Chapter 12

A review of feature extraction and classification algorithms for image RSVP-based BCI

Zhengwei Wang[1], Graham Healy[1], Alan F. Smeaton[1], and Tomas E. Ward[1]

Abstract

In this chapter, we introduce an architecture for rapid serial visual presentation (RSVP)-based brain–computer interface (BCI) systems that use electroencephalography (EEG). Hereafter, we will refer to the coupling of the RSVP protocol with EEG to support a target-search BCI as RSVP-EEG. Our focus in this chapter is on a review of feature extraction and classification algorithms applied in RSVP-EEG development. We briefly present the commonly deployed algorithms and describe their properties based on the literature. We conclude with a discussion on the future trajectory of this exciting branch of BCI research.

12.1 Introduction

The rapid serial visual presentation (RSVP) is a method that can be used to extend the brain–computer interface (BCI) approach to enable high throughput target image recognition applications [1–3]. Using electroencephalography (EEG) signals to label or rank images is of practical interest as many types of images cannot be automatically labeled by a computer [2]. A common example here is to enhance the performance of satellite imagery analysts, by performing selection to get a smaller number of images for later and more detailed inspection [1]. In the RSVP target-search paradigm (see Figure 12.1), there is a rapid succession of images presented on screen, in which only a small percentage contain target images. Images are typically presented to participants at a very fast speed on a monitor (5–12 images per second). These infrequent target images are known to elicit the P300 event-related potential (ERP), a type of brain response that has a well-established history of study [4]. The idea is that the participant is unaware when a target stimulus is going to appear; hence, its presentation on screen elicits the P300 ERP reflecting the orientation of participant's attention to the

[1]Insight Centre for Data Analytics, School of Computing, Dublin City University, Ireland

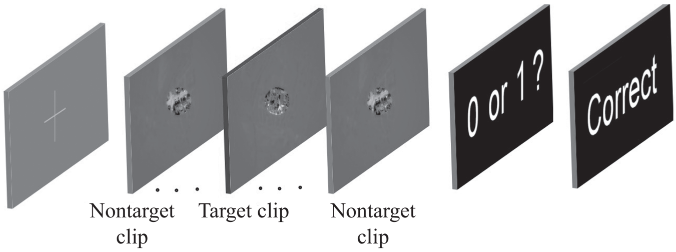


Figure 12.1 RSVP paradigm protocol in Section 12.1 [1]

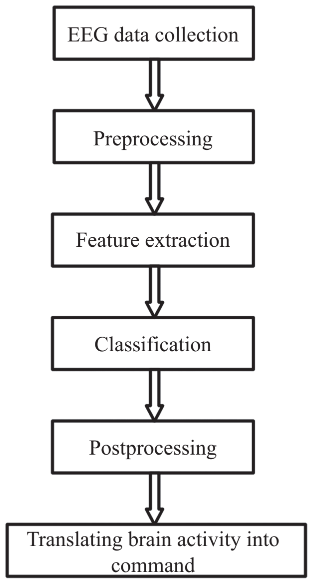


Figure 12.2 Block diagram of a typical BCI system in Section 12.1

stimulus. These brain activity related responses, when extracted, can be used through application of signal processing and machine-learning techniques to enable labeling and/or ranking of images.

In order to use an RSVP-EEG BCI in this way, a user must be capable of responding with brain activity patterns that can be identified automatically. In this regard, the use of an "oddball" paradigm to elicit P300 ERP responses is ideal as targets searched for tend to be infrequent in many datasets and the response has characteristic features.

Figure 12.2 shows the stages of a typical BCI system. Preprocessing, feature extraction, classification and postprocessing are classification system components. Changes in any one of these components can alter the performance of a BCI system. Preprocessing refers to denoising signals, i.e., filtering, artifact rejection, normalization, etc. Feature extraction and classification both belong to the machine-learning section and are essential elements of BCI systems. Postprocessing refers to the use of context information to eliminate outliers which can improve the performance of a classifier.

To date, a number of thorough reviews of classification techniques for BCI have been published [5,6], but none have been specifically dedicated to the review of feature extraction and classification algorithms used for RSVP-BCI research. More broadly, the RSVP target-detection problem is part of a wider field of study that investigates single-trial detection methods [1,2,7].

In Section 12.2, we give a brief introduction to the RSVP-EEG experimental setup. We then show several spatiotemporal signals that are typically present in RSVP-EEG, which have discriminative properties, e.g., the P300 and N200. It should be noted that the common objective of all BCI systems is to maximize classification accuracy rather than providing an interpretation of the underlying neurophysiology.

Finally, we describe the preprocessing step and the problems of RSVP-EEG data availability in the literature which has impact on algorithm development, reproducibility and benchmarking.

In Section 12.3, we outline common strategies used to extract useful features from RSVP-EEG data, namely, spatial filters achieved using (un)supervised techniques such as independent component analysis (ICA) which aims to find a linear representation of non-Gaussian data so as to maximize a statistical independence metric, time-frequency representation which decomposes RSVP-EEG to the time-frequency domain and some other feature extraction methods. Spatial filtering allows for dimensionality reduction by transforming high spatial dimension EEG to a subspace according to different optimization objectives, e.g., improving the signal-to-noise ratio (SNR). Reducing data dimensionality in this way is often essential to overcoming issues with having relatively fewer training examples than there are a high number of features—a scenario commonly referred to as the "curse of dimensionality."

In Section 12.4, we explore a number of commonly used classification strategies which covers both linear and nonlinear techniques. Linear classification techniques include linear discriminant analysis (LDA), Bayesian linear regression (BLR), logistic regression (LR) and support vector machine (SVM). Nonlinear classification approaches are mainly focused on artificial neural networks (NNs).

Therefore, and overall, this paper aims to survey the different feature extraction and classification techniques used in RSVP-based BCI research and to identify their critical properties, shortcomings and advantages. It also provides newcomers to the RSVP-BCI area with an introduction—a framework within which an analysis of RSVP-EEG data can be understood.

12.2 Overview of RSVP experiments and EEG data

12.2.1 RSVP experiment for EEG data acquisition

Data acquisition for RSVP-EEG experiments is typically carried out using two computers. One computer is used for stimulus presentation and the other for recording and monitoring of EEG data from participants. A typical setup is shown in Figure 12.3.

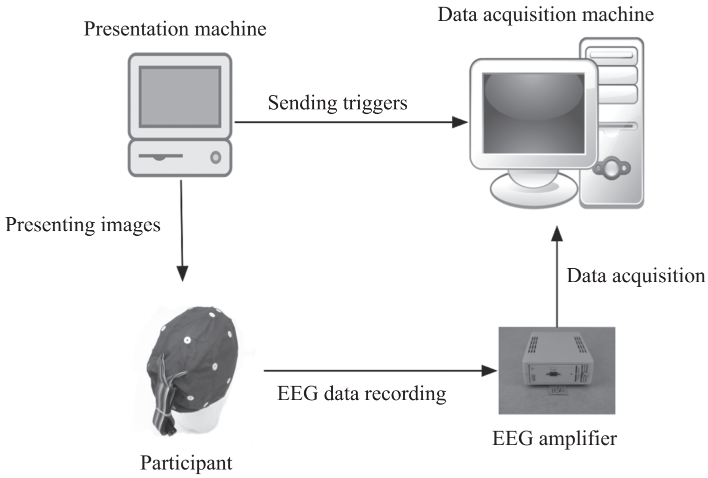


Figure 12.3 RSVP experiment setup in Section 12.2.1

The EEG amplifier is used for recording the EEG signals measured from the participant. When displaying the image sequence to participants, a timestamp for each image must be recorded and aligned with the multichannel time-series EEG captured on the acquisition computer. These are commonly referred to as triggers.

In RSVP-based BCI research, triggers are normally sent from the presentation software (e.g., PsychoPy, E-prime) either to the EEG acquisition device directly [7,8] via a physical port or to the acquisition software [9]. Due to the fast presentation speeds involved with RSVP-EEG, careful attention should be given to ensure that stimulus presentation timings are as expected. We suggest the validation of software triggers in an RSVP experiment against triggers captured using an optical sensor exposed to the presentation screen as this can help to resolve subtle timing issues that may be present [10].

12.2.2 Brief introduction to RSVP-EEG pattern

The most widely used pattern in the EEG signals acquired during RSVP-BCI is the P300 ERP. The P300 is a complex endogenous response that can be subdivided into a novelty-related P3a component and a posterior occurring component commonly encountered in RSVP-search referred to as the P3b. The discovery of the P300 arose from the confluence of increased technological capability for signal averaging applied to human neuroelectric measures and the impact of information theory on psychological research [11]. The P300 is often characterized by its amplitude and latency, where it is defined as the largest positive-going peak in the time range of 300–800 ms following a stimulus presentation. Its latency and amplitude can vary depending on stimulus modality, task conditions, subject age and other factors [4]. Figure 12.4 shows an example of a P300 (P3b) response at channel Pz in one RSVP search

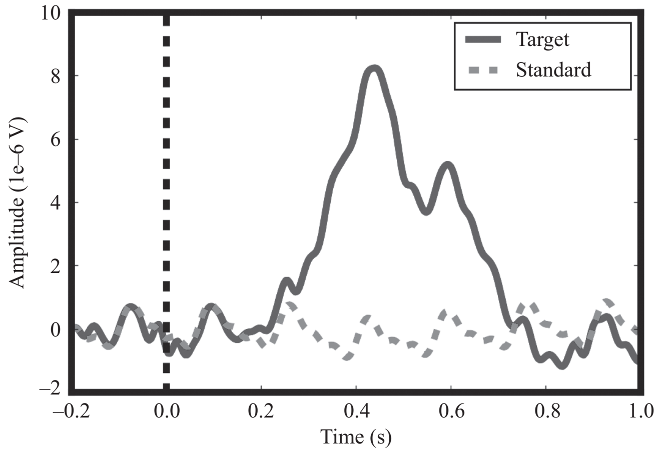


Figure 12.4 P300 response example at the Pz channel in an RSVP experiment in Section 12.2.2, the EEG signal has been band-passed between 0.1 and 30 Hz

task. It can be seen that the P300 peak occurs at around 480 ms. The periodic oscillation that can be seen in the ERP average for standard images is due to a steady-state visual evoked potential response [12]. To facilitate presentation of the concepts involved in RSVP-EEG, we make use of the neurally augmented image labeling strategies (NAILS) dataset [13]. This EEG dataset is part of an open data challenge carried out in 2017 [14].[1]

It is worth noting that not all participants display the stereotyped P300 response, with some displaying characteristics such as low amplitude components leading to unfavorable SNR properties. Reasons for this will not be explored here, but further information can be found in [15].

The P300 is not the only ERP that is commonly encountered when using an RSVP target search paradigm. Earlier, ERPs (notably the N200) are often present alongside the P3 [16] and can be useful in providing discriminative information for classification. In Figure 12.5, it can be seen that both early- and later-time regions generate discriminative ERP-related activity across participants [13].

[1]EEG data from 1 to 9 participants in NAILS was used in this work. Data collection was carried out with approval from Dublin City University's Research Ethics Committee (DCU REC/2016/099). Each participant completed six different tasks (INSTR, WIND1, WIND2, UAV1, UAV2 and BIRD). For each task, participants were asked to search for specific target images from the presented images (i.e., an airplane has the role of target in UAV1 and UAV2 tasks, a keyboard instrument is the target for the INSTR task, while a windfarm is the target in WIND1 and WIND2 tasks, parrot being the target in BIRD task). Each task was divided into 9 blocks, where each block contains 180 images (9 targets/171 standards); thus, there were 486 target and 9,234 standard images available for each participant. Images were presented to participants at a 6-Hz presentation rate. EEG data was recorded using a 32-channel BrainVision actiCHamp at 1,000 samples/s sampling frequency, using electrode locations as defined by the 10–20 system.

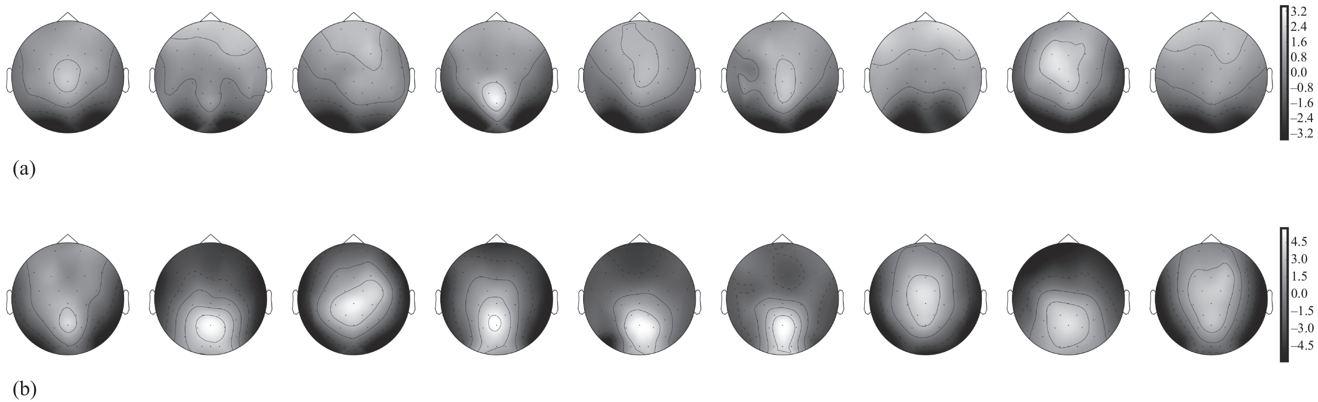


Figure 12.5 *ERPs for each participant corresponding to time regions of 200–340 and 370–700 ms in Section 12.2.2. These are selected for presentation to emphasize the presence of discriminative ERP-related activity in these time regions across participants, namely, time regions coinciding with P200, N200 and P300 ERP activity. (a) Early time region ERP for nine participants from left to right and (b) later time region ERP for nine participants from left to right*

Single-trial detection methods are not restricted solely to BCI-related contexts and are often found as part of a researcher's toolkit in developing an analysis pipeline when working with neuroscientific data. What such methods typically strive to accomplish is striking a balance between the neurophysiological interpretability of a model and its complexity. Transparent models are more likely to lend themselves to meaningful interpretations. While a strategy of enforcing simplistic models with few parameters may aid in interpretability, the purpose of RSVP-EEG is to maximize information throughout as defined by some performance metric.

In reality, a BCI can make use of nonneural signal sources present in the EEG. For example, some participants may, without realizing, blink their eyes upon seeing a target in an RSVP experiment, where the eye-blink will impart a large voltage deflection in the EEG. We are primarily concerned, however, with direct neural signal sources in this chapter and strategies to utilize these.

12.2.3 RSVP-EEG data preprocessing and properties

Preprocessing of some kind is generally a required step before any meaningful interpretation or use of the EEG data can be realized. Preprocessing typically involves re-referencing (changing the referencing channel), filtering the signal (by applying a bandpass filter to remove environmental noise or to remove activity in nonrelevant frequencies), epoching (extracting a time epoch typically surrounding the stimulus onset), trial/channel rejection (to remove those containing artifacts), etc. See [17] for further information. In RSVP-EEG, a common average reference or mastoid reference is often used. A bandpass filter (e.g. 0.1–30 Hz) is commonly applied in RSVP-EEG. The EEG signal is preferably analyzed as epochs (i.e., the whole EEG data is cut by using a fixed time window (e.g., 0–1,000 ms) corresponding to each trigger onset) and each segment is named as an epoch. These epochs can then be used for analysis (e.g., feature extraction, classification).

The presence of many artifacts such as those related to muscle movements in the EEG signal can be sometimes removed by using a bandpass filter as the frequencies of interest in RSVP-EEG do not always detrimentally overlap. During RSVP-EEG experiments, it can be very common for eye-blink behavior to occur in response to target images. This perhaps arises as a result of the participant withholding eye-blinks until a target is seen. While this may be favorable for improving the detection rate of targets, without inspection of the data it may lead to erroneous conclusions on what discriminative information is actually driving the performance of a classifier. One common strategy to investigate this involves identifying a spatial component in the EEG signal related to eye-blinks via ICA to determine if it is trial-locked to targets in any way. Additionally, ICA allows for such activity to be in part attenuated. An investigation of commonly used strategies (and subtle pitfalls) can be found in [18].

Before extracting features from RSVP-EEG data, some critical properties of EEG signals have to be considered concerning the design of an RSVP-based BCI system:

- Low SNR: EEG in noninvasive BCI has an inherently poor SNR and task-related ERPs are typically overwhelmed by strong ongoing EEG background activity in single trials;

- Curse of dimensionality: In RSVP-based BCI system, EEG data can have high-dimensionality spanning space and time. However, typically limited training sets are available especially considering the target image class which are usually infrequent;
- Overlapping epochs: There is substantial overlap between adjacent target epochs and standard epochs because of the short interstimulus interval used in the RSVP paradigm;
- Imbalanced datasets: Target images are overwhelmed by standard images in an RSVP application which leads to an imbalanced classification problem.

These critical properties have to be considered before feature extraction. The last one can be overcome through cost-sensitive learning [19], while the first three are inherent challenges in the design of an RSVP-based BCI system.

Two main differences between the RSVP-EEG paradigm and other ERP paradigms are that the former requires single-trial detection in the presence of overlapping epochs. Traditional ERP analysis typically computes a grand average ERP where phase-locked activity in the signal remains after averaging, whilst other non-locked background activity increasingly attenuates as more trials are averaged. For example, the P300 speller is an ERP paradigm that has been a benchmark for P300 BCI systems. In this paradigm, each desired symbol is spelt several consecutive times by a participant where the epochs corresponding to each row/column are averaged over the trials. This averaging process is able to improve the EEG SNR for the system because averaging reduces the contribution of random background EEG oscillations [20]. This repetition of an image stimulus is not always applicable in the RSVP-EEG paradigm because it can introduce unintentional behaviors such as a participant attending to an image due it being a salient repetition rather than it being a target. In single trial detection, low SNR is a challenge for the detection of discriminating ERP activity. Furthermore, the overlapping epoch's problem may contribute to overfitting when training a machine-learning model [21]. In summary, low SNR and overlapping epochs are two challenging problems for RSVP-EEG when compared to other ERP paradigms.

12.2.4 Performance evaluation metrics

A machine-learning model's performance can be evaluated by a variety of evaluation metrics. Area under the receiver operating characteristic (ROC) curve (AUC) is widely used as it illustrates the discriminative ability of a binary classifier system as its discrimination threshold is varied [22]. One may want to adjust this threshold for example to optimize for fewer false positives at the cost of more false negatives. ROC-AUC is the most widely used evaluation metric in RSVP-EEG research [1,2,7]. However, ROC-AUC score may not be suitable when evaluating some real-world systems because it gives an unified measure of the performance of a classifier across all potential thresholds and in effect sidesteps the issue of the impact of threshold selection.

Balanced accuracy (BA) is well suited for evaluating RSVP-EEG systems that utilize binary classifications [23]. BA can be calculated as below:

$$BA = \frac{1}{2}\,(sensitivity + specificity) \tag{12.1}$$

where $sensitivity = (TP/(TP + FN))$ and $specificity = (TN/(TN + FP))$.

Choosing evaluation metrics critically depends on the application. For example, if the classification system is used to rank target above standard images, then AUC would be the preferred evaluation metric. If the classification system is designed to give a binary classification (target vs standard), then BA can be a good evaluation metric. Both ROC-AUC and BA are robust to targets/standards ratio imbalances in dataset.

12.3 Feature extraction methods used in RSVP-based BCI research

The challenge for feature extraction methods is to find intrinsic characteristics of the EEG signals that relate to certain cognitive responses. Feature extraction in BCI systems plays an important role since it can greatly affect the SNR and the classification strategy used, which in turn determines the performance of the BCI.

This section focuses on extracting RSVP-EEG from three aspects: (1) spatial filtering (supervised and unsupervised); (2) time-frequency representation; (3) other feature extraction methods.

12.3.1 Spatial filtering

12.3.1.1 Supervised spatial filtering

As mentioned in the previous section, RSVP-EEG data suffers from low SNR and often high spatial dimensionality. Spatial filtering is an efficient technique for mitigating these concerns. In the area of BCI research, xDAWN [24], beamformer [25] and common spatial pattern (CSP) [26] are widely used for generating supervised spatial filters. In this chapter, we focus on three methods for generating spatial filters: xDAWN, CSP and LDA beamformer. For xDAWN, the goal is to maximize the signal-to-signal-plus-noise ratio (SSNR), whereas for CSP, the goal is to maximize the ratio between the discriminative activity and the common activity, leading to optimal variances for the discrimination of two types of stimulus EEG signals. LDA beamformer is used for source signal reconstruction where it maximizes the SNR.

Problem formulation: Let $\mathbf{X} \in \mathbb{R}^{C\times T}$ be EEG epochs corresponding to each image stimulus, where C is channel number and T is epoch time length. The problem of spatial filtering is to find a set of projection vectors (each comprised of weights for each channel) $\mathbf{w} \in \mathbb{R}^{C\times n}$ (n is the number of components) to project $\mathbf{X}$ to a subspace, where $\mathbf{w}$ is calculated by different algorithms, i.e., xDAWN, beamformer, CSP, etc.

$$\mathbf{X}_{\text{sub}} = \mathbf{w}'\mathbf{X} \tag{12.2}$$

CSP: CSP generates sets of channel weightings that can be used to project multichannel EEG data into a low-dimensional subspace, where this transformation can maximize the variance of two-class signal matrices. Let $X_+(i) \in \mathbb{R}^{C\times T}$ and $X_-(i) \in \mathbb{R}^{C\times T}$ be the ith event locked EEG epochs (C is the channel number and T is the time length) in two experimental conditions, i.e., $X_+(i)$ for the target image condition and $X_-(i)$ for the standard image condition. Covariance matrices in the two conditions can be estimated as

$$\mathbf{\Sigma}_c = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{X}_c(i)\mathbf{X}_c'(i)}{trace(\mathbf{X}_c(i)\mathbf{X}_c'(i))} \quad (c \in \{+,-\}) \tag{12.3}$$

where "′" denotes the matrix transposition and $\mathbf{\Sigma}_c \in \mathbf{R}^{\mathbf{C}\times\mathbf{C}}$. The CSP optimization problem can be formulated as

$$\{\max,\min\}_{\mathbf{w}\in\mathbb{R}^C} \quad \frac{\mathbf{w}'\mathbf{\Sigma}_+\mathbf{w}}{\mathbf{w}'\mathbf{\Sigma}_-\mathbf{w}} \tag{12.4}$$

This optimization problem is given by the simultaneous digitalization of the two covariance matrices. This can be achieved by solving the generalized eigenvalue problem:

$$\mathbf{\Sigma}_+\mathbf{w} = \lambda\mathbf{\Sigma}_-\mathbf{w} \tag{12.5}$$

Note: The objective of CSP is to maximize the variance in one class while minimizing the variance in the other class. Maximizing the variance in this way corresponds to maximizing the frequency-power of target-related activity in the signal. When using CSP, multiple spatial filters will be obtained and crossvalidation is normally deployed to choose a spatial filter(s). This strategy can be adapted to use multiple different band-passed versions of the same EEG signal epoch to leverage different sources of discriminative information present across different frequencies (that often also differ in spatial characteristics). CSP is widely applied in motor imagery-based BCI [27]. In Yu's work, CSP has been applied for producing spatial filters for RSVP-based BCI [8].

xDAWN: The xDAWN algorithm has been successfully applied in the P300 speller BCI application [24]. The basic goal of xDAWN is to enhance the SSNR of the responses corresponding to the target stimulus. Let recorded signals be $\boldsymbol{X} \in \mathbb{R}^{N_t\times N_s}$, where N_t is the time length of recorded EEG signals and N_s is the number of channels. It considers the following model:

$$\mathbf{X} = \mathbf{DA} + \mathbf{H} \tag{12.6}$$

where $\mathbf{D} \in \mathbb{R}^{N_t \times N_e}$ (N_e is the number of temporal samples of ERP corresponding to the target stimulus) is the real Toeplitz matrices and $\mathbf{A}$ is the ERP response to the target. $\mathbf{D}$ has its first column elements set to zero except for those that correspond to a target stimulus onset and $\mathbf{H}$ is the on-going EEG activity.

The problem statement for xDAWN becomes how to estimate the spatial filter for (12.6) such that the synchronous response is enhanced by spatial filtering:

$$\mathbf{XU} = \mathbf{DAU} + \mathbf{HU} \tag{12.7}$$

where $\mathbf{U} \in \mathbb{R}^{N_s \times N_f}$ (N_f is the number of spatial filters). The optimized solution can be achieved by

$$\hat{\mathbf{U}} = \arg\max_{\mathbf{U}} \frac{Trace(\mathbf{U}'\hat{\mathbf{A}}'\mathbf{D}'\mathbf{D}\hat{\mathbf{A}}\mathbf{U})}{Trace(\mathbf{U}'\mathbf{X}'\mathbf{X}\mathbf{U})} \tag{12.8}$$

where $\hat{\mathbf{A}}$ is the least squares estimation of response $\mathbf{A}$. More details about the computation method can be found in [24].

> **Note:** Separately from CSP, the numerator in xDAWN in (12.8) is the ERP response rather than target EEG epochs, i.e., ERP response being the mean value of target EEG epochs. xDAWN aims to enhance the SSNR of the response corresponding to the target stimulus and it is originally designed for enhancing the P300 evoked potential for the P300 speller BCI [24]. Similar to CSP, xDAWN generates multiple spatial filters as well. It is suggested to use crossvalidation to determine the number of spatial filters. In recent published work, xDAWN has been applied to RSVP-based BCI for spatial filtering [28].

LDA beamformer: LDA beamformer has been proposed to maximize the SNR of EEG in a way which is robust to correlated sources [29]. The generation of a spatial filter using LDA beamformer is comprised of three steps: (1) spatial pattern estimation; (2) covariance matrix estimation; (3) spatial filter optimization. Let column vectors $\mathbf{p}_1 \in \mathbb{R}^{C\times 1}$ and $\mathbf{p}_2 \in \mathbb{R}^{C\times 1}$ be the spatial pattern of a specific component in two different experimental conditions, where C is the number of channel. We denote the difference pattern as $\mathbf{p} := \mathbf{p}_1 - \mathbf{p}_2$ and the covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{C\times C}$. The optimization problem for the LDA beamformer can be stated as

$$\begin{aligned} &\underset{\mathbf{w}}{\text{minimize}} \;\; \mathbf{w}'\mathbf{\Sigma}\mathbf{w} \\ &\text{s.t.} \;\; \mathbf{w}'\mathbf{p} = 1 \end{aligned} \tag{12.9}$$

and the optimized solution can be determined as

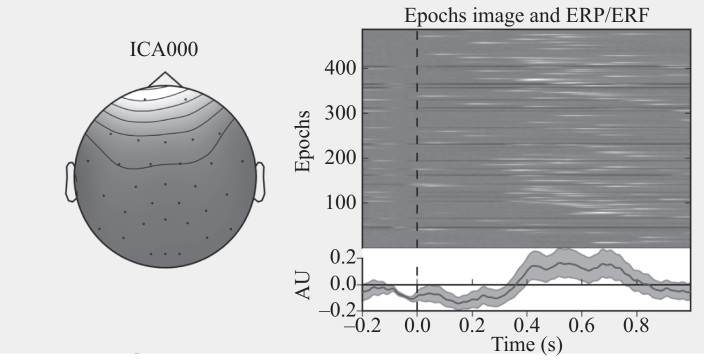
$$\mathbf{w} = \mathbf{C}^{-1}\mathbf{p}\left(\mathbf{p}'\mathbf{C}^{-1}\mathbf{p}\right)^{-1} \tag{12.10}$$

> **Note:** Separately from the previous two methods, the LDA beamformer method generates an optimal spatial filter (only one spatial filter) that maximizes the SNR. One optimal projection vector may not be able to fully capture all available information from the original EEG epoch due to the sources of variability such as different spatial characteristics of early and late target-discriminative ERPs across tasks and participants. Therefore, multiple time window LDA beamformers (where the researcher trains the LDA beamformer in multiple time windows) are often applied for improved performance with RSVP-based EEG data.

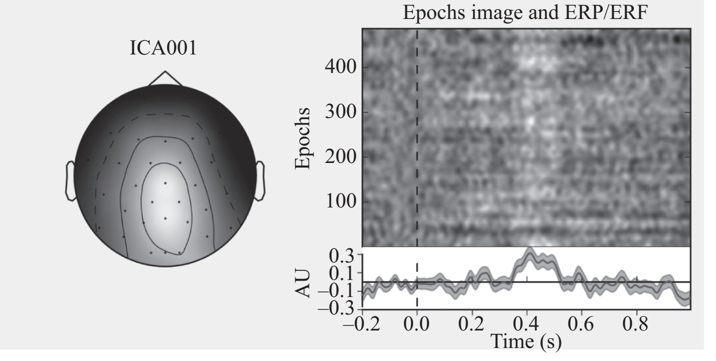
So far we have introduced three supervised approaches to generate useful spatial filters. After applying the spatial filter(s) to the original epoch, it can be appreciated that the dimensionality of the projected subspace has been reduced significantly and that this subspace signal may have different properties (optimized SSNR for xDAWN, optimized SNR for LDA beamformer, maximum difference of variance between two classes for CSP) depending on which algorithm has been used for generating the spatial weightings. This projected subspace can then be used as the basis of a feature set for training a practical classifier. It is worth noting that the overall effect of employing spatial filtering methods is an improvement in the SNR, a reduction in computation cost and a more favorable situation for many classification algorithms that suffer issues with high-dimensional feature vectors (particularly when few training examples are available). These are desirable properties of a signal processing pipeline for RSVP-based BCI.

12.3.1.2 Unsupervised spatial filtering

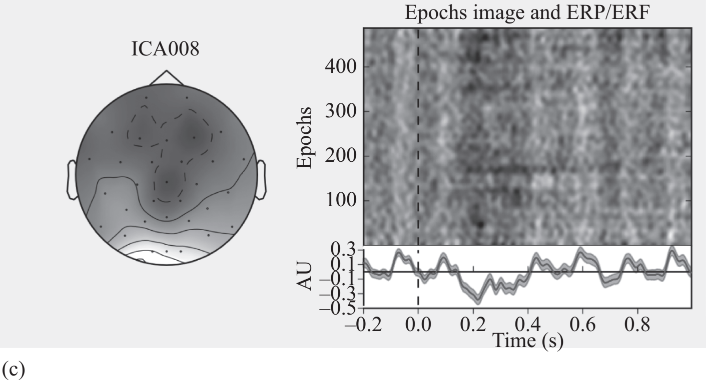
ICA: EEG source activity refers to the time-varying far-field potentials arising within an EEG source and volume-conducted to the scalp electrodes. The recorded EEG signals are then, according to this interpretation, the summation of neural activity, contributions of nonbrain sources such as scalp muscle, eye movement and cardiac artifacts, plus (ideally small) electrode and environmental noise [30]. Successful separation of contributions from these nonneural activity related sources can improve the SNR of the signals of interest. ICA is a technique that can aid here and can be used to find linear representations such that time-series signals obtained via its components' projections are statistically independent from each other (or as independent as possible). Such a representation is capable of capturing the essential structure of the data in many applications, including feature extraction and signal separation [31]. Essentially, ICA produces a matrix of spatial filters. ICA has been widely applied to EEG signal fields for denoising [32] and artifact removal [33]. Figure 12.6 illustrates the characteristics of three independent components (ICs) corresponding to eyeblinks, P300 activity and other trial locked ERP activity, respectively. From the IC plots (left), it is noticeable that different signal sources have different IC localizations, i.e., eyeblink is distributed frontally while the P300 is localized posteriorly. From an ERP image and



(a)



(b)



(c)

Figure 12.6 Examples of ICA components (left) and ERP images (right) in Section 12.3.1.2 for eye blink related activity (a), posterior P300 activity (b) and other trial-locked EEG activity (c). The results are generated using one participant's dataset (from NAILS) and only includes only RSVP target related trials

ERP time series plot (right hand side top and bottom plots), it can be seen that eyeblink activity and P300-related activity are locked to target stimuli. Moreover, it can be seen that the eyeblink time-locked activity is noticeably prominent at around 400 ms which is close to the time region where we also see P300 activity. This indicates that this participant sometimes blinked their eyes when presented with a target image likely as a result of an active effort to suppress eyeblinks up to that point in case they missed a target. Eyeblink artifacts in such instances can potentially be beneficial for the classification process. However, we are going to remove these as we only consider signals of direct neural origin in this exposition. It is worth emphasizing here that ICA successfully resolved the signal into a distinct physiologically interpretable source in this instance. Published work by Bigdely-Shamlo *et al.* demonstrated ICA successfully applied to RSVP-based BCI generating ICs and independent time course templates (ITs) for each IC. These ITs were selected as features for training the classifier [1] quite successfully.

Principal component analysis: Principal component analysis (PCA) is a statistical technique which uses eigenvalue decomposition to convert a set of correlated variables into a set of linearly uncorrelated variables where each of the resultant variables is referred to as a principal component [34]. For multivariate datasets, notably data in a high-dimensional space, PCA can be particularly effective for dimensionality reduction. PCA has been applied in EEG signal analysis for dimensionality reduction [35] and the production of spatial filters [36]. In RSVP-based BCI literature, PCA has only been applied for feature dimension reduction to date [1,37].

12.3.2 Time-frequency representation

Feature extraction in BCI can be achieved in the time domain, the frequency domain and the combined time-frequency domain. In the time domain, time regions coinciding with ERPs such as the P300 are used when extracting features for single-trial event detection [3]. Frequency domain features such as the amplitudes of μ (8–13 Hz) and β (14–26 Hz) are widely used in sensorimotor control BCI as it has been shown changes occur in these when a participant imagines (or engages) in certain types of movements. However, frequency domain features have not been used in the image RSVP paradigm in the literature to date. Time-frequency representations can be generated using methods such as short-time Fourier transform (STFT) and wavelet transforms. The wavelet transform method is often preferred over STFT in many instances as it produces a time-frequency decomposition of a signal over a range of characteristic frequencies that separates individual signal components more effectively than the STFT method does. A number of other properties of the source signal such as stationarity assumptions should be considered when utilizing one approach over another [38].

A set of common types of time-frequency features found in RSVP-based EEG has been proposed in Meng's work [39]. In Figure 12.7, we show the time-frequency mean power generated using target EEG epochs from the NAILS dataset. It can be seen that both high power and strong ITC appear in low frequencies (0–6 Hz) and in two time

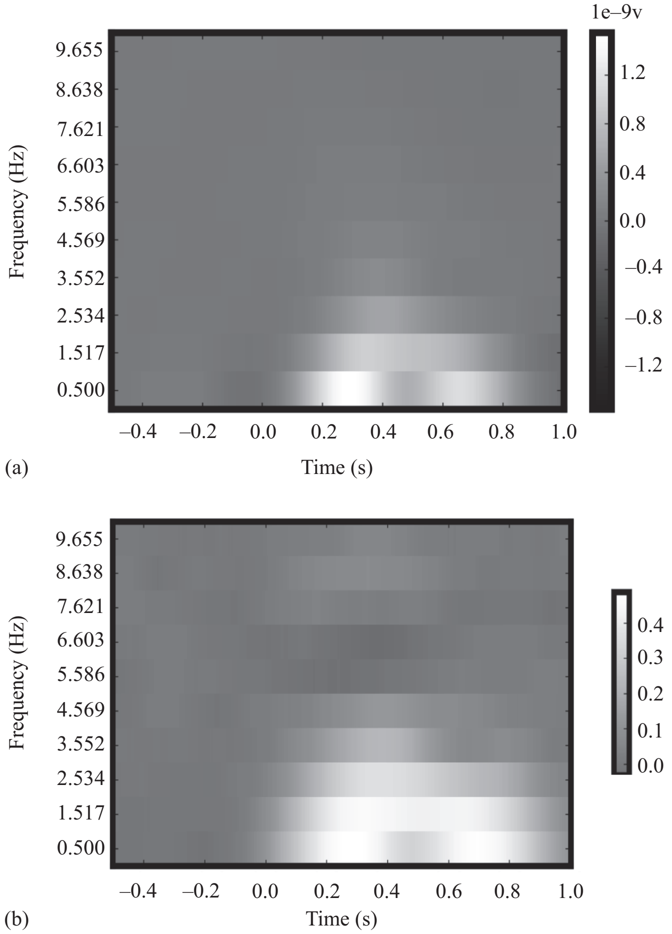


Figure 12.7 Time-frequency representation example corresponding to target images at the Pz channel for averaging nine participants in an RSVP experiment using Morlet wavelet transform in Section 12.3.2, the EEG signal has been band-passed between 0.1 and 30 Hz. (a) Time-frequency mean power and (b) intertrial coherence.

regions (200–400 and 600–800 ms). The discriminative ERSP-related activity appears in both an early time region and a later time region. The time-frequency representation strongly depends on the type of the image stimulus and the experimental environment.

12.3.3 Other feature extraction methods

In Huang's work, EEG signals from the stimulus onset to 500 ms poststimulus were extracted for each channel and concatenated to form a feature vector [7]. This resulted in each trial containing 32 $\times$ 129 features (where 32 is the number of channels and 129 is the number of time points). This strategy of building feature vectors as a concatenation over time regions (and channels) of interest is commonly found in the RSVP literature and often yields good results. Hierarchical discriminant component analysis has been proposed in [40], where this method estimates EEG signatures of target detection events using multiple linear discriminators, each trained at a different time window relative to the image onset. Since EEG signals contain both spatial and temporal information, a spatiotemporal representation for RSVP-based EEG data has been proposed by Alpert [37]. This representation is divided into two steps: (1) LDA is applied at each timestamp to produce the spatial weights and a spatial weight matrix is then used for mapping original epoch to a new space, and (2) PCA is then used for dimensionality reduction based on the temporal domain, i.e., for each independent channel.

12.3.4 Summary

Feature extraction is an essential step when designing a BCI system because pertinent features can significantly improve performance of the resulting classifier and additionally it can significantly reduce the computational cost. In RSVP-based BCI, discriminative ERP-related activity often occurs in both early and late time region. Feature extraction for RSVP-based BCI is best designed by considering these ERPs' properties.

12.4 Survey of classifiers used in RSVP-based BCI research

This section surveys the classifiers used for recognition of target and standard events in RSVP-based BCI systems. Due to the fact that the nonlinear problem has not been well explored in RSVP-based EEG data in the literature so far, this section is divided into linear and NN classifiers. Since deep-learning technology is very popular currently in other application domains such as computer vision and natural language processing, we introduced some deep-learning methods in the NN section.

12.4.1 Linear classifiers

Linear classifiers are widely used for designing BCI applications due to their good performance, often simple implementation and low computational complexity. Four main linear classifiers will be in introduced in this section, namely, LDA, BLR, LR and SVM. In this section, we consider our model as

$$y = \mathbf{w}'\mathbf{x} + b \tag{12.11}$$

where y is classifier output, $\mathbf{x}$ is the feature vector and b is the threshold.

12.4.1.1 Linear discriminant analysis

LDA is a supervised subspace learning method which is based on the Fisher criterion, and it is equivalent to least squares regression (LSR) if the regression targets are set

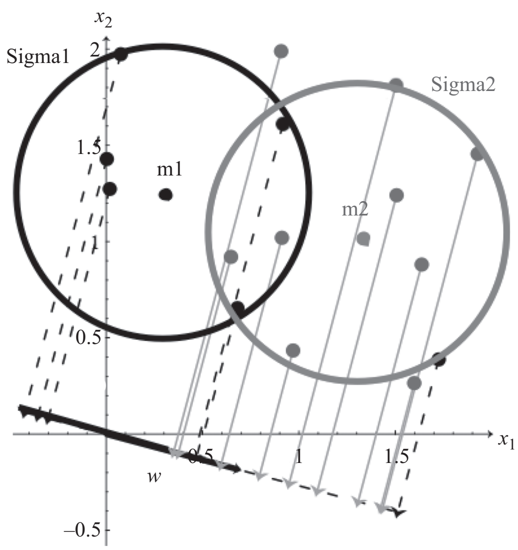


Figure 12.8 Projection of two different classes onto a line by LDA in Section 12.4.1.1 [43]

to N/N_1 for samples from class 1 and $-(N/N_2)$ for samples from class 2 (where N is the total number of training samples, N_1 is the number of samples from class 1 and N_2 is the number of samples from class 2) [41]. It aims to find an optimal linear transformation **w** that maps **x** to a subspace in which the between-class scatter is maximized while the within-class scatter is minimized in that subspace. The optimization problem for LDA is to maximize the cost function as below:

$$J = \frac{\mathbf{w}'\mathbf{S}_B\mathbf{w}}{\mathbf{w}'\mathbf{S}_W\mathbf{w}} \tag{12.12}$$

where $\mathbf{S}_B$ is the between-class scatter and $\mathbf{S}_W$ is the within-class scatter. Regularization is often applied in order to avoid the singular matrix problem of $\mathbf{S}_W$ [42]. Figure 12.8 shows the LDA implementation on two different classes with equal covariance and different mean values. The solid line is the projected subspace **w** where these two classes will be projected by LDA. This transformation enables the best separation between two classes on the subspace **w**. Details of the method can be found in Duda's book [43].

LDA has very low computational complexity which makes it suitable for online BCI systems. As mentioned earlier, classification of RSVP-based EEG data suffers from the imbalanced data set problem. In Xue's work [44], he showed that there is no reliable empirical evidence to support that an imbalanced data set has a negative effect on the performance of LDA for generating the linear transformation vector. Consequently, LDA is suitable and has been successfully used in RSVP-based BCI

[1,39]. LDA typically suffers from issues particularly when dealing with data sources that contain issues such as outliers. For this reason, regularization strategies are typically employed but are not covered here [42].

12.4.1.2 Bayesian linear regression

BLR, also named Bayesian linear discriminant analysis, can be seen as an extension of LDA or LSR. In BLR, regularization for parameters is used for preventing overfitting caused by high dimensional and noisy data. BLR assumes the parameter distribution and target distribution are both Gaussian [41]. We introduce LSR as a starting point for the description of BLR. Given the linear model in (12.11), the solution of LSR can be stated as

$$\mathbf{w} = (\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{X}\mathbf{y} \tag{12.13}$$

Note that $y = (N/N_1)$ for class 1 and $y = -(N/N_2)$ for class 2 here (threshold can be determined by adding a column with all one as the first column in $\mathbf{X}$). LSR does not consider the parameter distribution in this case, and it maximizes the likelihood. For BLR, it considers the parameter distribution and maximizes the posterior. Given the prior target distribution $p(y) \sim \mathcal{N}(\mu, \beta^{-1})$ and parameter distribution $p(\mathbf{w}) \sim \mathcal{N}(0, \alpha^{-1}\mathbf{I})$ (where β and α are the inverse variance), BLR gives the optimized estimation for the parameter:

$$\mathbf{w} = \beta(\beta\mathbf{X}\mathbf{X}^{\mathrm{T}} + \alpha\mathbf{I})^{-1}\mathbf{X}\mathbf{y} \tag{12.14}$$

It can be seen that the optimization of BLR is added with the prior information of parameter and data. Hence, the optimization depends on the hyperparameters β and α. In real-world applications, the hyperparameters can be tuned using crossvalidation or the maximum likelihood solution with an iterative algorithm [41,45]. BLR has been proven to have very good performance in BCI research [46,47].

12.4.1.3 Logistic regression

LR models the conditional probability as a linear regression of feature inputs. Considering the linear regression model of RSVP-based EEG data in (12.11), the logistic model can be constructed as

$$p(\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}'\mathbf{x}+b}} \tag{12.15}$$

The optimization problem of an LR can be constructed by minimizing the cost function as below:

$$J(\mathbf{w}, b) = -\frac{1}{m}\sum_{i=1}^{m}[y_i \log(p(\mathbf{x}_i)) + (1 - y_i)\log(1 - p(\mathbf{x}_i))] \tag{12.16}$$

where $y \in \{0, 1\}$ and m is the sample number of two classes. Similar to SVM mentioned earlier, LR can be modified by penalizing different cost terms to each class and the cost function can be modified as below:

$$J(\mathbf{w}, b) = -\frac{1}{n_0 + n_1} \sum_{i=1}^{n_0+n_1} [n_0 y_i \log(p(\mathbf{x}_i)) + n_1(1 - y_i) \log(1 - p(\mathbf{x}_i))] \quad (12.17)$$

where n_0 and n_1 are the numbers of standard and target image clips, respectively.

LR is part of a broader family of generalized linear models (GLMs), where the conditional distribution of the response falls in some parametric family, and the parameters are set by the linear predictor. LR is the case where the response is binomial and it can give the prediction of the conditional probability estimation. LR is easily implemented and has been successfully applied for RSVP-based BCI research [48,49].

12.4.1.4 Support vector machine

An SVM aims to select the hyperplane which maximizes the margins (i.e., the distance from the nearest training samples). In order to overcome the imbalanced classification problem, a weighted support vector machine is proposed [50]. An SVM can be used for linear and nonlinear classification by using the "kernel trick." This consists of mapping data to other spaces using a kernel function $\kappa(x_i, x_j)$. For linear classification, kernel function can be chosen as $\kappa(x_i, x_j) = \langle x_i, x_j \rangle$. For nonlinear classification, Gaussian kernel, $\kappa(x_i, x_j) = \exp(-(\|x_i - x_j\|^2/2\delta^2))$, is widely used in the classification area.

SVM has a small number of hyperparameters which need to be tuned manually and this is often done by using crossvalidation. There is already considerable use of SVMs in RSVP-based BCI research [8,51,52].

12.4.1.5 Other machine-learning algorithms

Here, we have introduced four linear classification algorithms that are widely used in the RSVP-based BCI research area. There are numerous machine-learning algorithms available currently, and we encourage readers to experiment with them. Scikit-learn is a machine-learning library in Python, and it provides lots of machine-learning algorithm implementations. Linear classification methods can be found in the GLMs list.

> **Note:** For the four classifiers above, every method involves hyperparameters except LDA (unless using a regularized version that relies on some parameter selection). We suggest using grid search or maximum likelihood estimation to determine these hyperparameters. K-fold validation can be used for the evaluation of each set of tuned hyperparameters. Importantly, classifier performance should be evaluated on a withheld testset such as an experimental block that does not overlap with the data used for model training or hyperparameter selection.

12.4.1.6 Summary

To summarize, we have introduced four types of linear classifier from the viewpoints of different optimization objectives. LDA aims to find the subspace which gives the best separation between two classes after projection. BLR uses prior information about data distribution and weights, constraining the estimated weights closer to zeros, which helps to stop overfitting. SVM finds the optimal hyperplane maximizing the margins and it can be applied to nonlinear cases by using the "kernel trick." LR can predict the conditional probability. We recommend using LDA (a regularized form) and BLR for RSVP-based BCI because of their low computational cost and good performance. Recent work in RSVP-based BCI research shows that BLR outperforms LDA and SVM [47].

12.4.2 Neural networks

NNs is yet another category of classifiers that are increasingly used in BCI research. NN comprises several artificial neurons that can enable nonlinear decision boundaries.

This section is divided into two parts. The first describes the multilayer perceptron (MLP), the most widely used NN, and then some deep-learning techniques are introduced.

12.4.2.1 Multilayer perceptron

An MLP is minimally comprised of three layers of neurons, namely an input layer, one or several hidden layers and an output layer [53]. In each hidden layer, each neuron is connected to the output of each neuron in the previous layer and its output is the input of each neuron in the next layer. Considering RSVP-based BCI, there is only one output in the output layer. Parameters in MLP can be updated by the BackPropagation algorithms which involve computing partial derivatives to update parameters in the direction of a gradient which decrease the overall loss function [54]. A number of what are called gradient descent optimization algorithms exist for this purpose [55].

NN and MLP are very flexible classifiers that can be applied to a great variety of problems because they are universal approximators [5]. Hence, MLP can be applied for almost all machine-learning problems including binary classification or multiclass classification or modeling. The main disadvantage of MLP is overfitting [56] due to the limited training samples especially for targets in RSVP-based EEG data. Therefore, one has to be careful when designing a MLP architecture and regularization is often required [57].

12.4.2.2 Some deep-learning techniques

Modern deep learning provides a powerful framework for supervised learning [58]. With more layers and more neurons in layers, a deep network can represent increasingly complex nonlinear patterns, and it has been successfully applied to many fields including computer vision [59], natural language processing [60], etc.

Since deep-learning implementations in the area of EEG is still rare, we will introduce three representative methods in the deep-learning field, namely, convolutional neural networks (CNNs), recurrent neural networks (RNN) and deep belief nets (DBN).

Convolutional neural networks

CNN is a type of NN that employs a mathematical operation called convolution specialized for processing a grid of values where the arrangement of the values is not arbitrary such as is the case with an image where pixels near to any one pixel tend to be correlated in a meaningful way [61]. Convolutional networks are simply NNs that use convolutions in place of general matrix multiplications in at least one of their layers [58]. CNN has been tremendously successful in many practical applications [62–64]. CNN leverages three properties that improve learning, namely sparse interactions, parameter sharing and equivariant representations [58]. Sparse interactions are accomplished by the convolution operation while choosing the kernel smaller than the input size. This property enables meaningful features to be extracted from input data. Parameter sharing refers to the fact that each member of the kernel is used at every position of the input with the same parameters. Equivariant representations means that if the input changes, the output changes in the same way [58].

Since a CNN is capable of extracting features from the input data automatically, CNN has become the most widely used deep-learning architecture in RSVP-EEG research [28,65–67]. All of these works have shown that CNN is effective in combining the spatial filtering and the classification steps in an unified way. CNN is also the winning solution to the NAILS competition in the 13th NTCIR conference [21] which shows better performance than traditional methods.

Recurrent neural networks

Different from CNN specialized for processing a grid of values, an RNN is a family of NNs which is designed for processing sequential data such as speech signals [68]. RNN processes the sequence which contains vectors $\mathbf{x}(t)$ with the time step index t ranging from 1 to τ [58].

There are several implementations of RNN in EEG signal analysis and classification [69–71]. However, RNN implementations on RSVP-based BCI have not been stated in the literature thus far. Since EEG is sequential and the P300 has a temporal property, it is an open question if an RNN can be used effectively in RSVP-based BCI.

Deep belief nets

A deep belief network (DBN) is a generative graphical model which comprises multiple layers of latent variables (“hidden units”), with connections between the layers but not between units within each layer [72]. It provides an efficient way to learn a multiple-layered restricted Boltzmann machine [73].

A DBN has shown efficacy in RSVP-based BCI from Ahmed’s work and it can extract discriminant features from RSVP-based EEG data as well [74].

12.4.2.3 Summary

In this section, we have introduced the use of NN techniques for classification. The first part introduced an MLP framework which is the most classically used NN framework. The MLP has a simple implementation and it is flexible but easily suffers from issues related to overfitting.

Deep learning is very popular and has had great success in many real-world applications but often requires very large volumes of data. With better data acquisition and more advanced generative models [75], it is possible to train better deep network models for RSVP-EEG leveraging very large datasets.

The main difficulty with EEG signals is the potentially very large feature dimensionality when considering all combinations of channel, frequency and time features. This makes feature extraction a very complex step that must cater to inter-subject and intertask variability. In traditional RSVP-based BCI systems, feature extraction and classification are always separated. Deep learning provides a potentially unified way to accomplish this.

12.5 Conclusion

In this chapter, we have introduced the main parts in a typical RSVP-based BCI system including RSVP data acquisition, data preprocessing, feature extraction and classification. We focused on the machine-learning architecture (feature extraction and classification) as it is the most important part of a typical BCI system.

We have shown that discriminative patterns in RSVP-based EEG data can appear in different time regions (both early and late), i.e., the P300 is not the only ERP being elicited in the RSVP paradigm. Therefore, we suggest designing a feature extraction method that takes into account the properties of the discriminative ERPs for a given task. A good feature extraction method can not only improve the classifier's performance in the later stage but also reduce the computational cost. In this chapter, we introduce existing feature extraction methods used in the literature so far. We stated the objective of each method and a direct comparison between those methods will be part of future work.

The other part of the machine-learning architecture in an RSVP-EEG BCI system is the classifier. We divided our discussion on classifiers into both linear classifiers and NNs. The choice of the classifier remains difficult and depends mainly on the number of available trials and feature vector dimensionality. Linear classifiers remain popular as they have low computational complexity, are easy to implement and have good performance in classification accuracy. NNs possibly outperform linear classifiers with a large number of trials as deep NNs are able to capture high level features related to the variability of the EEG signals across participants and over time. However, acquiring EEG data is time consuming and the variability in the EEG of a specific participant can change over time, which indicates that the number of trials in RSVP-based BCI is limited. Therefore, the choice of classification method should be capable of training a model with a limited amount of available data. In this aspect, linear classifiers are more preferable than NNs due to fewer parameters in the model which in turn can help to prevent overfitting on the noisy and limited RSVP-based EEG data. Here, we suggest to use LDA and BLR in RSVP-based BCI research as they are easy to implement, efficient and have good performance.

The area of RSVP-EEG stretches back well over a decade and there has been significant progress in this time. With the emergence of deep-learning approaches, computer vision recognition applications are able to perform at or even above a human level raising questions about whether people are still needed to perform image labeling tasks. We believe that when labeled image datasets are limited, these computer vision system may not perform very well as a typical component to their success is the availability of very large labeled image datasets. In this way, RSVP-EEG may assist in more efficiently labeling large datasets of image content to support this process. Similarly, many image labeling tasks may require subjective (or expert) knowledge about the image that cannot be easily learned by a deep-learning architecture but that may be readily detected when using an RSVP-EEG system. We see these systems as being able to work in a synergistic manner rather than competitively.

Acknowledgment

This work is funded as part of the Insight Centre for Data Analytics which is supported by Science Foundation Ireland under Grant Number SFI/12/RC/2289.

References

[1] Nima Bigdely-Shamlo, Andrey Vankov, Rey R Ramirez, and Scott Makeig. Brain activity-based image classification from rapid serial visual presentation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(5):432–441, 2008.

[2] Adam D Gerson, Lucas C Parra, and Paul Sajda. Cortically coupled computer vision for rapid image search. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):174–179, 2006.

[3] Paul Sajda, Adam Gerson, and Lucas Parra. High-throughput image search via single-trial event detection in a rapid serial visual presentation task. In *Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on*, pages 7–10. IEEE, 2003.

[4] John Polich. Updating p300: an integrative theory of p3a and p3b. *Clinical Neurophysiology*, 118(10):2128–2148, 2007.

[5] Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 4(2):R1, 2007.

[6] Klaus-Robert Müller, Matthias Krauledat, Guido Dornhege, Gabriel Curio, and Benjamin Blankertz. Machine learning techniques for brain–computer interfaces. *Biomedical Technology*, 49(1):11–22, 2004.

[7] Yonghong Huang, Deniz Erdogmus, Misha Pavel, Santosh Mathan, and Kenneth E Hild Ii. A framework for rapid visual image search using single-trial brain evoked responses. *Neurocomputing*, 74(12–13):2041–2051, 2011.

[8] Ke Yu, Kaiquan Shen, Shiyun Shao, Wu Chun Ng, Kenneth Kwok, Xiaoping Li. Common spatio-temporal pattern for single-trial detection of event-related potential in rapid serial visual presentation triage. *IEEE Transactions on Biomedical Engineering*, 58(9):2513–2520, 2011.

[9] Swartz Center for Computational Neuroscience UCSD. Lab streaming layer [online], available from http://github.com/sccn/labstreaminglayer/wiki. 2016.

[10] Zhengwei Wang, Graham Healy, Alan F Smeaton, and Tomas E Ward. An investigation of triggering approaches for the rapid serial visual presentation paradigm in brain computer interfacing. In *Signals and Systems Conference (ISSC), 2016 27th Irish*, pages 1–6. IEEE, 2016.

[11] Henri Begleiter. *Evoked brain potentials and behavior*, volume 2. New York: Springer Science & Business Media, 2012.

[12] Anthony M Norcia, Lawrence Gregory Appelbaum, Justin M Ales, Benoit R Cottereau, and Bruno Rossion. The steady-state visual evoked potential in vision research: a review. *Journal of Vision*, 15(6):4–4, 2015.

[13] Graham Healy, Zhengwei Wang, Cathal Gurrin, Tomás E Ward, and Alan F Smeaton. *An EEG image-search dataset: a first-of-its-kind in IR/IIR*. In Proceedings of CHIIR Workshop on Challenges in Bringing Neuroscience to Research in Human-Information Interaction, Oslo, Norway, March 2017.

[14] Graham Healy, Tomás Ward, Cathal Gurrin, and Alan F Smeaton. Overview of NTCIR-13 NAILS task. In *The Thirteenth NTCIR Conference (NTCIR-13)*, Tokyo, Japan, 5–8 Dec 2017.

[15] Carmen Vidaurre and Benjamin Blankertz. Towards a cure for BCI illiteracy. *Brain Topography*, 23(2):194–198, 2010.

[16] Brian F Odonnell and Ronald A Cohen. *The N2-P3 complex of the evoked potential and human performance*. NASA. Langley Research Center, Mental-State Estimation, pages 269–286, 1987.

[17] Steven J Luck. *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press, 2014.

[18] Michael Plöchl, José P Ossandón, and Peter König. Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, 6, 2012.

[19] Charles Elkan. The foundations of cost-sensitive learning. In *International Joint Conference on Artificial Intelligence*, volume 17, pages 973–978. Lawrence Erlbaum Associates Ltd, 2001.

[20] Reza Fazel-Rezai, Brendan Z Allison, Christoph Guger, Eric W Sellers, Sonja C Kleih, and Andrea Kübler. P300 brain computer interface: current challenges and emerging trends. *Frontiers in Neuroengineering*, 5, 2012.

[21] Amelia J Solon, Stephen M Gordon, Brent J. Lance and Vernon J. Lawhern. Deep learning approaches for p300 classification in image triage: Applications to the NAILS task. In *Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies, NTCIR-13, Tokyo, Japan*, pages 5–8. 2017.

[22] James A Hanley and Barbara J McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36, 1982.

[23] Kay Henning Brodersen, Cheng Soon Ong, Klaas Enno Stephan, and Joachim M Buhmann. The balanced accuracy and its posterior distribution. In *Pattern recognition (ICPR), 2010 20th International Conference on*, pages 3121–3124. IEEE, 2010.

[24] Bertrand Rivet, Antoine Souloumiac, Virginie Attina, and Guillaume Gibert. xDAWN algorithm to enhance evoked potentials: application to brain–computer interface. *IEEE Transactions on Biomedical Engineering*, 56(8):2035–2043, 2009.

[25] Yaqub Jonmohamadi, Govinda Poudel, Carrie Innes, Daniel Weiss, Rejko Krueger, and Richard Jones. Comparison of beamformers for EEG source signal reconstruction. *Biomedical Signal Processing and Control*, 14:175–188, 2014.

[26] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Muller. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2008.

[27] Yijun Wang, Shangkai Gao, and Xiaornog Gao. Common spatial pattern method for channel selection in motor imagery based brain–computer interface. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 5392–5395. IEEE, 2006.

[28] Hubert Cecotti, Miguel P. Eckstein, and Barry Giesbrecht. Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering. *IEEE Transactions on Neural Networks & Learning Systems*, 25(11):2030–2042, 2014.

[29] Matthias S Treder, Anne K Porbadnigk, Forooz Shahbazi Avarvand, Klaus-Robert Müller, and Benjamin Blankertz. The LDA beamformer: optimal estimation of ERP source time series using linear discriminant analysis. *NeuroImage*, 129:279–291, 2016.

[30] Scott Makeig and Julie Onton. ERP features and EEG dynamics: an ICA perspective. *Oxford Handbook of Event-Related Potential Components. New York, NY: Oxford*, 2009.

[31] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4):411–430, 2000.

[32] Marzieh Mohammadi, Sepideh Hajipour Sardouie, and Mohammad Bagher Shamsollahi. Denoising of interictal EEG signals using ICA and time varying AR modeling. In *Biomedical Engineering (ICBME), 2014 21th Iranian Conference on*, pages 144–149. IEEE, 2014.

[33] Tzyy-Ping Jung, Scott Makeig, Colin Humphries, *et al.* Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(2):163–178, 2000.

[34] Jonathon Shlens. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014.

[35] Muhammad Naeem, Clemens Brunner, and Gert Pfurtscheller. Dimensionality reduction and channel selection of motor imagery electroencephalographic data. *Computational Intelligence and Neuroscience*, 2009, 2009.

[36] T Zanotelli, SA Santos Filho, and Carlos J Tierra-Criollo. Optimum principal components for spatial filtering of EEG to detect imaginary movement by coherence. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 3646–3649. IEEE, 2010.

[37] Galit Fuhrmann Alpert, Ran Manor, Assaf B Spanier, Leon Y Deouell, and Amir B Geva. Spatiotemporal representations of rapid visual target detection: a single-trial EEG classification algorithm. *IEEE Transactions on Biomedical Engineering*, 61(8):2290–2303, 2014.

[38] Mike X Cohen. *Analyzing neural time series data: theory and practice*. Cambridge, MA: MIT Press, 2014.

[39] Jia Meng, Lenis Mauricio Meriño, Nima Bigdely Shamlo, Scott Makeig, Kay Robbins, and Yufei Huang. Characterization and robust classification of EEG signal from image RSVP events with independent time-frequency features. *PLoS One*, 7(9):e44464, 2012.

[40] Adam D Gerson, Mads Dyrholm, An Luo, *et al.* Spatiotemporal linear decoding of brain state. *Signal Processing Magazine IEEE*, 25(1):107–115, 2008.

[41] Christopher M. Bishop. *Pattern recognition and machine learning*. New York: Springer. 2007.

[42] Jerome H Friedman. Regularized discriminant analysis. *Journal of the American Statistical Association*, 84(405):165–175, 1989.

[43] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. New York: John Wiley & Sons, 2012.

[44] Jing Hao Xue and Michael Titterington. Do unbalanced data have a negative effect on LDA? *Pattern Recognition*, 41(5):1558–1571, 2008.

[45] David JC MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992.

[46] Ulrich Hoffmann, Jean-Marc Vesin, Touradj Ebrahimi, and Karin Diserens. An efficient P300-based brain–computer interface for disabled subjects. *Journal of Neuroscience Methods*, 167(1):115–125, 2008.

[47] Hubert Cecotti and Anthony J Ries. Best practice for single-trial detection of event-related potentials: application to brain–computer interfaces. *International Journal of Psychophysiology*, 111:156–169, 2017.

[48] Paul Sajda, Adam D Gerson, Marios G Philiastides, and Lucas C Parra. Single-trial analysis of EEG during rapid visual discrimination: enabling cortically-coupled computer vision. *Towards Brain–Computer Interfacing*, pages 423–44, 2007.

[49] Yonghong Huang, Deniz Erdogmus, Santosh Mathan, and Misha Pavel. Boosting linear logistic regression for single trial ERP detection in rapid serial visual presentation tasks. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 3369–3372. IEEE, 2006.

[50] Edgar E Osuna. *Support vector machines: training and applications*. Available from http://hdl.handle.net/1721.1/7290. 1997.

[51] Graham Healy and Alan F Smeaton. Optimising the number of channels in EEG-augmented image search. *Proceedings of the 25th BCS Conference on Human-Computer Interaction*. British Computer Society, 2011.

[52] Eva Mohedano, Kevin Mcguinness, Graham Healy, Noel E O'Connor, Alan F Smeaton, Amaia Salvador, Sergi Porta, and Xavier I-Nieto. Exploring EEG for object detection and retrieval. In *ACM on International Conference on Multimedia Retrieval*, pages 591–594, 2015.

[53] Albert Nigrin. *Neural networks for pattern recognition*. Oxford: Oxford University Press, 1995.

[54] Hecht-Nielsen. Theory of the backpropagation neural network. In *International Joint Conference on Neural Networks*, volume 1, pages 593–605, 1989.

[55] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[56] Dineshbalu Balakrishnan and Sadasivan Puthusserypady. Multilayer perceptrons for the classification of brain computer interface data. In *Bioengineering Conference, 2005. Proceedings of the IEEE Northeast*, pages 118–119, 2005.

[57] Anil K Jain, Robert PW Duin, and Jianchang Mao. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 22(1):4–37, 2000.

[58] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. Cambridge, MA: MIT Press, 2016.

[59] Gustavo Carneiro, Jacinto C. Nascimento, and António Freitas. The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. *IEEE Transactions on Image Processing*, 21(3):968–982, 2012.

[60] Jianfeng Gao, Xiaodong He, and Li Deng. Deep learning for web search and natural language processing. Available from https://www.microsoft.com/en-us/research/publication/deep-learning-for-web-search-and-natural-language-processing/, 2015.

[61] Sergey Zagoruyko and Nikos Komodakis. Learning to compare image patches via convolutional neural networks. In *Computer Vision and Pattern Recognition*, pages 4353–4361, 2015.

[62] Steve Lawrence and C Lee Giles, Ah Chung Tsoi, and Andrew D Back. Face recognition: a convolutional neural-network approach. *IEEE Transactions on Neural Networks*, 8(1):98–113, 1997.

[63] Baotian Hu, Zhengdong Lu, Hang Li, and Qingcai Chen. Convolutional neural network architectures for matching natural language sentences. *Advances in Neural Information Processing Systems*, 3:2042–2050, 2015.

[64] Ming Liang and Xiaolin Hu. *Recurrent convolutional neural network for object recognition*. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

[65] Hyungtae Lee and Heesung Kwon. Single-trial EEG RSVP classification using convolutional neural networks. In *SPIE Defense + Security*, page 983622, 2016.

[66] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. EEGnet: a compact convolutional network for EEG-based brain–computer interfaces. *arXiv preprint arXiv:1611.08024*, 2016.

[67] Ran Manor and Amir B. Geva. Convolutional neural network for multi-category rapid serial visual presentation BCI. *Frontiers in Computational Neuroscience*, 9:146, 2015.

[68] Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernocký, and Sanjeev Khudanpur. Recurrent neural network based language model. In *INTER-SPEECH 2010, Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September*, pages 1045–1048, 2010.

[69] Elif Derya Übeyli. Analysis of EEG signals by implementing eigenvector methods/recurrent neural networks. *Digital Signal Processing*, 19(1):134–143, 2009.

[70] Nihal Fatma Güler, Elif Derya Übeyli, and İnan Güler. Recurrent neural networks employing Lyapunov exponents for EEG signals classification. *Expert Systems with Applications*, 29(3):506–514, 2005.

[71] Elliott M Forney and Charles W Anderson. Classification of EEG during imagined mental tasks by forecasting with Elman recurrent neural networks. In *International Joint Conference on Neural Networks*, pages 2749–2755, 2011.

[72] Geoffrey E Hinton. Deep belief networks. *Scholarpedia*, 4(6):5947, 2009.

[73] Geoffrey E Hinton. A practical guide to training restricted Boltzmann machines. *Momentum*, 9(1):599–619, 2012.

[74] Shaheen Ahmed, Lenis Mauricio Merino, Zijing Mao, Jia Meng, Kay Robbins, and Yufei Huang. A deep learning method for classification of images RSVP events with EEG data. In *IEEE Global Conference on Signal and Information Processing*, pages 33–36, 2013.

[75] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, *et al.* Generative adversarial networks. *Advances in Neural Information Processing Systems*, 2014.