PART 2

Signal Processing and Machine Learning

6

Electroencephalography Data Preprocessing

6.1. Introduction

Electroencephalography[1] provides measurements of electrical potential in the form of a time signal for each electrode. Although they only partially reflect the brain's underlying electrophysiological phenomena (see Chapter 3), these measurements contain a significant amount of information, which is used in clinical diagnosis and cognitive sciences, as well brain computer interfaces. The information provided by EEG is varied in nature: it can show variations of electric potential in amplitude at certain times and at certain frequencies, or variations in its spatial distribution. Accordingly, to observe these phenomena, a variety of methods are used in order to analyze signals in time, in frequency or in space. Finally, statistical analysis is also necessary to assess the stability of the information extracted with respect to different parameters (in time, through subjects, through different populations, between different experimental conditions).

Before being interpreted, an EEG should be properly preprocessed in order to denoise or filter the measurements and to extract different

Chapter written by Maureen CLERC.

1 This whole chapter, which deals with electroencephalography, also applies to magnetoencephalography described in Chapter 2.

Brain–Computer Interfaces 1: *Foundations and Methods*, First Edition.
Edited by Maureen Clerc, Laurent Bougrain and Fabien Lotte.

components of interest to the study at hand. In this chapter, we will see that the data can be represented from a variety of viewpoints (temporal, frequency, time–frequency, spatial, statistical), offering many possibilities for preprocessing.

Section 6.2 focuses on EEG acquisition, which in practice reduces the amount of available information. The subsequent sections discuss different representations in time and frequency in sections 6.3 and 6.4 respectively, and then, in section 6.5, combined time–frequency representations, which make it possible to analyze non-stationary signals. Section 6.6 discusses a variety of spatial representations, from topographic representations on the scalp's surface, to the reconstruction of electrical brain sources. Finally, section 6.7 focuses on statistical analyses aimed at extracting or separating components.

Some methods presented in this chapter have also been addressed in Chapter 2 on brain imaging, of which they are an integral part (time–frequency representations, source localization, independent component analysis (ICA)).

6.2. Principles of EEG acquisition

Because of its very low amplitude, which is in the order of 10–100 μV, EEG data can only be measured through an amplifier. This amplifier provides one time-dependent signal per measured channel. Originally, EEG amplifiers were analog, using a pen to trace signals on a roll of paper, much like a seismograph. Nowadays, most amplifiers are digital: they perform an analog–digital conversion and supply data in the form of sampled and quantized signals.

6.2.1. *Montage*

An EEG carries out measurements through a set of electrodes placed on the scalp, as explained in Chapter 1. Data come in the form of time signals, each corresponding to an acquisition channel. However, a channel's signal does not directly correspond to an electrode's electrical potential, since only *electrical potential differences* are measurable. There are different ways, called *montages*, of measuring these potential differences. The most common types are monopolar and bipolar.

In a monopolar montage, the same reference is used for each channel. This reference can be derived from a particular electrode (called the *reference electrode*), or from the average electric potential over all electrodes (called the *average reference*).

The bipolar montage measures differences in potential between neighboring electrodes. Each channel therefore has a different reference. A bipolar channel is especially sensitive to the electrical activity coming from sources located between the two electrodes.

The choice of montage used to be crucial in the times when analog amplifiers traced measurements on strips of paper. But with digital amplifiers, it is now possible to process the data *a posteriori* in order to transform one montage to another.

6.2.2. *Sampling and quantification*

With digital amplifiers, EEG recordings are not continuous but discrete. Specifically, the data are sampled, i.e. measured at instants separated by the *sampling interval*. This is usually the same throughout the duration of the recording. The *sampling rate* (or sampling frequency) is the inverse of the sampling interval. For example, a signal sampled at 100 Hz has a sampling interval of 10 ms.

The sampling frequency (or interval) determines the temporal resolution of observable phenomena, as well as the extent of the analyzable frequency spectrum (see section 6.4.1). Shannon's sampling theory calls for a sampling frequency greater than twice the target maximum frequency [MAL 08].

The analog–digital conversion performed by EEG amplifiers quantifies the signals: their amplitude is coded with a finite number of bits. For instance, 16 bits provide 2^{16} values to encode the whole dynamic range of the EEG. With 16 bits, values between ± 600 μV have an amplitude precision of 0.0183 μV.

6.3. Temporal representation and segmentation

The most basic representation of EEG data provided directly by the amplifier is temporal. Raw EEG data comes in the form of time *signals*, each

corresponding to a channel. Observing these signals during their acquisition is important in order to ascertain that they are of good quality.

6.3.1. *Segmentation*

The signals measured by the amplifier can be accompanied by additional time stamps known as *events*. Each event is associated with a date (a time sample) and a label specifying its nature. Events are sometimes marked by the experimenter using a specific software, during data acquisition or review. In brain–computer interfaces, it is more common for marking to be performed automatically during acquisition. The technical details of this marking process are explained in Chapter 7 of Volume 2 [CLE 16]

Events provide temporal information that makes it possible to define segments of particular interest for data analysis through a time window (these segments are also referred to as "epochs"). This procedure is called data *segmentation* or *epoching* . In an EEG data set, we thus have *epochs* or *trials* of similar events. With these trials (also called realizations in statistical terms), and machine learning methods, it will be possible to automatically classify new data whose nature is unknown *a priori*.Trials are also important for EEG analysis, since they make it possible to obtain enough data to reduce the effect of noise. Averaging the repetitions together yields "evoked potentials" (also see Chapter 2).

6.3.2. *Time domain preprocessing*

Noise reduction: Some parts of the signal can be eliminated simply because they display undesirable behavior (e.g. because of displaced electrodes that have lost contact with the scalp and only measure noise).

Component separation: The signal components of interest can be identified directly in temporal signals, especially in low-frequency components (such as the P300), or in large amplitude changes (such as slow cortical potentials, or the spikes observed in epilepsy). However, the activity measured on the sensors results from a combination of several brain sources, and as we shall see in section 6.6.2, these temporal components are generally better identified after spatial filtering.

6.4. Frequency representation

Since the EEG recordings in humans were carried out by Hans Berger, its oscillatory character has been evident. We have seen in previous chapters that EEG contains oscillations (alpha and beta waves, etc.) from different brain regions. To represent the EEG spectrum, that is to say its frequency variations, it is necessary to use the Fourier transform [BRI 01].

6.4.1. *Fourier transform*

We define a function's (continuous) Fourier transform as

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-i\omega\, t) dt$$

Consider a signal f measured on the interval $[0,\ T]$ with N discrete samples separated by τ (the sampling interval): $f[n] \doteq f(n\tau)$, $n = 0 \ldots N-1$. The discrete Fourier transform is defined for $k = 0$ to $N-1$ by

$$\hat{f}[k] = \sum_{n=0}^{N-1} f[n] \exp^{-2i\pi kn/N}$$

The discrete Fourier transform of f is a sequence of complex numbers $\hat{f}[k] = c_k \exp(i\phi_k)$, where c_k (respectively ϕ_k) represents the amplitude (respectively phase) spectrum.

The definition of the discrete Fourier transform involves only N samples of the signal, and due to this finite length, it is mathematically defined as the Fourier series of an N-periodic Dirac comb obtained by *periodization* of $f[n], n = 0, \ldots, N-1$. This is of considerable importance if we consider regularity properties, because periodization can artificially introduce discontinuities. The Fourier transform of a discontinuous function decreases slowly at high frequencies, causing artificial components to appear in its spectrum. This frequency alteration is called the Gibbs phenomenon.

It is therefore preferable to apply the Fourier transform to signals whose periodization is not discontinuous: before applying the Fourier transform, it is often recommended to multiply the signal by a window function, equal to zero at both ends $t = 0$ and $t = T$. This window (called a *tapering* window)

has the effect of blurring the periodization discontinuity without dramatically altering the original signal's Fourier transform. Many window functions can accomplish this (Gaussian, Hamming, Blackman, Hanning, etc.), and it is necessary to compare each one's relative advantage in terms of frequency resolution, frequency spreading and amplitude modifications.

A signal $f[n] = f(n\tau)$, $n = 0 \ldots N-1$ sampled at frequency $F = 1/\tau$ (the inverse of the sampling interval τ), has a discrete Fourier transform that includes frequencies between $-F/2$ and $F/2$ with a frequency resolution (the distance between two samples) of $F/N \approx \frac{1}{N\tau} = \frac{1}{T}$. It is therefore important to pay attention to both the sampling frequency and the signal length when it comes to its spectrum. To summarize:

– the sampling rate limits the extent of the analyzable Fourier spectrum, as mentioned above;

– on the other hand, the signal duration T determines the spectrum's frequency resolution, i.e. the minimum frequency distance between components such that they are distinguishable.

The most common digital method for calculating the discrete Fourier transform is the fast Fourier transform (FFT), a recursive algorithm that is valuable because of its low computational complexity: it only grows by a factor of $N \log N$ with the number of samples N. It is important to recall that when the signal length N increases, the estimation of its spectrum does not automatically become more precise. Straightforward estimators of spectral density are indeed biased, and it is necessary to resort to a periodogram to correct for this phenomenon [BRO 91, BRI 01].

6.4.2. *Frequency filtering*

Noise reduction: it is possible to eliminate artifacts that occur in specific frequency bands through filtering. A high-pass filter removes low frequencies, which is useful to correct for occasionally observed low-frequency fluctuations, especially with dry electrodes. A low-pass filter removes high frequencies, eliminating some sources of noise, such as electrical activity coming from muscles. The power line current produces an artifact on the

EEG channels at a given frequency[2], which can be eliminated by a band-stop filter (notch filter).

Component separation: many neurophysiological BCI markers are in a specific frequency band, as seen in Chapter 4, which may even be specific to each subject (for example the motor rhythm μ). However, for steady-state visual evoked potentials, the frequency of interest is not subject specific, since it is usually twice the stimulation frequency. Frequency analysis of the signal is important in order to filter signals in the appropriate band. Filtering must, however, be applied carefully because, for non-stationary signals, it may introduce artifacts. This may be the case, for example, for epileptic spikes, which, when filtered at high frequency, may suggest the presence of oscillations that in fact result only from the Gibbs phenomenon [BÉN 10].

6.5. Time–frequency representations

The Fourier transform is a powerful tool for decomposing EEG signals into oscillatory components, and thus for analyzing brain rhythms. However, it is rare for these oscillations to remain stationary for the duration of the recording. Rather, it is common to observe transient events, i.e. oscillation "bursts". For this reason, it is common practice to perform frequency analysis in a time window centered around a chosen moment, while allowing this time window to move along the time axis. This is known as a *sliding window*, and leads to a time–frequency representation indexed by two variables: the time point where the decomposition is centered and the frequency of each Fourier component.

6.5.1. *Time–frequency atom*

Representing a time–frequency signal involves analyzing it around a time–frequency position (u, ξ). A time–frequency atom is the building block that localizes the analysis in time and frequency [FLA 93, MAL 08]. It is a normalized function ϕ ($\|\phi\|^2 = \int_{-\infty}^{\infty} |\phi(t)|^2 \, dt = 1$), centered in time around $u = \int_{-\infty}^{\infty} t \, |\phi(t)|^2 \, dt$, and in frequency around $\xi = \frac{1}{2\pi} \int_{-\infty}^{\infty} \omega \, |\hat{\phi}(\omega)|^2 \, d\omega$.

2 50 Hz in Europe, Asia and Africa; 60 Hz in North America.

Analyzing a signal f around (u, ξ) amounts to calculating the inner product $\langle f, \phi \rangle$ of f and ϕ (which involves the complex conjugate of ϕ, denoted ϕ^*):

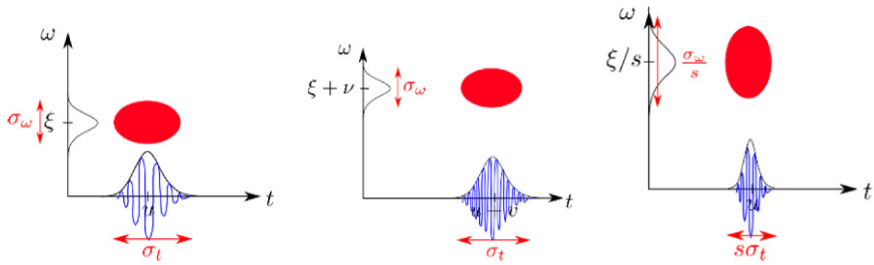
$$\langle f, \phi \rangle = \int_{-\infty}^{\infty} f(t)\, \phi^*(t)\, dt$$

Suppose that the function ϕ is localized around a position u (i.e. it vanishes outside a neighborhood of u). Then, $\langle f, \phi \rangle$ depends only on the values taken by $f(t)$ in the neighborhood of $t = u$.

Similarly, in terms of frequency, using Parseval's formula, the inner product can be rewritten as:

$$\langle f, \phi \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega)\, \hat{\phi}^*(\omega)\, d\omega$$

If the function $\hat{\phi}$ is localized around a frequency ξ, then $\langle f, \phi \rangle$ depends only on the values taken by $\hat{f}(\omega)$ in the neighborhood of $\omega = \xi$.



a) Original atom centered at (u, ξ)

b) Atom after translation and modulation to $(u + v, \xi + \nu)$

c) Atom after scaling (contraction by $s < 1$)

Figure 6.1. *The translation and modulation of an atom alters its time–frequency position without changing its time–frequency spread, whereas scaling alters both localization and spread*

The temporal spread of a time–frequency atom around its average position u is given by σ_t:

$$\sigma_t^2 = \int_{-\infty}^{\infty} (t - u)^2\, |\phi(t)|^2\, dt$$

The smaller the temporal spread, σ_t, the more the function is concentrated around its average position. The translation of a function changes its average position, without changing its time spread (Figure 6.1(b)).

Similarly, we can measure the frequency spread σ_ω of ϕ around ξ:

$$\sigma_\omega^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} (\omega - \xi)^2 \, |\hat{\phi}(\omega)|^2 \, d\omega \qquad [6.1]$$

Modulating a function (i.e. multiplying it by a complex exponential $e^{i\nu t}$) moves its average frequency by a factor of ν, but it does not change its frequency spread (Figure 6.1(b)).

There is a limit to joint time–frequency concentration. The more a function is concentrated in time, the less it is concentrated in frequency. This principle is expressed mathematically (under certain regularity conditions for the function ϕ) by the equation: $\sigma_t \, \sigma_\omega \geq \frac{1}{2}$. Gaussian functions are the only functions for which this relation is in fact an equality: they have a maximal time–frequency concentration.

NOTE.– In the context of quantum physics, a similar result, called "uncertainty principle", has been shown by Weyl: with increasing certainty about a particle's position comes increased uncertainty about its momentum.

6.5.2. *Short-time Fourier transform*

The short-term Fourier transform (STFT) consists of constructing a time-frequency atom around (u, ξ) by translating and modulating a symmetric real function g, supported within a time interval: $g_{u,\xi}(t) = g(t - u)e^{i\xi t}$. The function g can, for example, be a window function which is smoothed at its extremities to avoid discontinuities.

The STFT of f is obtained by computing the scalar product with the atom $g_{u,\xi}$:

$$\langle f, \, g_{u,\xi} \rangle = \int_{-\infty}^{\infty} f(t) g(t - u) e^{-i\xi t} dt$$

The resolution of the STFT around an instant u and a frequency ξ is uniform in the time–frequency plane. It is independent of u and ξ, but it depends on the

concentration of g in time (its support) and frequency (the decay of $\hat{g}(\omega)$ at high frequencies, which is related to the regularity of g).

The *spectrogram*, energy density in time–frequency, is the squared amplitude of the STFT: $P_S f(u,\xi) = |\langle f,\, g_{u,\xi}\rangle|^2$.

The STFT is very useful for extracting features of brain activity for brain–computer interfaces. It is important to choose an analysis window g whose time and frequency concentrations are well adapted to the phenomena to explore.

6.5.3. *Wavelet transform*

As we have seen, the STFT has a uniform solution in time–frequency, which depends on the time–frequency concentration of g. But for certain types of signals, which have non-uniform variations at different frequencies or at different times, it is sometimes preferable to analyze the phenomena with a resolution that depends on frequency. For such signals, the wavelet transform may be an appropriate choice.

A wavelet is a normalized time–frequency atom ψ, whose average is zero $\int \psi(t)\,dt = 0$, and whose average position is also zero $\int t|\psi(t)|^2\,dt = 0$. A family of time–frequency atoms is constructed by translating ψ to position u and applying a scaling factor $s > 0$:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right)$$

Since ψ is centered around 0, $\psi_{u,s}$ is centered around $t = u$. Recalling that the time spread σ_t of ψ is defined by $\sigma_t^2 = \int_{-\infty}^{+\infty} t^2\,|\psi(t)|^2\,dt$, a change in variable $t \mapsto \frac{t-u}{s}$ shows that the time spread of the atom $\psi_{u,s}$ is $s\sigma_t$:

$$\int_{-\infty}^{+\infty} (t-u)^2\,|\psi_{u,s}(t)|^2\,dt = s^2\,\sigma_t^2 \qquad [6.2]$$

Note η the average frequency of $\hat{\psi}$: $\eta = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \omega\,|\hat{\psi}(\omega)|^2\,d\omega$. The Fourier transform $\psi_{u,s}$ corresponds to scaling $\hat{\psi}$ by a factor of $1/s$, and modulating it by u: $\hat{\psi}_{u,s}(\omega) = \sqrt{s}\,\hat{\psi}(s\omega)\,\exp(-i\omega u)$. The average frequency of ψ_{us} is

therefore η/s, and its frequency spread is $\frac{\sigma_\omega}{s}$:

$$\frac{1}{2\pi}\int_0^{+\infty}\left(\omega-\frac{\eta}{s}\right)^2\left|\hat{\psi}_{u,s}(\omega)\right|^2 d\omega=\frac{\sigma_\omega^2}{s^2}$$

where σ_ω is the frequency spread of the wavelet ψ defined by equation [6.1].

Figure 6.1(c) illustrates these changes in the time–frequency spread induced by scaling.

Wavelet transform or multiscale filtering: the wavelet transform of f at time position u and scale s is defined by its inner product with $\psi_{u,s}$

$$\langle f,\psi_{u,s}\rangle=\int_{-\infty}^{+\infty} f(t)\,\frac{1}{\sqrt{s}}\,\psi^*\left(\frac{t-u}{s}\right)dt \qquad [6.3]$$

The wavelet transform $\langle f,\psi_{u,s}\rangle$, denoted by $W_s(u)$, amounts to filtering f with a filter scaled by s. Indeed, the scalar product [6.3] above can be seen as the convolution of f with the function $\tilde{\psi}_s(t)=\frac{1}{\sqrt{s}}\,\psi^*(-t/s)$. The Fourier transform of $W_s(u)$ with respect to u can thus be written as a product:

$$\widehat{W_s}(\xi)=\sqrt{s}\hat{f}(\xi)\hat{\psi}^*(s\omega)$$

This can be interpreted as the filtering of f by a high-pass filter scaled by s, since a wavelet ψ has an average of zero, so $\hat{\psi}(0)=\int_{-\infty}^{\infty}\psi(t)\,dt=0$.

The *scalogram* is the time–frequency energy of f $|\langle f,\psi_{u,s}\rangle|^2=\left|\langle f,\,\psi_{u,\frac{\eta}{\xi}}\rangle\right|^2$.

The time–frequency resolution of the wavelet transform and the scalogram are related to the time–frequency spread of atoms $\psi_{u,s}$. According to our previous remarks on the relationship between time–frequency spread and scale, the temporal resolution increases as the frequency increases (or when the scale s decreases). Conversely, the frequency resolution decreases as the inverse of the central frequency.

6.5.4. *Time–frequency transforms of discrete signals*

We have described the short-time Fourier transforms and wavelet transforms for continuous functions $f(t)$, but in practice they must be applied to discrete signals of the form $\{f[n],\ n = 1 \ldots N\}$.

Time–frequency transforms of those discrete signals, whether wavelets or STFTs, may take two different forms, depending on the way the time–frequency plane is sampled.

In an *orthogonal transform*, exactly N samples are used in the time–frequency plane, i.e. the transform has the same dimensionality as the initial signal.

In a *redundant transform*, the time–frequency plane is sampled more densely, resulting in more samples than the original N samples of f.

The redundant transform may seem less interesting, because it uses more data to represent the same information. However, only a redundant transform provides the property of invariance by translation, which is very useful if there is uncertainty on the exact time when the activity of interest occurs.

Orthogonal wavelet transforms can be calculated with the fast wavelet transform, a fast algorithm with $N \log N$ complexity [MAL 08]. This type of wavelet transform is useful for signal or image compression applications (see JPEG 2000), but for EEG signal analysis, redundant wavelet transforms are preferable.

6.5.5. *Toward other redundant representations*

Time–frequency atoms are convenient for representing localized events in both time and frequency, as we have seen. When grouped together in a "dictionary" (term for a large collection of atoms providing a redundant representation of signals), the atoms most accurately representing the signals can be selected by algorithms that promote sparsity. As examples of such algorithms, we can mention Matching Pursuit [MAL 93] and its multitrial [BÉN 09] or multichannel [GRI 08] extensions.

Despite their importance, we must admit that due to their smooth, symmetric time-course, conventional time–frequency atoms (e.g. Gabor wavelets) are quite different from the neurophysiological events they seek to represent. It is often necessary to use several atoms to represent a single event, which is detrimental to interpretation. However, dictionaries can be enriched in order to contain atoms of the desired shape [SPU 15].

Since the desired shape is rarely known beforehand, the measured signals themselves should be used to learn atoms' waveforms, through dictionary learning [MAI 09]. For this purpose, iterative methods have been developped that alternatively update waveforms and their associated coefficients in the signals. This type of analysis, which is well adapted to the signals of interest, should be further developed in the future [HIT 13].

6.6. Spatial representations

In order to characterize the electrical activity observed by EEG, it is interesting to associate it with the brain regions where it originates. The position of the different electrodes can be used for this: primary visual phenomena may be found on occipital electrodes and motor phenomena on central electrodes. The spatial localization of electrical potential on the scalp provides coarse information about the origin of electrical activity, called a *topographical representation* (section 6.6.1). For a more detailed spatial analysis, it is common to instead rely on a spatial filtering or localisation (sections 6.6.2 and 6.6.3).

6.6.1. *Topographic representations*

With a monopolar montage (potential differences between each sensor and a common reference), and a so-called "high (spatial) resolution" EEG, the distribution of electrodes on the scalp provides a map of the electrical potential, much like a temperature map of the globe. In fact, in order to obtain such a representation, it is necessary to interpolate the data between the measurement electrodes. This interpolation can be done with spherical splines (which assume the electrodes to be distributed on a sphere), or with two-dimensional splines on a mesh matching the shape of the scalp, which is a better approximation [PER 87].

How many electrodes are required for a "high-resolution EEG"? With 20 electrodes, it is possible to obtain a coarse topographic representation, and 64 electrodes or more produce a fairly accurate continuous map of potential. It is not reasonable to exceed 256 electrodes (the densest system currently on the market) for two reasons. On the one hand, it would be difficult to obtain measurements without creating isopotentials (in cases where electrodes are placed using conductive gel), and on the other hand that number is sufficient to properly represent the potential field present on the scalp. Indeed, due to the low conductivity of the skull, which stands in between sources and sensors, the electrical potential has rather low spatial frequencies on the scalp [NUN 06]. According to Shannon's theory, it can be accurately represented with a low-density spatial sampling. In contrast to EEG, magnetoencephalography has higher spatial frequencies and several hundred sensors are useful to accurately represent the magnetic field.

6.6.2. *Spatial filtering*

The electrical potential measured by EEG results from the superposition of different sources of cerebral activity, which create variations in electric potential on the scalp through volume conduction. Separating these different sources is difficult, but worthwhile because it yields signals that are closer to the underlying electrical brain activity. Spatial filtering offers this possibility of "deconvolving" the sensor measurements, like a pair of glasses compensates for short sightedness.

Before presenting the possible strategies, we should mention that the source mixture is instantaneous (Maxwell quasistatic). Consider M brain sources of amplitude j_m. For each time sample n, the potential measured at each sensor k is of the form[3] $f_k[n] = \sum_{m=1}^{M} g_{km} j_m[n]$. This relationship is a linear one that can be written as a matrix–vector product $f = \mathbf{G}\, J$, where the matrix $\mathbf{G}$ is called *gain matrix*. Each element g_{km} of the gain matrix represents the potential on the sensor indexed by k resulting from the presence of a unitary source (of magnitude 1) at the position of the source indexed by m.

3 In this model, measurement noise is ignored.

6.6.2.1. *Surface Laplacian*

The electric potential, denoted V, satisfies the relation:

$$\nabla \cdot (\sigma \nabla V) = \nabla \cdot \mathbf{J}^{\mathrm{p}} \qquad [6.4]$$

where σ denotes the conductivity of the head and $\mathbf{J}^{\mathrm{p}}$ are the primary sources. In areas where there are no primary electrical sources (e.g. skull and scalp), $\nabla \cdot (\sigma \nabla V) = 0$. So the Laplacian of the potential vanishes in those regions: $\Delta V = 0$.

On a surface S with normal vector $\mathbf{n}$, the surface Laplacian $\Delta_S V$ is connected to the volumic Laplacian by $\Delta V = \Delta_S V + \frac{\partial^2 V}{\partial \mathbf{n}^2}$. On the scalp, the surface Laplacian provides an interesting spatial filter because $\Delta_S V$ is close to the normal current on the skull. Indeed, as the Laplacian cancels out within the scalp, we have $\Delta_S V = -\frac{\partial^2 V}{\partial \mathbf{n}^2}$. But since air is non-conductive, by normal current conservation, on the scalp we have $\frac{\partial V}{\partial \mathbf{n}} = 0$, so $-\frac{\partial^2 V}{\partial \mathbf{n}^2}$ is approximately proportional to the normal derivative of potential on the skull. To compute the surface Laplacian, we must know V on the scalp, as well as the shape of the scalp's surface. The "surface Laplacian" most commonly used in BCI, which simply subtracts from a channel the average of its neighbors, is a very coarse approximation of the actual $\Delta_S V$, which simply subtracts the average of its neighbors from each channel. Figure 6.2 shows several possible choices for the neighbors, the "large" (respectively "narrow") Laplacian filters being sensitive to deep (respectively superficial) cortical sources.

6.6.2.2. *Cortical current density*

We have seen that the surface Laplacian is close to the normal current on the skull. Some techniques proceed even further to reconstruct the normal current on the cortical surface, which is even closer to the primary sources. This is possible because, in areas where there are no sources (the scalp, the skull), the potential is harmonic ($\Delta V = 0$), and mathematically, there is existence and uniqueness of the continuation of a function which is harmonic within a domain. However, the problem is "ill-posed" because the solution of this continuation problem is unstable: small changes to data on the boundary can lead to major changes in the function within the domain. But regularized numerical approximations have been proposed, which use, for example, finite boundary elements [HE 02, CLE 07].

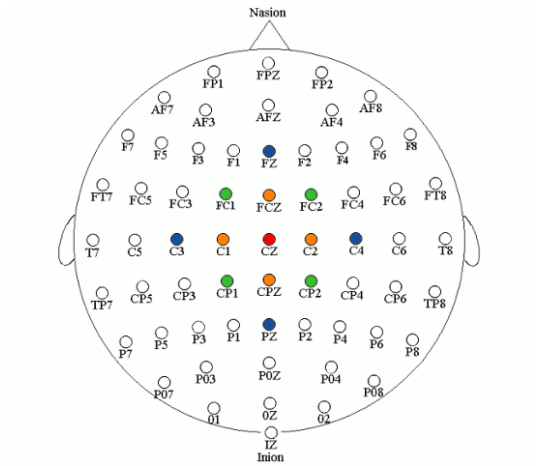


Figure 6.2. *Three Laplacian filters applied to Cz. The average of the signal measured at C3, C4, Fz and Pz (blue) is withdrawn to the Cz channel (red) for wide Laplacian filtering. The electrodes used for a medium and narrow Laplacian filter are green and orange, respectively. For a color version of this figure, see www.iste.co.uk/clerc/interfaces1.zip*

6.6.3. *Source reconstruction*

Beyond the normal current density, it can be even more interesting to identify the distribution of primary sources $\mathbf{J}^{\mathrm{p}}$ responsible for the observed measurements. The source reconstruction problem is ill-posed: it is unstable with respect to measurements, and in some cases the distribution of sources may be non-uniquely determined by the boundary measurements. However, mathematical results from potential theory guarantee uniqueness in the presence of exact measurements on a dense (continuous) portion of the boundary. The sources should obey one of the two models below (Figure 6.3):

– sources distributed orthogonally to the surface of the cortex: the brain activity that can be measured on EEG comes from pyramidal neurons (Chapter 3) and it can be represented by an extended source on a surface S, oriented in the normal direction $\mathbf{n}(\mathbf{r})$ at each position $\mathbf{r}$ of S, and whose amplitude is denoted $j(\mathbf{r})$:

$$\mathbf{J}^{\mathrm{p}}(\mathbf{r}) = j(\mathbf{r})\mathbf{n}(\mathbf{r})\delta_S(\mathbf{r})$$

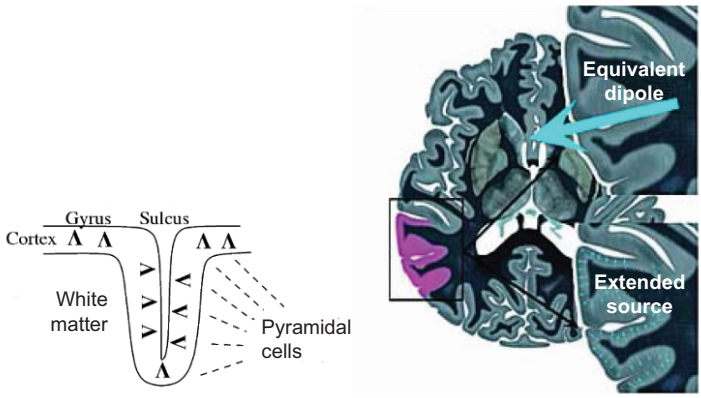


Figure 6.3. *Left: Cortical activity is locally perpendicular to the cortical surface. Right: according to the resolution at which it is described, the same cortical activity can be represented by an extended source (top inset) or an equivalent dipole. For a color version of this figure, see www.iste.co.uk/clerc/interfaces1.zip*

Since $\mathbf{n}(\mathbf{r})$ is known, source reconstruction consists of estimating the distribution of amplitudes $j(\mathbf{r})$ on the cortical surface:

– sources formed by a superposition of single dipoles:

$$\mathbf{J}^{\mathrm{p}}(\mathbf{r}) = \sum_{m=1}^{M} \mathbf{q}_m \delta_{\mathbf{p}_m}(\mathbf{r})$$

This model is acceptable when the activity is quite focal, possibly with several (M) foci. Source reconstruction consists of estimating the number M of dipoles, their position $\mathbf{p}_m$ and moment $\mathbf{q}_m$ (a vector representing amplitude and orientation).

Chapter 2 has already mentioned and illustrated source localization. In a nutshell, the sources J are estimated by minimizing a quantity of the form $\|f - \mathbf{G}\,J\|$ measuring a distance between measurements and their prediction, assuming the source were J. Source reconstruction requires precise knowledge of the gain matrix $\mathbf{G}$, mentioned in section 6.6.2, which connects the brain sources to sensor measurements. To calculate the gain matrix, it is necessary

to solve the forward problem [6.4], and therefore, model the electrical conductivity σ of all tissues composing the head. Three types of geometric models exist for conductivity, as shown in Figure 6.4, each corresponding to a family of numerical methods:

– models of nested spheres use numerical methods based on spherical harmonic expansions [DE 93];

– surfacic models, whose surfaces are assumed to separate tissues with homogeneous conductivity, use numerical methods based on finite boundary elements [KYB 05];

– volumic models, whose conductivity can be inhomogeneous and even anisotropic, use numerical methods based on 3D finite elements, with tetrahedral, hexahedral meshes [WAB 07] or meshless, implicit formulations [VAL 10].

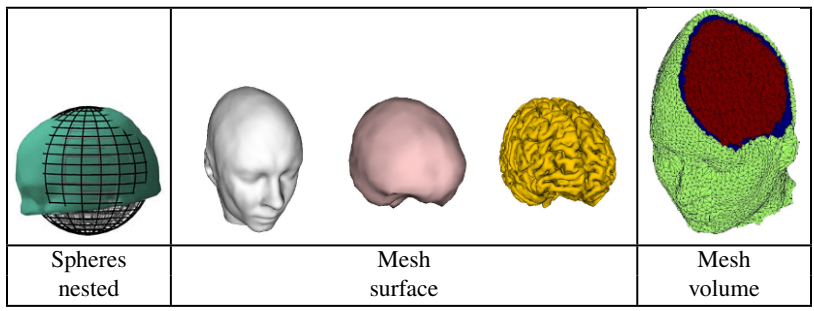


Figure 6.4. *Geometric representations of head conductivity used to compute the EEG forward problem. For a color version of this figure, see www.iste.co.uk/clerc/interfaces1.zip*

6.6.4. *Using spatial representations in BCI*

In BCI, the use of the Laplacian filter has become common, but more advanced spatial representations such as source reconstruction or cortical density are seldom used [NOI 08]. Indeed, they require geometrical data (the geometry of head tissues and electrode positioning), and their complex processing pipelines still require specific expertise. Moreover, some of these methods are still too expensive in terms of computing time to be applied online. Still, the benefit of the brain source reconstruction is to provide an

excellent spatial and temporal separation of brain characteristics [BUR 15] and to produce generalizable features across subjects. This has been highlighted for error potential detection [DYS 15]. Restorative BCIs (which aim to restore efficient brain activity) should look for features in source space in order to better promote positive changes in brain activity. In the future, BCI should benefit from advances in source reconstruction methods, since this also addresses a need for better understanding the neurological mechanisms of BCI.

6.7. Statistical representations

It is sometimes difficult to identify relevant information from a set of measurements, or even from the various preprocessed representations described above. Statistical representations are useful for analyzing sets of signals measured simultaneously or through several trials (or realizations in statistical language). In Chapter 8, we will see that statistical analysis is, for example, essential for the preprocessing of extracellular potentials.

In the case of EEG, many families of statistical representations have been proposed (Bayesian methods, ergodic theory, etc.), and two of them, particularly used for in BCI, will be presented here: principal component analysis (PCA) and independent component analysis (ICA), which has already been discussed in Chapter 2.

6.7.1. *Principal component analysis*

A set of time courses measured by K channels $\{f_k[n], n = 1 \dots N\}_{k=1 \dots K}$ can be seen as a cloud of N points of IR^K. PCA estimates an orthogonal transformation (i.e. rotation) of $\mathbb{R}^K$ such that in the new arrangement, the axes are aligned in the direction of the point cloud's greatest dispersion. This is equivalent to diagonalizing the empirical covariance matrix of the data.

PCA must be conducted on data with a high signal-to-noise ratio (for example after averaging across several trials).

6.7.2. *Independent component analysis*

ICA is a blind source separation method. It assumes that the EEG comes from a mixture of K *statistically independent* sources s_j, for each channel $k = 1, \ldots, K$,

$$f_k[n] = \sum_{j=1}^{K} a_{kj} s_j[n]$$

The mixing coefficients can be organized in a matrix $\mathbf{A}$. ICA finds a "demixing" matrix $\mathbf{W}$ that approximates the inverse of the mixing matrix $\mathbf{W} \sim \mathbf{A}^{-1}$, so as to maximize the statistical independence of the reconstructed sources

$$\hat{s}_k[n] = \sum_{j=1}^{K} w_{kj} f_j[n]$$

A heuristic for approaching the mutual independence of signals is to search for $\mathbf{W}$ that maximizes the non-Gaussianity of signals $\hat{s}_k$, or that minimizes the mutual information between the $\hat{s}_k$ [HYV 00].

6.7.3. *Using statistical representations in BCI*

PCA requires data with a high signal to noise ratio, which, in practice, has already been averaged over several trials so for BCI it can be applied to training data. But the PCA rotation, once estimated, can be applied online to new data, in order to enhance it.

The same principle as PCA is used in the common spatial patterns (CSP) algorithm for binary classification (see Chapter 7). The CSP algorithm operates a joint diagonalization of the covariance matrices of each of the two classes, with the aims to maximize the distance between their covariance. Recent work has focused on information geometry in the space of symmetric positive definite matrices [CON 15]. Robust detection algorithms have been derived with this approach; for instance the "Riemannian Potato" for detecting artifacts [BAR 13].

ICA extracts components from EEG signals that can sometimes be independent of brain activity and hence considered artifacts.

Electrooculographic (EOG) activity is particularly well eliminated by ICA. For this, the demixing matrix is estimated from a prerecorded dataset. EOG components are identified by their frontal topography and time course. These components are subtracted through a projection, which consists of applying the demixing matrix, canceling out EOG components and finally reapplying the mixing matrix.

The methods described in this section are limited to two dimensions (sensors $\times$ time). More sophisticated methods can operate in more dimensions, e.g. multivariate analysis with PARAFAC in three dimensions (sensors $\times$ time $\times$ frequency) [MIW 04]. Mixed effects methods can be used to separate different signal variability sources [HUA 08], with applications in BCI detection [SPI 15].

6.8. Conclusions

In BCIs, preprocessing can be applied either offline (to analyze a prerecorded dataset) or online (directly during BCI interaction).

All forms of preprocessing described above are notably applied offline in order to check that the information required for BCIs is present in the signals and to specify the features to be exploited. Preprocessing in the temporal (or time–frequency), spatial and statistical domains each employs complementary principles, and they are therefore often combined. It is quite common to reconstruct dipolar sources from cross-trial averaged signals; to apply a time–frequency analysis to reconstructed distributed sources; to perform ICA and then to analyze the topographies of different components to localize corresponding sources in the brain.

The order in which processing is performed can sometimes have an effect, but not when the processing is linear, for example applying a Laplacian filter before or after cross-trial averaging produces the same results.

This chapter has presented an overview of useful forms of preprocessing for BCI and explained the basics of its main methods and their limitations and extensions. Of course, these methods need not be confined to preprocessing, since they may more generally be used as information extraction methods for EEG signals. It is certainly also interesting to apply them retrospectively to

interpret brain activity modifications occurring before, during and after BCI use.

6.9. Bibliography

[BAR 13] BARACHANT A., ANDREEV A., CONGEDO M., "The Riemannian potato: an automatic and adaptive artifact detection method for online experiments using Riemannian geometry", *TOBI Workshop IV*, Sion, Switzerland, pp. 19–20, January 2013.

[BAB 08] BABILONI F., CINCOTTI F., CARDUCCI F. *et al.*, "Spatial enhancement of EEG data by surface laplacian estimation: the use of magnetic resonance imaging-based head models", *Clinical Neurophysiology*, vol. 112, pp. 724–727, 2008.

[BRO 91] BROCKWELL P.J., DAVIS R.A., *Time Series: Theory and Methods*, Springer Series in Statistics, Springer-Verlag, New York, 1991.

[BÉN 09] BÉNAR C., PAPADOPOULO T., TORRÉSANI B. *et al.*, "Consensus matching pursuit for multi-trial EEG signals", *Journal of Neuroscience Methods*, vol. 180, pp. 161–170, 2009.

[BRI 01] BRILLINGER D.R., *Time Series: Data Analysis and Theory*, Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, 2001.

[BUR 15] BURLE B., SPIESER L., ROGER C. *et al.*, "Spatial and temporal resolutions of eeg: is it really black and white? a scalp current density view", *International Journal of Psychophysiology*, vol. 97, no. 3, pp. 210–220, 2015.

[CON 15] CONGEDO M., AFSARI B., BARACHANT A., "Approximate joint diagonalization and geometric mean of symmetric positive definite matrices", *PLoS ONE*, vol. 10, no. 4, p. e0121423, 2015.

[CLE 07] CLERC M., KYBIC J., "Cortical mapping by laplace-cauchy transmission using a boundary element method", *Inverse Problems*, vol. 23, pp. 2589–2601, 2007.

[CLE 16] CLERC M., BOUGRAIN L., LOTTE F., *Brain–Computer Interfaces 2: Technology and Application*, ISTE, London and John Wiley and Sons, New York, 2016.

[DE 93] DE MUNCK J.C., PETERS M.J., "A fast method to compute the potential in the multisphere model", *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 11, pp. 1163–1174, 1993.

[DYS 15] DYSON M., THOMAS E., CASINI L. *et al.*, "Online extraction and single trial analysis of regions contributing to erroneous feedback detection", *NeuroImage*, vol. 121, pp. 146–158, 2015.

[FLA 93] FLANDRIN P., *Temps-Fréquence*, Hermès, Paris, 1993.

[GRI 08] GRIBONVAL R., RAUHUT H., SCHNASS K. *et al.*, "Atoms of all channels, unite! average case analysis of multi-channel sparse recovery using greedy algorithms", *The Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 655–687, 2008.

[HE 02] HE B., ZHANG X., LIANG J. *et al.*, "Boundary element method-based cortical potential imaging of somatosensory evoked potentials using subjects' magnetic resonance images", *NeuroImage*, vol. 16, pp. 564–576, 2002.

[HIT 13] HITZIGER S., CLERC M., GRAMFORT A. *et al.*, "Jitter-adaptive dictionary learning – application to multi-trial neuroelectric signals", *ICLR – 1st International Conference on Learning Representations*, May 2013.

[HUA 08] HUANG Y., ERDOGMUS D., PAVEL M. *et al.*, "Mixed effects models for eeg evoked response detection", *IEEE Workshop on Machine Learning for Signal Processing, MLSP*, pp. 91–96, 2008.

[HYV 00] HYVÄRINEN A., OJA E., "Independent component analysis: algorithms and applications", *Neural Networks*, vol. 13, pp. 411–430, 2000.

[KYB 05] KYBIC J., CLERC M., ABBOUD T. *et al.*, "A common formalism for the integral formulations of the forward eeg problem", *IEEE Transactions on Medical Imaging*, vol. 24, pp. 12–28, 2005.

[MAL 08] MALLAT S., *A Wavelet Tour of Signal Processing*, Academic Press, New York, 2008.

[MIW 04] MIWAKEICHI F., MARTINEZ-MONTES E., VALDÉS-SOSA P.A., *et al.*, "Decomposing EEG data into space-time-frequency components using parallel factor analysis", *NeuroImage*, vol. 22, no. 3, pp. 1035–1045, 2004.

[MAI 09] MAIRAL J., PONCE J., SAPIRO G. *et al.*, "Supervised dictionary learning", *Advances in Neural Information Processing Systems*, pp. 1033–1040, 2009.

[MAL 93] MALLAT S., ZHANG Z., "Matching pursuit with time-frequency dictionaries", *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3414, 1993.

[NOI 08] NOIRHOMME Q., KITNEY R., MACQ B., "Single-trial EEG source reconstruction for brain–computer interface", *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 5, pp. 1592–1603, 2008.

[NUN 06] NUNEZ P., SRINIVASAN R., *Electric Fields of the Brain*, Oxford University Press, Oxford, New York, 2006.

[PER 87] PERRIN F., BERTRAND O., PERNIER J., "Scalp current density mapping: value and estimation from potential data", *IEEE Transactions on Biomedical Engineering*, vol. 34, no. 4, pp. 283–288, 1987.

[SPI 15] SPINNATO J., ROUBAUDM-C., BURLE B. *et al.*, "Detecting single-trial EEG evoked potential using a wavelet domain linear mixed model: application to error potentials classification", *Journal of Neural Engineering*, vol. 12, no. 3, p. 036013, 2015.

[SPU 15] SPUSTEK T., JEDRZEJCZAK W., BLINOWSKA K., "Matching Pursuit with Asymmetric Functions for Signal Decomposition and Parameterization", *PloS one*, vol. 10, no. 6, 2015.

[VAL 10] VALLAGHÉ S., PAPADOPOULO T., "A trilinear immersed finite element method for solving the electroencephalography forward problem", *SIAM Journal on Scientific Computing*, vol. 32, no. 4, pp. 2379–2394, 2010.

[WAB 07] WOLTERS C.H., ANWANDER A., BERTI G. *et al.*, "Geometry-adapted hexahedral meshes improve accuracy of finite-element-method-based eeg source analysis", *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 8, pp. 1446–1453, 2007.