *Rows* represent the hypothesized classes (p', n'); i.e., output of the decoder. P, N: Total of samples of classes **p** and **n**, respectively. (*Bottom*) Different performance measures that can be extracted from the confusion matrix. (*Right*) ROC space. *Each symbol* depicts different performance levels (FPR, TPR): ▲: (0.2,0.6), ▼: (0.1,0.9), ●: (0.5,0.8), ♦: (0.6,0.4). *Colored lines* correspond to three examples of ROC curves for different AUC levels.

$$\mathrm{ITR} = \frac{\log_2 N + p\log_2 p + (1-p)\log_2\left(\frac{1-p}{N-1}\right)}{T}, \quad (23.9)$$

where N is the number of possible output classes and p is the accuracy of the decoder. Some variations of this measure have been proposed to consider the a priori probabilities of each class (Schlögl et al., 2007b) or to consider the case where the output of the classifier is not assigned to any of the possible classes (e.g., when the posterior probability for all classes is close to chance level) (Millán et al., 2004).

To sum up, evaluating the decoding performance of a BCI decoder is not straightforward. As seen above there are multiple metrics that can be used reflecting different aspects of the decoding process. The choice of the decoder can be tuned to optimize one of these metrics based on the application requirements. For instance, based on the impact of misclassifications, one may privilege a decoder that ensures a low FPR even if this decreases the overall accuracy. Some composite metrics have thus been proposed to include the cost of misclassifications directly in the performance evaluation (Quitadamo et al., 2012).

In general all these metrics have their advantages and disadvantages and none of them can entirely reflect the decoder behavior. Further discussions on this topic can be found in the literature: see Schlögl et al. (2007b), Haufe et al. (2014), or Thompson et al. (2014) for overall reviews on the topic; Thomas et al. (2013) for motor imagery BCIs; Quitadamo et al. (2012), Lotte and Jeunet (2018), and Seno et al. (2010) for P300-based spellers; and Rousselet and Pernet (2012), Antelis et al. (2013), and Spuler et al. (2015) for regression decoders and the use of correlation metrics.

Moreover, it should be remembered that these metrics have been mainly conceived for pattern recognition applications where samples are extracted from time-invariant distributions and the order of commands does not affect the overall system behavior. These assumptions do not hold in the BCI field where samples are generated by the dynamics of the brain activity, which are inherently affected by the interaction between the human and the machine, and in consequence by previous outcomes of the BCI decoder (Lemm et al., 2011). In addition, it should be noted that the metrics discussed here concerns only the performance of the BCI decoder, which is only one part of the overall human–machine interaction. Hence, they may not entirely reflect how successful the entire system will be in achieving the intended tasks, nor how satisfied the user will be with such a performance. Therefore, a full assessment of the BCI should also consider evaluating the degree of achievement of the task-related goals as well as the corresponding human factors.

SUMMARY, DISCUSSION, AND OUTLOOK

Since the first generation of BCIs, signal processing and machine learning approaches have endowed these systems with significantly greater reliability and decoding capabilities. Indeed, benefits of data-driven algorithms have been largely proven. When compared to hand-picked features and simple thresholding decoders, machine learning can find combinations of features and decision planes much more finely grained than a human counterpart.

As a result, a large number of studies have been devoted to evaluate the BCI decoders and to advance machine learning algorithms so as to improve the decoding performance. Nevertheless, most of these efforts have neglected the fact that the decoding process is only a part of the entire human–machine interaction loop. As a result, they have generally not been successfully translated into a significant improvement of the BCI system

performance or reliability (Chavarriaga et al., 2017). A common mistake is to focus solely on the evaluation of the decoding performance using previously recorded data (e.g., using cross-validation in offline analysis). However, offline performance assessments are not necessarily a good estimator of how the system will perform during closed-loop interaction between the human and the machine. Unfortunately, differences between calibration and online performance are rarely reported in literature. One exception is the work by Leeb et al. (2013b), in which the criteria to move from the calibration to the online phase was for subjects to have a performance higher than 0.4, measured using the Youden index.[c] However, as many as 18 out of 23 subjects in the study had considerably lower performances at the beginning of the online phase (cf. fig. 6 in Leeb et al., 2013b). As a matter of fact, multiple human factors can modulate the user's brain activity including cognitive workload, levels of attention, motivation, or fatigue (Guillot et al., 2005; Kleih et al., 2010; Nijboer et al., 2010). Furthermore, the amount of experience the user has will also affect their proficiency in BCI controlling skills (Leeb et al., 2013b; Kaiser et al., 2014). An underlying assumption in most approaches is that users will adapt throughout the BCI use, and the neural features will be strengthened yielding better control of the system (i.e., acquisition of the BCI skill) or promoting beneficial neuroplasticity changes in the realm of rehabilitation (Ramos-Murguialday et al., 2013; Soekadar et al., 2015; Biasiucci et al., 2018). For instance, a recent study demonstrated consistent and significant performance improvement, contingent with enhancement of feature discriminability, over long training periods without decoder recalibration (Perdikis et al., 2018).

Therefore, the ultimate performance evaluation (and the eventual recalibration) of a BCI system has to be performed during closed-loop use, ideally over an extended period of time. This imposes a challenge, since during BCI use in a real application the ground truth labels of the executed commands are not typically available. A possibility in these cases is to interleave sessions of cued operations to assess performance (and eventually recalibrate the decoder); use contextual information about the task to infer labels in a self-supervised manner (Orsborn et al., 2012; Grizou et al., 2014a, b; Kao et al., 2014; Kindermans et al., 2014; Iturrate et al., 2015b; Jarosiewicz et al., 2015; Perdikis et al., 2015; Zeyl et al., 2016; Hübner et al., 2017; Huebner et al., 2018); or use error-related signals to update the decoders (Spüler et al., 2012a, b). Whenever contextual information is not available, fully unsupervised approaches that do not exploit label information can still be used to unbias the system (Vidaurre et al., 2011).

It is noteworthy, particularly in the case of noninvasive BCIs, that low performances are often explained as a matter of the subject's "BCI illiteracy" (Vidaurre and Blankertz, 2009; Blankertz et al., 2010). This term, introduced about a decade ago, to explain why BCIs often fail to work for a significant number of subjects, has been used in scores of publications ever since. Nonetheless, it has derogatory connotations as it suggests that some subjects have inherent characteristics that prevent them from successfully using these interfaces (Thompson, 2019). The more likely explanations of these failures are inadequate choices in terms of the paradigms, training procedures, and decoder design (Jeunet et al., 2016; Chavarriaga et al., 2017). Thus, the term BCI illiteracy should be avoided and careful discussions on the possible causes of poor performances are required when analyzing BCI results.

Furthermore, the ability of decoding neural activity is not only important for the control of external devices but also provides valuable insights on how specific processes are encoded in the brain. Single-trial decoding from BCI has been used to study diverse phenomena such as cortical dynamics (Ganguly et al., 2011; Koralek et al., 2012) and learning (Ganguly and Carmena, 2009; Sadtler et al., 2014). This application of BCIs requires interpretability of the methods and features used for decoding. In consequence, black-box methods that perform complex, nonlinear transformations of the data may not be the most appropriate for increasing our knowledge of the underlying phenomena exploited by the BCI systems.

In this chapter we have summarized the most common signal processing and machine learning approaches used to process and decode neural activity for BCI purposes. Besides the decoding process, machine learning approaches can also be used to improve the interaction at other parts of the brain–machine loop. One example is the use of reinforcement learning approaches based on decoded ErrP. In these cases decoding of an ErrP signal, reflecting that the user considers the last BCI action as erroneous, is used to adapt the system so as to reduce the likelihood of performing this action again (DiGiovanna et al., 2009; Chavarriaga and Millán, 2010; Iturrate et al., 2015a; Zander et al., 2016; Kim et al., 2017; Salazar-Gomez et al., 2017; Ehrlich and Cheng, 2018). Another strategy to exploit machine learning to improve BCI performance is through the use of shared control (Saeedi et al., 2017), where the BCI-decoded information is not used as a direct command, but combined by an intelligent controller with

[c]The Youden index (Youden, 1950) is computed as $YI = sensitivity + specificity - 1$.

other sources of information such as environmental information provided by external sensors, or by the history of previous commands. A common approach is the use of probabilistic methods to better interpret the output of the decoder. A straightforward example is the use of language models to filter our potential misclassifications in BCI typewriters based on the previous characters written (Blankertz et al., 2007; Martens et al., 2011; Perdikis et al., 2014). Similarly, sensory fusion has also been used to take into account environmental information in mobility applications such as BCI-controlled wheelchairs (Iturrate et al., 2009; Carlson and Millán, 2013), telepresence robots (Escolano et al., 2012; Leeb et al., 2015), or for reaching and grasping prostheses (Kim et al., 2006).

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation NCCR Robotics. I.I. also acknowledges support from the EPFL Fellows fellowship program co-funded by Marie Curie, FP7 Grant agreement no. 291771. This paper only reflects the authors' views and funding agencies are not liable for any use that may be made of the information contained herein.

References

Acunzo DJ, Mackenzie G, van Rossum MCW (2012). Systematic biases in early ERP and ERF components as a result of high-pass filtering. J Neurosci Methods 209: 212–218. https://doi.org/10.1016/j.jneumeth.2012.06.011.

Agashe HA, Paek AY, Zhang Y et al. (2015). Global cortical activity predicts shape of hand during grasping. Front Neurosci 9: 121.

Antelis JM, Montesano L, Ramos-Murguialday A et al. (2013). On the usage of linear regression models to reconstruct limb kinematics from low frequency EEG signals. PLoS One 8: e61976. https://doi.org/10.1371/journal.pone.0061976.

Baccino T, Manunta Y (2005). Eye-fixation-related potentials: insight into parafoveal processing. J Psychophysiol 19: 204–215.

Bertrand O, Perrin F, Pernier J (1985). A theoretical justification of the average reference in topographic evoked potential studies. Electroencephalogr Clin Neurophysiol 62: 462–464. https://doi.org/10.1016/0168-5597(85)90058-9.

Biasiucci R, Leeb I, Iturrate S et al. (2018). Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke. Nat Commun 9: 2421. https://doi.org/10.1038/s41467-018-04673-z.

Bigdely-Shamlo N, Mullen T, Kreutz-Delgado K et al. (2013). Measure projection analysis: a probabilistic approach to EEG source comparison and multi-subject inference. Neuroimage 72: 287–303.

Blankertz B, Krauledat M, Dornhege D et al. (2007). A note on brain actuated spelling with the Berlin brain-computer interface. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 4555 LNCS, no. PART 2, pp. 759–768. https://doi.org/10.1007/978-3-540-73281-5_83.

Blankertz B, Sannelli C, Halder S et al. (2010). Neurophysiological predictor of SMR-based BCI performance. Neuroimage 51: 1303–1309. https://doi.org/10.1016/j.neuroimage.2010.03.022.

Blankertz B, Lemm S, Treder M et al. (2011). Single-trial analysis and classification of ERP components—a tutorial. Neuroimage 56: 814–825. https://doi.org/10.1016/j.neuroimage.2010.06.048.

Borghini G, Astolfi L, Vecchiato G et al. (2014). Measuring neuro-physiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci Biobehav Rev 44: 58–75. https://doi.org/10.1016/j.neubiorev.2012.10.003.

Bouton CE, Shaikhouni A, Annetta NV et al. (2016). Restoring cortical control of functional movement in a human with quadriplegia. Nature 533: 247–250. https://doi.org/10.1038/nature17435.

Boye AT, Kristiansen UQ, Billinger M et al. (2008). Identification of movement-related cortical potentials with optimized spatial filtering and principal component analysis. Biomed Signal Process Control 3: 300–304.

Bradberry TJ, Gentili RJ, Contreras-Vidal JL (2010). Reconstructing three-dimensional hand movements from noninvasive electroencephalographic signals. J Neurosci 30: 3432–3437. https://doi.org/10.1523/JNEUROSCI.6107-09.2010.

Branco MP, Freudenburg ZV, Aarnoutse EJ et al. (2017). Decoding hand gestures from primary somatosensory cortex using high-density ECoG. Neuroimage 147: 130–142.

Brouwer A-M, Hogervorst MA, van Erp JBF et al. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. J Neural Eng 9: 045008. https://doi.org/10.1088/1741-2560/9/4/045008.

Bundy DT, Zellmer E, Gaona CM et al. (2014). Characterization of the effects of the human dura on macro-and micro-electrocorticographic recordings. J Neural Eng 11: 016006.

Carlson T, Millán JdR (2013). Brain-controlled wheelchairs: a robotic architecture. IEEE Rob Autom Mag 20: 65–73.

Carmena JM, Lebedev MA, Crist RE et al. (2003). Learning to control a brain–machine interface for reaching and grasping by primates. PLoS Bio 11: E42. https://doi.org/10.1371/journal.pbio.0000042.

Castermans T, Duvinage M, Cheron G et al. (2014). About the cortical origin of the low-delta and high-gamma rhythms observed in EEG signals during treadmill walking. Neurosci Lett 561: 166–170. https://doi.org/10.1016/j.neulet.2013.12.059.

Cecotti H, Graser A (2011). Convolutional neural networks for P300 detection with application to brain–computer interfaces. IEEE Trans Pattern Anal Mach Intell 33: 433–445. https://doi.org/10.1109/TPAMI.2010.125.

Chai R, Ling SH, San PP et al. (2017). Improving EEG-based driver fatigue classification using sparse-deep belief networks. Front Neurosci 11: 103. https://doi.org/10.3389/fnins.2017.00103.

Chaudhary U, Xia B, Silvoni S et al. (2017). Brain–computer interface-based communication in the completely locked-in state. PLoS Biol 15: e1002593.

Chavarriaga R, Millán JdR (2010). Learning from EEG error-related potentials in non-invasive brain–computer interfaces. IEEE Trans Neural Syst Rehabil Eng 18: 381–388. https://doi.org/10.1109/TNSRE.2010.2053387.

Chavarriaga R, Sobolewski A, Millán JdR (2014). Errare machinale est: the use of error-related potentials in brain–machine interfaces. Front Neurosci 8: 208. https://doi.org/10.3389/fnins.2014.00208.

Chavarriaga R, Fried-Oken M, Kleih S et al. (2017). Heading for new shores! Overcoming pitfalls in BCI design. Brain Comput Interfaces 4: 60–73. https://doi.org/10.1080/2326263X.2016.1263916.

Cincotti F, Mattia D, Aloise F et al. (2008). High-resolution EEG techniques for brain–computer interface applications. J Neurosci Methods 167: 31–42. https://doi.org/10.1016/j.jneumeth.2007.06.031.

Cohen MX (2014). Analyzing neural time series data: theory and practice, MIT Press.

Cohen MX (2015). Comparison of different spatial transformations applied to EEG data: a case study of error processing. Int J Psychophysiol 97 (3): 245–257. https://doi.org/10.1016/j.ijpsycho.2014.09.013.

Collinger JL, Wodlinger B, Downey JE et al. (2013). High-performance neuroprosthetic control by an individual with tetraplegia. Lancet 381: 557–564. https://doi.org/10.1016/S0140-6736(12)61816-9.

Debener S, Ullsperger M, Siegel M et al. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J Neurosci 25: 11730–11737. https://doi.org/10.1523/JNEUROSCI.3286-05.2005.

Delorme A, Palmer J, Onton J et al. (2012). Independent EEG sources are dipolar. PLoS One 7: e30135. https://doi.org/10.1371/journal.pone.0030135.

Dias CL, Sburlea AI, Müller-Putz GR (2018). Masked and unmasked error-related potentials during continuous control and feedback. J Neural Eng 15: 036031.

Diedrichsen J, Shadmehr R (2005). Detecting and adjusting for artifacts in fMRI time series data. Neuroimage 27: 624–634. https://doi.org/10.1016/j.neuroimage.2005.04.039.

DiGiovanna J, Mahmoudi B, Fortes J et al. (2009). Coadaptive brain–machine interface via reinforcement learning. IEEE Trans Biomed Eng 56: 54–64. https://doi.org/10.1109/TBME.2008.926699.

Duda RO, Hart PE, Stork DG (2000). Pattern classification, second edn. Wiley-Interscience, 0471056693. November.

Edelman B, Baxter B, He B (2015). EEG source imaging enhances the decoding of complex right hand motor imagery tasks. IEEE Trans Biomed Eng 63: 4–14. https://doi.org/10.1109/TBME.2015.2467312.

Ehrlich SK, Cheng G (2018). Human-agent co-adaptation using error-related potentials. J Neural Eng 15: 066014.

Escolano C, Antelis JM, Minguez J (2012). A telepresence mobile robot controlled with a noninvasive brain–computer interface. IEEE Trans Syst Man Cybern B Cybern 42: 793–804.

Fawcett T (2006). An introduction to ROC analysis. Pattern Recogn Lett 27: 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

Fernández-Rodríguez A, Velasco-Á lvarez F, Ron-Angevin R (2016). Review of real brain-controlled wheelchairs. J Neural Eng 13: 061001.

Friston KJ, Williams S, Howard R et al. (1996). Movement-related effects in fMRI time-series. Magn Reson Med 35: 346–355. https://doi.org/10.1002/mrm.1910350312.

Ganguly K, Carmena JM (2009). Emergence of a stable cortical map for neuroprosthetic control. PLoS Biol 7: e1000153. https://doi.org/10.1371/journal.pbio.1000153.

Ganguly K, Dimitrov DF, Wallis JD et al. (2011). Reversible large-scale modification of cortical networks during neuroprosthetic control. Nat Neurosci 14: 662–667. https://doi.org/10.1038/nn.2797.

Gilja V, Nuyujukian P, Chestek CA et al. (2012). A high-performance neural prosthesis enabled by control algorithm design. Nat Neurosci 15: 1752–1757. https://doi.org/10.1038/nn.3265.

Grizou J, Iturrate I, Montesano L et al. (2014a). Calibration-free BCI based control. In: Proceedings of the 28th AAAI conference on artificial intelligence (AAAI), pp, 1213–1220.

Grizou J, Iturrate I, Montesano L et al. (2014b). Interactive learning from unlabeled instructions. In: Proceedings of the thirtieth conference on uncertainty in artificial intelligence (UAI).

Grosse-Wentrup M, Buss M (2008). Multiclass common spatial patterns and information theoretic feature extraction. IEEE Trans Biomed Eng 55: 1991–2000. https://doi.org/10.1109/TBME.2008.921154.

Guger C, Daban S, Sellers E et al. (2009). How many people are able to control a P300-based brain–computer interface (BCI)? Neurosci Lett 462: 94–98. https://doi.org/10.1016/j.neulet.2009.06.045.

Guillot A, Haguenauer M, Dittmar A et al. (2005). Effect of a fatiguing protocol on motor imagery accuracy. Eur J Appl Physiol 95: 186–190. https://doi.org/10.1007/s00421-005-1400-x.

Gwin JT, Gramann K, Makeig S et al. (2010). Removal of movement artifact from high-density EEG recorded during walking and running. J Neurophysiol 103: 3526–3534. https://doi.org/10.1152/jn.00105.2010.

Hardoon DR, Szedmak S, Shawe-Taylor J (2004). Canonical correlation analysis: an overview with application to learning methods. Neural Comput 16: 2639–2664.

Haufe S, Meinecke F, Görgen K et al. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87: 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067.

Hochberg LR, Bacher D, Jarosiewicz B et al. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. Nature 485: 372–375. https://doi.org/10.1038/nature11076.

Hübner D, Verhoeven T, Schmid K et al. (2017). Learning from label proportions in brain–computer interfaces: online unsupervised learning with guarantees. PloS One 12: e0175856.

Huebner D, Verhoeven T, Mueller K-R et al. (2018). Unsupervised learning for brain–computer interfaces based on event-related potentials: review and online comparison [research frontier]. IEEE Comput Intell Mag 13: 66–77.

Hyvärinen A, Karhunen J, Oja E (2004). Independent component analysis, vol. 46 John Wiley & Sons.

Iturrate I, Antelis JM, Kubler A et al. (2009). A noninvasive brain-actuated wheelchair based on a p300 neurophysiological protocol and automated navigation. IEEE Trans Robot 25: 614–627.

Iturrate I, Chavarriaga R, Montesano L et al. (2014). Latency correction of error potentials between different experiments reduces calibration time for single-trial classification. J Neural Eng 11: 036005. https://doi.org/10.1088/1741-2560/11/3/036005.

Iturrate I, Chavarriaga R, Montesano L et al. (2015a). Teaching brain–machine interfaces as an alternative paradigm to neuroprosthetics control. Sci Rep 5: 13893. https://doi.org/10.1038/srep13893.

Iturrate I, Grizou J, Omedes J et al. (2015b). Exploiting task constraints for self-calibrated brain–machine interface control using error-related potentials. PLoS One 10: e0131491.

Iturrate I, Chavarriaga R, Pereira M et al. (2018). Human EEG reveals distinct neural correlates of power and precision grasping types. Neuroimage 181: 635–644.

Jarosiewicz B, Sarma AA, Bacher D et al. (2015). Virtual typing by people with tetraplegia using a self-calibrating intracortical brain–computer interface. Sci Transl Med 7: 313ra179. https://doi.org/10.1126/scitranslmed.aac7328.

Jeunet C, Jahanpour E, Lotte F (2016). Why standard brain–computer interface (BCI) training protocols should be changed: an experimental study. J Neural Eng 13: 036024. https://doi.org/10.1088/1741-2560/13/3/036024.

Jochumsen M, Niazi IK, Dremstrup K et al. (2016). Detecting and classifying three different hand movement types through electroencephalography recordings for neurorehabilitation. Med Biol Eng Comput 54: 1491–1501.

Kaiser V, Bauernfeind G, Kreilinger A et al. (2014). Cortical effects of user training in a motor imagery based brain–computer interface measured by fNIRS and EEG. Neuroimage 85 (Pt. 1): 432–444. https://doi.org/10.1016/j.neuroimage.2013.04.097.

Kao J, Stavisky S, Sussillo D et al. (2014). Information systems opportunities in brain–machine interface decoders. Proc IEEE 102: 666–682. https://doi.org/10.1109/JPROC.2014.2307357.

Kim HK, Biggs J, Schloerb W et al. (2006). Continuous shared control for stabilizing reaching and grasping with brain–machine interfaces. IEEE Trans Biomed Eng 53: 1164–1173.

Kim SK, Kirchner EA, Stefes A et al. (2017). Intrinsic interactive reinforcement learning–using error-related potentials for real world human–robot interaction. Sci Rep 7: 17562.

Kindermans P-J, Tangermann M, Müller K-R et al. (2014). Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP speller. J Neural Eng 11: 035005.

Kleih SC, Nijboer F, Halder S et al. (2010). Motivation modulates the P300 amplitude during brain–computer interface use. Clin Neurophysiol 121: 1023–1031. https://doi.org/10.1016/j.clinph.2010.01.034.

Koralek AC, Jin X, Long JD et al. (2012). Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills. Nature 483: 331–335. https://doi.org/10.1038/nature10845.

Krusienski DJ, Sellers EW, Cabestaing F et al. (2006). A comparison of classification techniques for the P300 speller. J Neural Eng 3: 299–305. https://doi.org/10.1088/1741-2560/3/4/007.

Kubanek J, Miller K, Ojemann J et al. (2009). Decoding flexion of individual fingers using electrocorticographic signals in humans. J Neural Eng 6: 066001.

Lagerlund T, Sharbrough F, Busacker N (1997). Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. J Clin Neurophysiol 14: 73–82.

Lee K, Liu D, Perroud L et al. (2017). A brain-controlled exoskeleton with cascaded event-related desynchronization classifiers. Robot Auton Syst 90: 15–23. https://doi.org/10.1016/j.robot.2016.10.005.

Leeb R, Lancelle M, Kaiser V et al. (2013a). Thinking penguin: multi-modal brain–computer interface control of a VR game. IEEE Trans Comput Intell AI Games 5: 117–128. https://doi.org/10.1109/TCIAIG.2013.2242072.

Leeb R, Perdikis S, Tonin L et al. (2013b). Transferring brain–computer interfaces beyond the laboratory: successful application control for motor-disabled users. Artif Intell Med 59: 121–132. https://doi.org/10.1016/j.artmed.2013.08.004.

Leeb R, Tonin L, Rohm M et al. (2015). Towards independence: a BCI telepresence robot for people with severe motor disabilities. Proc IEEE 103: 969–982. https://doi.org/10.1109/JPROC.2015.2419736.

Lemm S, Blankertz B, Dickhaus T et al. (2011). Introduction to machine learning for brain imaging. Neuroimage 56: 387–399. https://doi.org/10.1016/j.neuroimage.2010.11.004.

Lew EY, Chavarriaga R, Silvoni S et al. (2014). Single trial prediction of self-paced reaching directions from EEG signals. Front Neurosci 8: 222.

López-Larraz E, Montesano L, Gil-Agudo Á et al. (2014). Continuous decoding of movement intention of upper limb self-initiated analytic movements from pre-movement EEG correlates. J Neuroeng Rehabil 11: 153.

Lotte F, Jeunet C (2018). Defining and quantifying users' mental imagery-based BCI skills: a first step. J Neural Eng15. https://doi.org/10.1088/1741-2552/aac577.

Lotte F, Congedo M, Lécuyer A et al. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. J Neural Eng 4: R1–R13. https://doi.org/10.1088/1741-2560/4/2/R01.

Makeig S, Debener S, Onton J et al. (2004). Mining event-related brain dynamics. Trends Cogn Sci 8: 204–210. https://doi.org/10.1016/j.tics.2004.03.008.

Marshall D, Coyle D, Wilson S et al. (2013). Games, gameplay, and BCI: the state of the art. IEEE Trans Comput Intell AI Games 5: 82–99. https://doi.org/10.1109/TCIAIG.2013.2263555.

Martens SM, Mooij JM, Hill NJ et al. (2011). A graphical model framework for decoding in the visual ERP-based BCI speller. Neural Comput 23: 160–182. https://doi.org/10.1162/NECO_a_00066.

Martin S, Brunner P, Iturrate I et al. (2016). Word pair classification during imagined speech using direct brain recordings. Sci Rep 6: 25803.

Martin S, Iturrate I, Millán JdR et al. (2018). Decoding inner speech using electrocorticography: progress and challenges toward a speech prosthesis. Front Neurosci 12: 422.

McFarland DJ, McCane LM, David SV et al. (1997). Spatial filter selection for EEG-based communication. Electroencephalogr Clin Neurophysiol 103: 386–394.

Milekovic T, Ball T, Schulze-Bonhage A et al. (2012). Error-related electrocorticographic activity in humans during continuous movements. J Neural Eng 9: 026007.

Millán JdR, Carmena J (2010). Invasive or noninvasive: understanding brain–machine interface technology [conversations in BME]. IEEE Eng Med Biol Mag IEEE 29: 16–22. https://doi.org/10.1109/MEMB.2009.935475.

Millán JdR, Renkens F, Mourino J et al. (2004). Noninvasive brain-actuated control of a mobile robot by human EEG. IEEE Trans Biomed Eng 51: 1026–1033. https://doi.org/10.1109/TBME.2004.827086.

Miller J, Zanos S, Fetz E et al. (2009). Decoupling the cortical power spectrum reveals real-time representation of individual finger movements in humans. J Neurosci 29: 3132–3137.

Müller-Putz G, Scherer R, Brunner C et al. (2008). Better than random: a closer look on BCI results. Int J Bioelectromagnetism 10: 52–55.

Niazi IK, Jiang N, Tiberghien O et al. (2011). Detection of movement intention from single-trial movement-related cortical potentials. J Neural Eng 8: 066009.

Nicolelis MAL, Dimitrov D, Carmena JM et al. (2003). Chronic, multisite, multielectrode recordings in macaque monkeys. Proc Natl Acad Sci U S A 100: 11041–11046. https://doi.org/10.1073/pnas.1934665100.

Nijboer F, Birbaumer N, Kübler A (2010). The influence of psychological state and motivation on brain–computer interface performance in patients with amyotrophic lateral sclerosis—a longitudinal study. Front Neurosci 4: . https://doi.org/10.3389/fnins.2010.00055.

Omedes J, Iturrate I, Montesano L et al. (2013). Using frequency-domain features for the generalization of EEG error-related potentials among different tasks. 2013 35th annual international conference of the IEEE. Engineering in Medicine and Biology Society (EMBC), pp. 5263–5266.

Omedes J, Iturrate I, Minguez J et al. (2015). Analysis and asynchronous detection of gradually unfolding errors during monitoring tasks. J Neural Eng 12: 056001.

Orsborn AL, Dangi S, Moorman HG et al. (2012). Closed-loop decoder adaptation on intermediate time-scales facilitates rapid BMI performance improvements independent of decoder initialization conditions. IEEE Trans Neural Syst Rehabil Eng 20: 468–477. https://doi.org/10.1109/TNSRE.2012.2185066.

Parra L, Christoforou C, Gerson A et al. (2008). Spatiotemporal linear decoding of brain state. IEEE Signal Process Mag 25: 107–115. https://doi.org/10.1109/MSP.2008.4408447.

Pasley BN, David SV, Mesgarani N et al. (2012). Reconstructing speech from human auditory cortex. PLoS Biol 10: e1001251.

Perdikis S, Leeb R, Williamson J et al. (2014). Clinical evaluation of BrainTree, a motor imagery hybrid BCI speller. J Neural Eng 11: 036003. https://doi.org/10.1088/1741-2560/11/3/036003.

Perdikis S, Leeb R, Chavarriaga R et al. (2015). Context-aware learning for finite mixture models. ArXiv. ArXiv ID: 1507.08272.

Perdikis S, Tonin L, Saeedi S et al. (2018). The cybathlon BCI race: successful longitudinal mutual learning with two tetraplegic users. PLoS Biol 16: 1–28. https://doi.org/10.1371/journal.pbio.2003787.

Pereira J, Sburlea AI, Müller-Putz GR (2018). EEG patterns of self-paced movement imaginations towards externally-cued and internally-selected targets. Sci Rep 8: 13394.

Pfurtscheller G, Lopes da Silva FH (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol 110: 1842–1857. https://doi.org/10.1016/S1388-2457(99)00141-8.

Pistohl T, Schulze-Bonhage A, Aertsen A et al. (2012). Decoding natural grasp types from human ECoG. Neuroimage 59: 248–260.

Pourtois G, Delplanque S, Michel C et al. (2008). Beyond conventional event-related brain potential (ERP): exploring the time-course of visual emotion processing using topographic and principal component analyses. Brain Topogr 20: 265–277. https://doi.org/10.1007/s10548-008-0053-6.

Quian Quiroga R (2007). Spike sorting. Scholarpedia 2: 3583.

Quiroga RQ, Panzeri S (2009). Extracting information from neuronal populations: information theory and decoding approaches. Nat Rev Neurosci 10: 173–185. https://doi.org/10.1038/nrn2578.

Quitadamo LR, Abbafati M, Cardarilli GC et al. (2012). Evaluation of the performances of different P300 based brain–computer interfaces by means of the efficiency metric. J Neurosci Methods 203: 361–368. https://doi.org/10.1016/j.jneumeth.2011.10.010.

Ramos-Murguialday A, Broetz D, Rea M et al. (2013). Brain–machine interface in chronic stroke rehabilitation: a controlled study. Ann Neurol 74: 100–108.

Randazzo L, Iturrate I, Chavarriaga R et al. (2015). Detecting intention to grasp during reaching movements from EEG. In: 37th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 1115–1118.

Reis PMR, Hebenstreit F, Gabsteiger F et al. (2014). Methodological aspects of EEG and body dynamics measurements during motion. Front Hum Neurosci 8. https://doi.org/10.3389/fnhum.2014.00156.

Rivet B, Souloumiac A, Attina V et al. (2009). xDAWN algorithm to enhance evoked potentials: application to brain–computer interface. IEEE Trans Biomed Eng 56: 2035–2043.

Rossi L, Foffani G, Marceglia S et al. (2007). An electronic device for artefact suppression in human local field potential recordings during deep brain stimulation. J Neural Eng 4: 96–106. https://doi.org/10.1088/1741-2560/4/2/010.

Rousselet GA (2012). Does filtering preclude us from studying ERP time-courses? Front Psychol 3: 131. https://doi.org/10.3389/fpsyg.2012.00131.

Rousselet GA, Pernet CR (2012). Improving standards in brain–behavior correlation analyses. Front Hum Neurosci 6: 119.

Sadtler PT, Quick KM, Golub MD et al. (2014). Neural constraints on learning. Nature 512: 423–426. https://doi.org/10.1038/nature13665.

Saeedi S, Chavarriaga R, Millán JdR (2017). Long-term stable control of motor-imagery BCI by a locked-in user through adaptive assistance. IEEE Trans Neural Syst Rehabil Eng 25: 380–391. https://doi.org/10.1109/TNSRE.2016.2645681.

Salazar-Gomez AF, DelPreto J, Gil S et al. (2017). Correcting robot mistakes in real time using EEG signals. 2017 IEEE international conference on robotics and automation (ICRA), IEEE, pp. 6570–6577.

Sani OG, Chavarriaga R, Shamsollahi MB et al. (2016). Detection of movement related cortical potential: effects of causal vs. non-causal processing. In: 2016 38th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Aug, pp. 5733–5736. https://doi.org/10.1109/EMBC.2016.7592029.

Sburlea AI, Müller-Putz GR (2018). Exploring representations of human grasping in neural, muscle and kinematic signals. Sci Rep 8: 16669.

Schirrmeister RT, Springenberg JT, Fiederer LDJ et al. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. Hum Brain Mapp 38: 5391–5420. https://doi.org/10.1002/hbm.23730.

Schlögl A, Keinrath C, Zimmermann D et al. (2007a). A fully automated correction method of EOG artifacts in EEG recordings. Clin Neurophysiol 118: 98–104. https://doi.org/10.1016/j.clinph.2006.09.003.

Schlögl A, Kronegg J, Huggins J et al. (2007b). Evaluation criteria in BCI research. In: G Dornhege, JdR Millán, T Hinterberger, DJ McFarland, K-R Müller (Eds.), Toward brain–computer interfacing. MIT Press, pp. 327–342. chapter 19.

Schwarz A, Ofner P, Pereira J et al. (2017). Decoding natural reach-and-grasp actions from human EEG. J Neural Eng 15: 016005.

Sellers EW, Ryan DB, Hauser CK (2014). Noninvasive brain–computer interface enables communication after brainstem stroke. Sci Transl Med 6: 257re7. https://doi.org/10.1126/scitranslmed.3007801.

Seno BD, Matteucci M, Mainardi LT (2010). The utility metric: a novel method to assess the overall performance of discrete brain–computer interfaces. IEEE Trans Neural Syst Rehabil Eng 18: 20–28. https://doi.org/10.1109/TNSRE.2009.2032642.

Soekadar SR, Birbaumer N, Slutzky MW et al. (2015). Brain–machine interfaces in neurorehabilitation of stroke. Neurobiol Dis 83: 172–179. https://doi.org/10.1016/j.nbd.2014.11.025.

Soekadar SR, Witkowski M, Gómez C et al. (2016). Hybrid EEG/EOG-based brain/neural hand exoskeleton restores fully independent daily living activities after quadriplegia. Sci Robot 1 (1): eaag3296. https://doi.org/10.1126/scirobotics.aag3296.

Spüler M, Bensch M, Kleih S et al. (2012a). Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI. Clin Neurophysiol 123: 1328–1337. https://doi.org/10.1016/j.clinph.2011.11.082.

Spüler M, Rosenstiel W, Bogdan M (2012b). Online adaptation of a c-VEP brain–computer interface (BCI) based on error-related potentials and unsupervised learning. PloS One 7: e51077.

Spüler M, Walter A, Rosenstiel W et al. (2014). Spatial filtering based on canonical correlation analysis for classification of evoked or event-related potentials in EEG data. IEEE Trans Neural Syst Rehabil Eng 22: 1097–1103. https://doi.org/10.1109/TNSRE.2013.2290870.

Spuler M, Sarasola-Sanz A, Birbaumer N et al. (2015). Comparing metrics to evaluate performance of regression methods for decoding of neural signals. Conf Proc IEEE Eng Med Biol Soc: 1083–1086. https://doi.org/10.1109/EMBC.2015.7318553.

Stulp F, Sigaud O (2015). Many regression algorithms, one unified model: a review. Neural Netw 69: 60–79. https://doi.org/10.1016/j.neunet.2015.05.005.

Sturm I, Lapuschkin S, Samek W et al. (2016). Interpretable deep neural networks for single-trial EEG classification. J Neurosci Methods 274: 141–145. https://doi.org/10.1016/j.jneumeth.2016.10.008.

Thomas E, Dyson M, Clerc M (2013). An analysis of performance evaluation for motor-imagery based BCI. J Neural Eng 10: 031001. https://doi.org/10.1088/1741-2560/10/3/031001.

Thompson MC (2019). Critiquing the concept of BCI illiteracy. Sci Eng Ethics 25 (4): 1217–1233. https://doi.org/10.1007/s11948-018-0061-1.

Thompson DE, Quitadamo LR, Mainardi L et al. (2014). Performance measurement for brain–computer or brain–machine interfaces: a tutorial. J Neural Eng11: 035001. https://doi.org/10.1088/1741-2560/11/3/035001.

Torres Valderrama A, Oostenveld R, Vansteensel MJ et al. (2010). Gain of the human dura in vivo and its effects on invasive brain signal feature detection. J Neurosci Methods 187: 270–279. https://doi.org/10.1016/j.jneumeth.2010.01.019.

Urigüen JA, Garcia-Zapirain B (2015). EEG artifact removal-state-of-the-art and guidelines. J Neural Eng 12: 031001. https://doi.org/10.1088/1741-2560/12/3/031001.

VanRullen R (2011). Four common conceptual fallacies in mapping the time course of recognition. Front Psychol 2: 365. https://doi.org/10.3389/fpsyg.2011.00365.

Vansteensel MJ, Pels EG, Bleichner MG et al. (2016). Fully implanted brain–computer interface in a locked-in patient with ALS. N Engl J Med 375: 2060–2066.

Varoquaux G (2018). Cross-validation failure: small sample sizes lead to large error bars. Neuroimage 180: 68–77. https://doi.org/10.1016/j.neuroimage.2017.06.061.

Vidaurre C, Blankertz B (2009). Towards a cure for BCI illiteracy: machine-learning based co-adaptive learning. Proceedings of the 7th NFSI & ICBEM.

Vidaurre C, Kawanabe M, von Bünau P et al. (2011). Toward unsupervised adaptation of LDA for brain–computer interfaces. IEEE Trans Biomed Eng 58: 587.

Waldert S, Preissl H, Demandt E et al. (2008). Hand movement direction decoded from MEG and EEG. J Neurosci 28: 1000–1008.

Widmann A, Schröger E (2012). Filter effects and filter artifacts in the analysis of electrophysiological data. Front Psychol 3: 233. https://doi.org/10.3389/fpsyg.2012.00233.

Winkler AM, Webster MA, Brooks JC et al. (2016). Non-parametric combination and related permutation tests for neuroimaging. Hum Brain Mapp 37: 1486–1511. https://doi.org/10.1002/hbm.23115.

Wolpaw JR, Birbaumer N, Heetderks WJ et al. (2000). Brain–computer interface technology: a review of the first international meeting. IEEE Trans Rehabil Eng 8: 164–173. https://doi.org/10.1109/TRE.2000.847807.

Yael D, Vecht JJ, Bar-Gad I (2018). Filter-based phase shifts distort neuronal timing information. eNeuro 5: 0261-17. https://doi.org/10.1523/ENEURO.0261-17.2018.

Youden WJ (1950). Index for rating diagnostic tests. Cancer 3: 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.

Zander TO, Krol LR, Birbaumer NP et al. (2016). Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity. Proc Natl Acad Sci 113: 14898–14903.

Zeyl T, Yin E, Keightley M et al. (2016). Partially supervised P300 speller adaptation for eventual stimulus timing optimization: target confidence is superior to error-related potential score as an uncertain label. J Neural Eng 13: 026008. https://doi.org/10.1088/1741-2560/13/2/026008.