Scuola Superiore Sant'Anna di Studi Universitari e Perfezionamento

Nicoló Valigi

Effects of napping on impulse control: insights from sleep Electrophysiology

Diploma di secondo livello in Ingegneria Industriale

Mentor:

Dr. Ugo Faraguna
SonnoLab
Università di Pisa

Tutor:

Prof.ssa Cecilia Laschi
Scuola Superiore Sant'Anna
Universitá di Pisa

Anno Accademico 2011-2012

A Ugo e al suo fantastico laboratorio.

Contents

Chapter 1

Introduction

Many questions around the neuroscience of sleep remain unanswered, even tough they may very well turn out to be fundamental in understanding how the human brain works. New insights in the physiology of nightly processes may provide new avenues for piercing the mysteries of our most complex organ.

Afternoon napping has been shown to improve cerebral performance in a variety of areas, including memory. In this work, we will focus on the executive functions of the brain, more specifically on impulse control and inhibition. Following up on previous literature that proved how some sleep processes are localized in specific cortical areas, we try to establish a relation between nap physiology and impulse control performance.

Through the use of electroencephalography and several signal processing techniques, we will try to uncover metrics and patterns that correlate sleep activity with increased performance in the impulse control test.

Chapter 2

Medical background

As far as I know, the only reason we need to sleep that is really, really solid is because we get sleepy.

William Dement

The importance of sleep can perhaps be best understood in terms of the many biological disfunctions that its absence brings about. These include decreases in higher-level cognitive abilities, difficulties in concentration, and emotional disorders. Extreme sleep deprivation has also been connected to premature death, as proved by a fundamental experiment on rats [Rechtschaffen and Bergmann, 2002].

On the other hand, the medical community still has to tackle a huge number of open questions about the function of sleep itself. It is currently unclear whether sleep has a single function, multiple ones which co-evolved together, or we just do not have enough parts to piece the puzzle together.

This chapter briefly reviews the medical literature on the fronts that are relevant to this research effort:

- basic notions on sleep physiology and electroencephalography;
- the relationship between learning and localized brain activity during sleep;
- previous work on the effect of afternoon napping on impulse control.

2.1 Sleep electroencephalography

This section classifies the EEG patterns present in the waking and sleeping human brain and associates them with sleep physiology.

During wakefulness, the EEG contains two main patterns of activity:

alpha activity consists of regular waves with a frequency between 8 and 12 Hz. α activity is more clearly evident when the subject is resting and in the absence of mental or physical activity.

beta activity consists of low amplitude waves with a frequency between 13 and 30 Hz. β waves are irregular, consistent with our interpretation of mental activity as the combination of many neural circuits that are engaged asynchronously. β waves are more clearly evident when the subject is attentive or thinking.

During the course of a night, sleepers go through a number of different stages, that correspond to different behaviors and EEG patterns. More specifically, sleep takes place in *cycles*. Just after the onset of sleep, the subject advances from stage 1 to stage 4, then regresses to stage 2 and goes through the REM (Rapid Eye Movement) phase. While this first cycle lasts around one hour, the following ones grow longer as the night progresses, with longer REM phases but shorter deep sleep.

Stage 1 is the lightest form of sleep, during which the subject can be easiliy awoken. During stage 1 sleep, muscular activity slows down, as do the eyes. From the EEG standpoint, stage 1 sleep is similar to wakefulness, but the activity has a slower frequency. Theta waves, with a frequency between 3.5 and 7.5 Hz, are clearly visible.

Stage 2 comprises about 50% of the total sleep time during the night. Besides θ waves, the EEG also shows *K complexes*, patterns composed of one large negative swing followed by a positive one. EEG traces also contain *sleep spindles*, burst of wavy activity around 12-14 Hz.

Stage 3 and 4 are collectively referred to as *deep sleep*. Delta waves start to appear in stage 3. They are characterized by large voltage swings and frequency lower than 3.5 Hz, and make up almost all of the EEG activity in stage 4. Interestingly, brain areas with the highest activity during waking also produce the strongest δ waves.

REM comprises about 20% of night-time sleep time. EEG shows low-amplitude and high-frequency patterns, corresponding to increased brain activity.

2.2 Local sleep, learning, and napping

The seminal paper in [Huber et al., 2004] showed how learning tasks during wakefulness can trigger local phenomena in the brain cortex that correlate to increased performance. More specifically, the participants were assigned a motor learning task that was subjectively indistinguishable from the control task but activated specific brain regions.

Cortical activity during the night following the session was monitored with an EEG device, which showed increased amount of Slow Wave Activity in those same brain regions that were involved in the learning task. Figure 2.1 shows such difference in cortical activity between the learning task and the control.

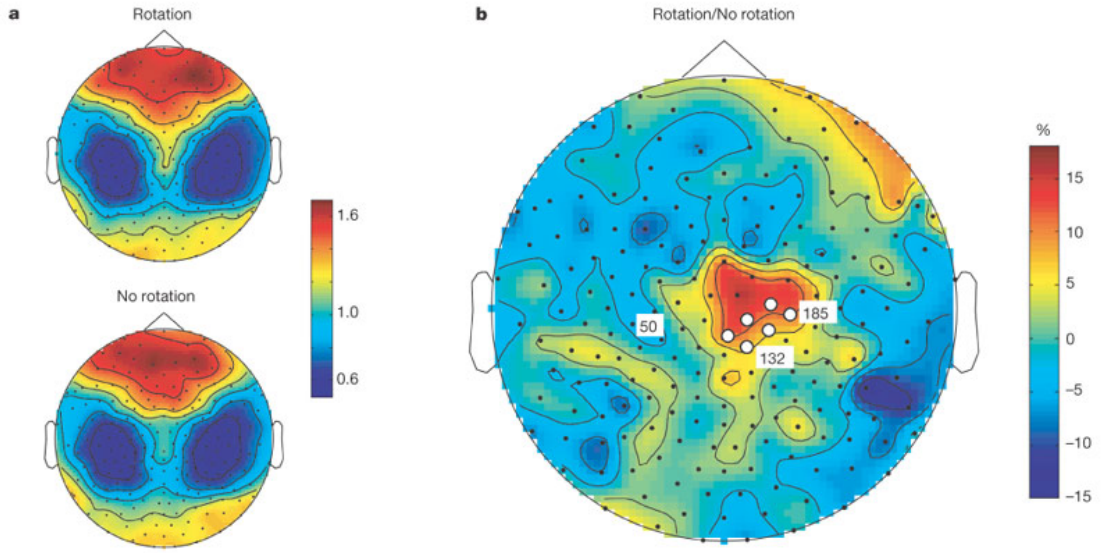


Figure 2.1: The intensity of slow wave activity as measured during the learning task, the control task, and their relative difference.

The important findings in this paper suggest that sleep homeostasis is regulated on local level, and may thus be related to learning processes, rather than other biological functions.

Napping

Napping has been strongly correlated with improvements in attention, memory, and wakefulness. For examples, [Garbarino et al., 2004] showed how afternoon napping can reduce the accident rate of professionals working the night shift.

Other studies, including [Akerstedt and Folkard, 1996] have proved that afternoon napping is regulated by circadian mechanisms, and is thus more effective just after lunchtime. Researchers have also linked napping to enhancements of declarative, procedural and motor memory. Such effects have been shown to depend on which sleep phases were experienced during the nap and thus on its length.

In connection with the sleep physiology outlined in section 2.1, the work in [Hayashi et al., 2005] has shown that naps shorter than 30 minutes mainly consist of stage 1 and stage 2. Furthermore, the authors have established that the wholesome properties of napping are due to the stage 2 period, even when its duration is restricted to 3 minutes.

2.3 Inhibition

The brain activities related to the management and control of other cognitive processes are often referred to as the *executive functions* of the brain. A widely accepted model identifies three main components:

working memory that holds relevant information for the task at hand and allows it to be updated when needed;

task switching , the ability to switch attention between one task and another;

inhibition of prepotent responses to external or internal stimuli.

Past research, such as [Miyake et al., 2000], has tried to assess the independence of individual executive functions and their effect on cognitive tasks.

However, this work is mainly concerned with the *inhibition* components and its relationship with sleep, and napping in particular.

From the psychological point of view, landmark works such as [Metcalfe and Mischel, 1999] have closely connected inhibition behaviors with willpower and self-control. From a physiological standpoint, much effort has been spent in the identification of the brain regions most closely connected with inhibition. It is now commonly thought that executive functions are housed in the prefrontal cortex. More specifically, studies using fMRI ([Rubia et al., 2003]) have shown that *motor* inhibition processes could be isolated in the right prefrontal cortex.

Chapter 3

The experiment

The experiment involved 14 subjects aged 26 ± 8.22 years, of which 8 men and 6 women. None of the participants reported regular napping in the afternoons.

Each subject participated in two testing sessions, spaced approximately one week apart, one of which was the control experiment.

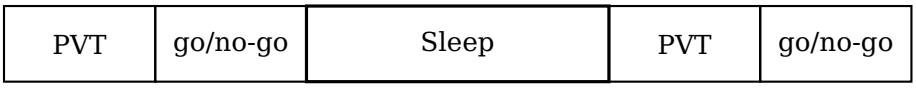


Figure 3.1: Session 1 (napping)

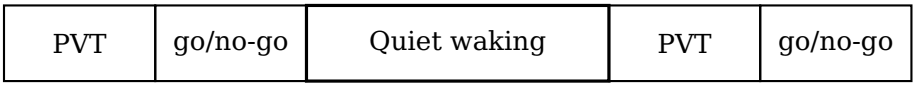


Figure 3.2: Session 2 (control)

The following sections detail each segment of the two experimental sessions.

PVT A Psychomotor Vigilance Test was used to measure vigilance and limit the effect of post-napping sleepiness in the go/no-go tests. We used a PC-based PVT software which, according to [Khitrov et al., 2014], had performance comparable to the industry gold standard, the PVT-192 hardware device. PVT measures a single quantity, the *response time*, from which other useful statistical measures can be extracted.

go/no-go To test stimulus inhibition performance, we used a go/no-go protocol similar to the one used in the landmark study of [Garavan et al., 1999]. According to this method, the subject is serially presented a sequence of the letters X and Y, at a speed of around 1/sec. The subject is instructed to click the mouse button when two consecutive letters are different (XY or YX) but make no such action when they

are equal (XX or YY). Each test consisted of 300 stimuli, 10% of which required a /emphno-go response, for a duration of 5 minutes.

sleep and quiet waking In the first session, subjects were encouraged to take a nap at around 3pm, lasting at most 1 hour. In the control session, they were allowed to use their phones but instructed to remain awake and not close their eyes. During both sessions, the subjects were monitored by EEG as described below.

Data acquisition

For the duration of the nap and the control waking period, we acquired 64-channel EEG data using a *Micromed BrainSpy Holter* EEG system. We used the standard 10-20 International System, with 2 additional electrodes on the eyes.

Chapter 4

Independent Component Analysis

Independent Component Analysis (ICA) is one of several signal processing techniques that solve the *blind source separation* problem. In particular, ICA has been applied very successfully in the analysis of EEG data. In the following sections, we will outline the mathematical background of ICA-based methods and investigate their usefulness in the context of our sleep experiment.

The most common didactical example of the usage of ICA is the *cocktail party* problem, in which we want to recover distinct audio signals given their sum. This corresponds to the real-life situation of recovering the individual voices of people talking at a party only by listening to a recording of their combination.

The following sections introduce the concept of ICA and compare it to other common approaches to signal analysis, such as Principal Component Analysis (PCA), and whitening. The material is largely based on the standard references on the subjects, mainly [Comon, 1994], [Hyvärinen et al., 2002] and on several of tutorials.

4.1 The ICA model

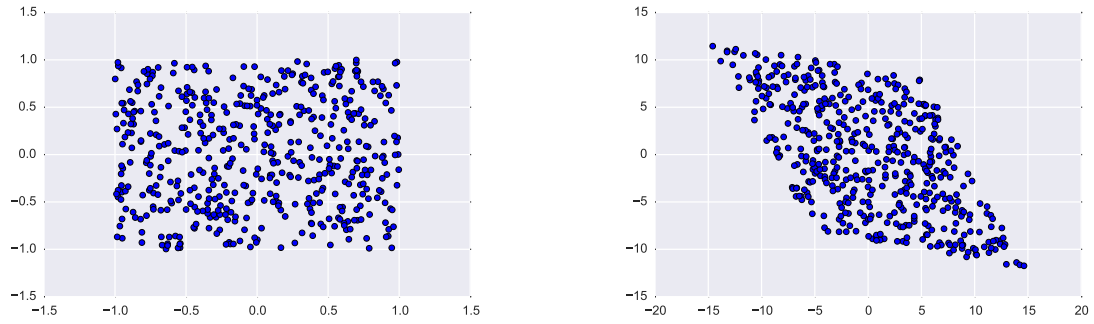
A *linear* mixture $\mathbf{x}$ of n independent components s_i can be expressed as:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{4.1}$$

Equation 4.1 is commonly refererred to as the generative *ICA model*. Both the *mixing matrix* $\mathbf{A}$ and the independent components *x* are unknown quantitites that must be recovered by the algorithm.

To visualize this, we can generate random 2-component vectors $[s_1, s_2]$ from an uniform bivariate distribution.

```
data = np.random.uniform(low=-1., high=1., size=(500, 2))
plt.scatter(*np.transpose(data))
```



(a) Example of bivariate uniform distribution.

(b) After the linear mixing, the vectors are not independent anymore.

Figure 4.1: Linearly mixing a bivariate uniform distribution loses the independence of its components.

Figure 4.1a confirms that s_1 and s_2 are indeed independent. On the other hand, we can suppose we can only observe the mixtures $\mathbf{x} = [x_1, x_2]$, whose two components are not independent anymore. Figure 4.1b shows one such mixture obtained using the mixing matrix given by:

$$A = \begin{bmatrix} -5 & 10 \\ 10 & -2 \end{bmatrix} \tag{4.2}$$

and the following code:

```
A = np.array([[-5,10],[10,-2]])
plt.scatter(*np.transpose(np.dot(data,A)))
```

As we will see in the following sections, the ICA transformation can retrieve the original (unmixed) components.

Required assumptions

The most important requirements for ICA to be applied successfully are the following:

- the components s_i must be statistically independent;
- the components s_i must be have *nongaussian* distributions
- the mixing matrix $\mathbf{A}$ is square

In the following sections, we will review each of these requirements.

Independent components

We consider two scalar random variables y_1 and y_2. By definition,

$$p_1(y_1) = \int p(y_1, y_2) dy_2 \tag{4.3}$$

We define x_1 and y_1 to be independent if and only if their joint probability distribution can be factorized as follows:

$$p(y_1, y_2) = p_1(y_1) p_2(y_2) \tag{4.4}$$

This means that the realization of y_1 does not affect the probability distribution of y_2, and viceversa.

Furthermore, given any two functions h_1 and h_2, the following equality holds if y_1 and y_2 are independent:

$$E\{h_1(y_1) h_2(y_2)\} = E\{h_1(y_1)\} E\{h_2(y_2)\} \tag{4.5}$$

Uncorrelatedness is a weaker form of independence in which $h_1(y_1) = y_1$ and $h_2(y_2) = y_2$.

Nongaussian independent components

Intuitively, one could say that Gaussian-distributed random variables do not have enough information in the higher order moments to allow ICA to make them out individually.

More specifically, we refer to the property of Gaussian distributions that uncorrelatedness implies independence. This means that ICA can not go any further than whitening and only determines the mixing matrix up to an orthogonal transformation (since any orthogonal transformation of the PDF does not alter the PDF itself at all). In other terms, the PDF of uncorrelated Gaussian variables has rotational symmetry, making it impossible to recover the direction of the columns of the mixing matrix $\mathbf{A}$.

Maximization of non-Gaussianity

An intuitive interpretation of ICA can be given in terms of maximization of *non-Gaussianity* of the independent components. We have already seen how ICA is impossible when the independent components are Gaussian distributed. To develop this idea further, we refer to the *central limit theorem*.

In particular, the central limit theorem states that the sum of independent random variables tends towards a Gaussian. Armed with this insight, we note that any linear mixture of the independent components is intuitively “more Gaussian“ than any of the IC themselves.

In this sense, we can define ICA as the search for the directions (IC) that maximize non-Gaussianity. There are $2n$ such local maxima in the optimization spectrum, corresponding to the set of s_i and $-s_i$.

Estimating non-Gaussianity

Kurtosis

Kurtosis is one way to quantify the non-Gaussianity of a probability distribution. Kurtosis is a normalized fourth moment:

$$\mathrm{kurt}(y) = E\{y^4\} - 3(E\{y^2\})^2 \tag{4.6}$$

If we suppose that the random variable has been normalized so that $E\{y^2\} = 1$, kurtosis is just $E\{y^4\} - 3$. While the kurtosis of a Gaussian distribution is zero, all other distributions can be grouped as *subgaussian* ($\mathrm{kurt} < 0$) and *supergaussian* ($\mathrm{kurt} > 0$). Supergaussian distributions have heavy tails, with high PDF both near zero and for large values of the variable. On the other hand, subgaussian distributions have flatter PDF, with constant values near zero and very low values for extreme values of the variable.

Kurtosis is a very simple method which is straightforward and efficient to compute. However, it is not very robust, in the sense that it is very sensitive to outliers (due to the fourth power).

Negative entropy

The negative entropy measure is a robust estimator for the gaussianity of a distribution, even though it is unfortunately complex to compute. From information theory, we know that, for a given variance, the Gaussian variable has the highest entropy. Statistically, *negentropy* is the most rational way to quantify gaussianity of a distribution:

$$J(\mathbf{y}) = H(\mathbf{y}_{\mathrm{gauss}}) - H(\mathbf{y}) \tag{4.7}$$

In real world applications, negentropy must be approximated in some way. The simplest one uses higher order moments and is mostly equivalent to the kurtosis criterion outlined above:

$$J(y) \sim \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} \mathrm{kurt}(y)^2 \tag{4.8}$$

and suffers from the same robustness issues we touched upon earlier.

To improve on this approximation, we can substitute y^3 and y^4 with non-polynomial functions G_1 and G_2:

$$J(y) \sim k_1(E\{G^1(y)\})^2 + k_2(E\{G^2(y)\} - E\{G^2(\nu)\})^2 \tag{4.9}$$

where ν is a Gaussian random variable that is used to normalize the metric.

Any function that grows more slowly than a third or fourth degree polynomial is interesting from a robustness perspective, for example:

$$G_1(y) = \frac{1}{\alpha_1} \log \cosh a_1 y$$
$$G_2(y) = -\exp(-y^2/2)$$

4.2 ICA and PCA

The goal of Principal Component Analysis is similar to ICA's but both the methodology and the outcomes are different. PCA focuses on dimensionality reduction, and achieves it by reducing redundancy as measured by correlations between data elements. On the other hand, ICA is based on the stronger concept of independence, and does not focus as much on dimensionality reduction.

For example, PCA can be used to reduce the amount of bits needed to transfer an 8x8 image by assembling all the pixel values into a 64-element vector. Since neighboring elements are strongly correlated, the first 10 principal components can be enough to reconstruct a faithful reproduction of the original signal.

Mathematically, PCA is an orthogonal transformation designed to extract a set of linearly uncorrelated variables (called *principal components* from a dataset. Such principal components are built in such a way that the first one shows the greatest variance in the data. In turn, each of the following components is aligned with the direction of maximum variance under the constraint of being orthogonal to the previous one.

PCA also has a straightforward geometrical interpretation: each principal component is an axis of the n-dimensional ellipsoid that best fits the dataset.

Let us consider a linear combination between the $\mathbf{x}$ vector and a vector $\mathbf{w}$ of weights:

$$y_1 = \mathbf{w}_1^T \mathbf{x} \tag{4.10}$$

The first principal component of $\mathbf{x}$ is the y_1 with the greatest variance, under the obvious constraint that the $\mathbf{w_1}$ has unit norm. Formally, the *PCA criterion* states that

$$J_1^{PCA}(\mathbf{w}_1) = E\{y_1^2\} = \mathbf{w}_1^T E\{\mathbf{x}\mathbf{x}^T\}\mathbf{w}_1 = \mathbf{w}_1 \mathbf{C}_x \mathbf{w}_1 \qquad \text{s.t.} \qquad ||\mathbf{w}_1|| = 1 \tag{4.11}$$

A study of Rayleigh quotients suggests that

$$\mathbf{w}_1 = \mathbf{e}_1 \tag{4.12}$$

so that

$$y_1 = \mathbf{e}_1^T \mathbf{x} \tag{4.13}$$

The following elements of the weight vector are computed by subtracting each principal component from $\mathbf{x}$. According to linear algebra, we find that:

$$\mathbf{w}_k = \mathbf{e}_k \tag{4.14}$$

Dimensionality reduction

The version of the PCA algorithm outlined above only maps the dataset to a space of new coordinates which are uncorrelated over the dataset. However, in practical applications we find that the importance of the principal components quickly decreases. This means that PCA can successfully used for *dimensionality reduction* to help make sense of higher dimensional data.

For instance, keeping the first two principal components selects the hyper-dimensional plane in which the variance of the dataset is highest. This helps in data exploration, as it can make any pattern more evident to an human observer.

Moreover, since we have seen how the principal component basis vectors $\mathbf{W}_i$ are eigenvectors of the covariance matrix, we can write

$$E\{y_m^2\} = E\{\mathbf{e}_m^T \mathbf{x}\mathbf{x}^T \mathbf{e}_m\} = \mathbf{e}_m^T \mathbf{C}_x \mathbf{e}_m = d_m \tag{4.15}$$

The formula above expresses the fact that the variances of the principal components are given by the eigenvalues of the covariance matrix. Principal components have zero mean by definition, so that a small value for the variance means they are very close to zero in norm.

The above discussion makes it clear how the sequence of eigenvalues of the covariance matrix can help in estimating the amount of information that would be lost when leaving out some of the last principal vectors.

4.3 ICA and Whitening

In this section, we show that whitening is a weaker version of ICA that relies on the concept of uncorrelatedness rather than independence. Two random variables y_1 and y_2 are uncorrelated if their covariance is zero:

$$E\{y_1 y_2\} - E\{y_1\}E\{y_2\} = 0 \tag{4.16}$$

Furthermore, we usually consider zero-mean random vectors, so that the covariance is equal to the correlation, and the uncorrelatedness condition can be written as:

$$E\{y_1\}E\{y_2\} = 0 \tag{4.17}$$

A zero-mean random vector $\mathbf{z} = (z_1, .., z_n)$ is *white* when its elements are uncorrelated and have unit variances. The problem can be stated as the search of a linear transformation $\mathbf{V}$:

$$\mathbf{z} = \mathbf{V}\mathbf{x} \tag{4.18}$$

that whitens the random vector $\mathbf{x}$.

We want to prove that:

$$\mathbf{V} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T \tag{4.19}$$

is a whitening transformation. Using the eigenvalue decomposition, we rewrite the covariance matrix as $\mathbf{C_X} = \mathbf{E}\mathbf{D}\mathbf{E}^T$, where $\mathbf{E}$ is the column-by-column matrix of eigenvectors and $\mathbf{D}$ is the diagonal matrix of eigenvalues of $\mathbf{C_X}$. Such a decomposition is guaranteed to exist since $\mathbf{C_X}$ is a positive semidefinite matrix. Moreover,

$$E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{V}E\{\mathbf{x}\mathbf{x}^T\}\mathbf{V}^T = \mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{E}\mathbf{D}\mathbf{E}^T\mathbf{E}\mathbf{D}^{-1/2} = \mathbf{I} \tag{4.20}$$

In some cases, it can be inefficient to re-compute the covariance matrix every time a new measurement is available. In others, the covariance matrix may not be stationary. Whatever the problem, on-line methods for PCA can be used to find a whitening transformation without the need to compute the covariance matrix.

Even though whitening is a simple and powerful transformation, it does solve the blind source problem. If the take a whitening transformation $\mathbf{V}$ as outlined above, we find that the generative model becomes:

$$\mathbf{z} = \mathbf{V}\mathbf{A}\mathbf{s} = \hat{\mathbf{A}}\mathbf{s} \tag{4.21}$$

Furthermore, any orthogonal transformation $\mathbf{U}$ such that $\mathbf{y} = \mathbf{U}\mathbf{z}$ satisfies the condition

$$E\{\mathbf{y}\mathbf{y}^T\} = E\{\mathbf{U}\mathbf{z}\mathbf{z}^T\mathbf{U}^T\} = \mathbf{U}\mathbf{I}\mathbf{U}^T = \mathbf{I} \tag{4.22}$$

meaning that $\mathbf{y}$ is also white if $\mathbf{z}$ is. In practice, this means that whitening only computes the independent components up to an orthogonal transformation.

In the following sections, we will see why whitening is an useful *preprocessing* step in ICA.

4.4 Preprocessing

When implementing ICA-based analysis workflows, some mathematical procedures are commonly employed to improve the quality of results. We will describe two of them.

Centering

All the mathematical derivations above are based on the assumption that the data has zero mean. This is not an limiting assumption, since the sample can always be centered by subtracting its sample mean. We also note that zero-mean mixtures $\mathbf{x}$ imply that the signals themselves must be zero-mean as well, since

$$E\{\mathbf{s}\} = \mathbf{A}^{-1}E\{\mathbf{x}\} \tag{4.23}$$

Whitening

Whitening is a very useful pre-processing step that can be carried out with a linear transformation. From the discussion above, we note that, if $\mathbf{V}$ is a whitening transformation, the new mixing matrix $\hat{\mathbf{A}} = \mathbf{VA}$ is orthogonal.

In concrete terms, this fact greatly reduces the dimensionality of the problem, since a $n \times n$ orthogonal matrix only has $n(n-1)/2$ degrees of freedom, not n^2.

Chapter 5

Data analysis

In terms of EEG activity, the results of the ICA analysis can be interpreted as follows:

- the rows of the $\mathbf{S}$ matrix contain the time evolution of the component activity
- the columns of the inverse $\mathbf{A}^{-}1$ mixing matrix are the scalp projections of the components

5.1 Artifact rejection

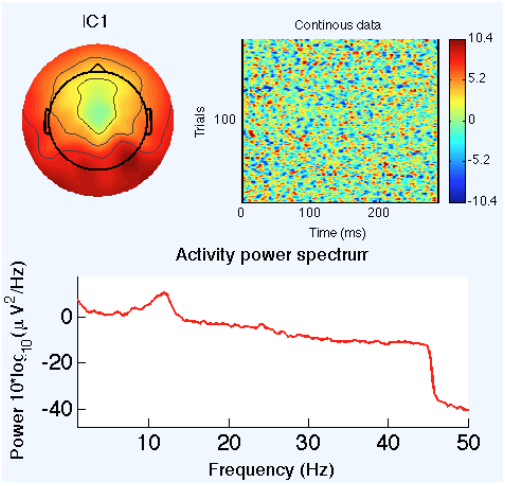
In the past, several methods have been proposed to clean up the EEG data and remove artifactual and spurious components, such as muscle noise, eye movements, and line noise. More recent developments have focused on leveraging the power of ICA to reject artifactual IC and recover a bigger portion of the original brain data.

In practice, some heuristics are needed to determine which of the independent components are due to artifactual events, and thus need to be removed. The following sections describe some of the methods we have collected from various sources.

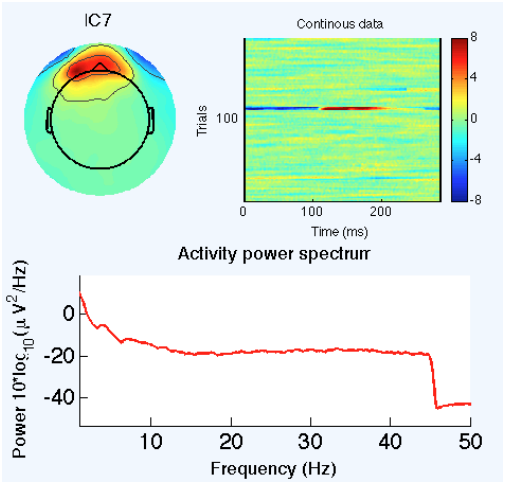
There exist also experimental methods for automatic rejection of bad epochs. They are usually based on thresholding of several parameters, including many that will be discussed below. A survey of such methods can be found in the NBT Wiki [NBT Tut. 2, 2015].

According to the NBT Wiki [NBT Tut. 1, 2015], the main steps involved in manual data cleaning with ICA are the following:

- run ICA
- detect artifactual IC
- reject filtered components
- finish processing



(a) This IC component is due to brain activity, as can be seen in the spectrum.



(b) This IC component is artifactual, due to its concentration in the eye region.

Figure 5.1: Examples of artifactual and brain-related components.

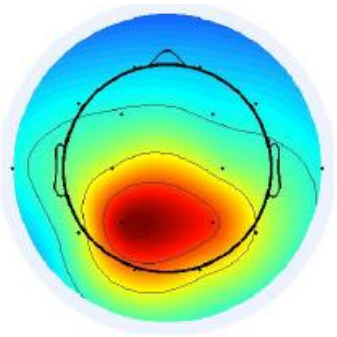
The detection of artifactual components is a tricky process that requires an experienced operator. Some guidelines for their identification are the following:

Localization and power spectrum Looking at the temporal plot of IC extracted from the raw data, it is quite hard to make out channels are due to brain activity and which are not. Usually, brain components have a clear rhythm, whereas artifactual components have sharp and/or abrupt variations in voltage. An clear example of such differences is shown in figure 5.1.

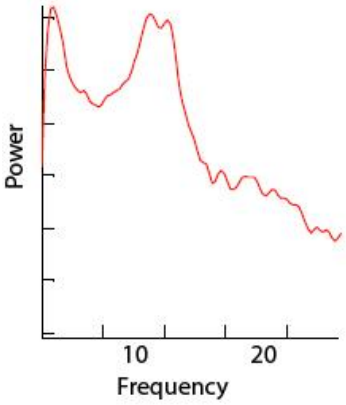
Gaussianity of the component Another way to identify artifactual components is to plot their probability density function: brain components usually are mostly Gaussian, while artifactual components are not. This difference is most evident on a QQ plot.

Examples of common components

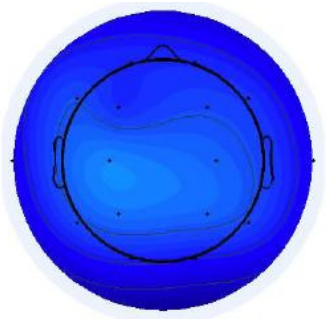
To aid in the identification of components, it is useful to refer to some common scenarios, such as heartbeats, eye blinks, and slow wave brain activity. In particular, figure 5.2 shows the topographic map and power spectra for several commonly occurring sources of scalp signals.



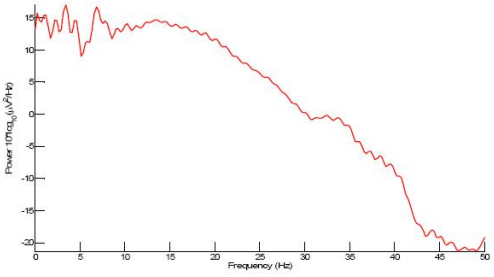
(a) Topographical map of an IC component due to α wave activity.



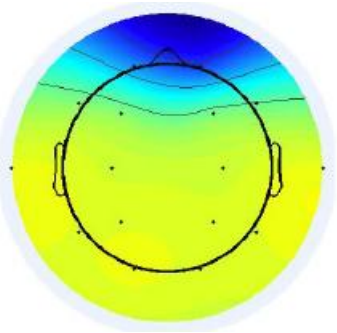
(b) Power spectrum of an IC component due to α wave activity.



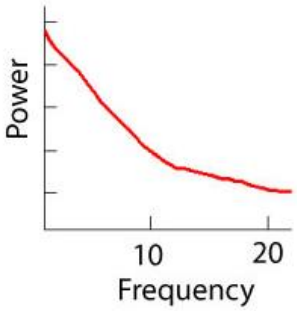
(c) Topographical map of an IC component due to the heartbeat.



(d) Power spectrum of an IC component due to the heartbeat.



(e) Topographical map of an IC component due to eye movements.



(f) Power spectrum of an IC component due to eye movements.

Figure 5.2: Fingerprints of independent components due to various kinds of physiological activity.

5.2 Analysis workflow

While the previous discussion focused on general concepts regarding EEG data cleaning and rejection, this section will detail the exact procedure that we followed before handling EEG data. The process consists of the following steps:

Preprocessing

1. First-order high pass filter at 0.1Hz to get rid of DC bias
2. Band pass filter between 0.3 and 30 Hz
3. Notch filter at 50 Hz to get rid of AC mains interference
4. ICA decomposition and rejection of artifactual channels
5. Regeneration of the original signal without the artifactual channels

Frequency domain analysis

1. Rejection of artifactual channels based on their power spectra (as outlined in the previous section) based on heuristics (such as power at high frequencies, or total power)

Slow wave analysis

1. *Chebichev* second-order filter between 0.5 and 4 Hz
2. Slow wave detection according to the guidelines in [Massimini, 2004]

5.3 The Multiple Comparison Problem

A typical EEG dataset for a single subject has very high dimensionality, due to the number of channels (64 to 128) and the time resolution of the instrument. Even in the frequency domain, one usually works with 0.16Hz frequency bins, and 30 second windowing period.

As outlined in [FT Tut. 1, 2015], the researcher must be extremely careful when drawing conclusions from such complex datasets. In the simple case in which single data points can be compared across different experimental conditions, common parametric tests such as the *t-test* can be used successfully.

However, when the researcher is looking for statistical evidence in the entire set of data points (in the order of thousands), some correction must be made to account for the *Multiple Comparison Problem*. Statistically, as the number of tested hypotheses increases, so does the likelihood of a rare event happening. This increases the likelihood of incorrectly rejecting the null hypothesis (a Type I error).

When testing multiple hypotheses, one usually considers the *Familywise error rate (FWER)*, defined as the probability of making one or more false discoveries

among all the hypotheses. There exist several procedures aimed at controlling the FWER.

The Bonferroni correction The Bonferroni correction is a straightforward way to account for multiple hypotheses. We state that the FWER can be controlled to FWER $\leq \alpha$, where α is the requested significance level, by choosing $pi \leq \frac{\alpha}{m}$, where m is the number of tested hypotheses. This statement can be proved by using the *union bound*, since the FWER is defined as the probability of making *at least one* Type-I error.

While the Bonferroni correction imposes no requirement on the structure of the test statistic, it unfortunately reduces the inference power of the data. In simple terms, one could look at the union bound as some sort of worst-case scenario in which the statistics are assumed to be all independent. For this reason, results obtained through the Bonferroni correction are usually conservative, especially as the number of tested hypotheses is large.

Due to this limitation, statisticians and practitioners alike have tried to develop alternative, more powerful, procedures that maintain more predictive power.

Resampling procedures

Two techniques based on resampling (see [Moore et al., 2010]) offer a more powerful alternative to the Bonferroni correction. Both procedures are powerful for a variety of reasons:

- They are *non-parametric*, i.e. they do not require any assumptions on the population distribution;
- Can be used in situations in which there is no analytic treatment to compute the p-values;
- Straightforward to use

Bootstrap The whole concept of frequentist inference is based on the sampling distribution of a test statistic (for example, the sample mean). The idea of boostrapping consists in sampling (with replacement) a number of new samples from the original data. These new samples are taken from the same probability distribution as the original distribution. This means that the bootstrap distribution of a statistic test equals the sampling distribution of the statistic.

In practice, one finds that the bootstrap distributions exhibits the theoretical properties of the sampling distribution. For example, the bootstrap distribution of the sample mean is Normal, according to the central limit theorem. In one sense,

the calculations required to compute the bootstrap distribution replace the theoretical machinery at the core of frequentist inference (central limit theory, mean, and standard deviation of the sample mean).

The rationale behind the bootstrap is that the different resamples can be interpreted as different realizations of the underlying random process that generated the original sample in the first place. The bootstrap distribution can be used to obtain confidence intervals for any test statistic, simply by performing a normal *t-test* on the bootstrap distribution itself.

Permutation testing Permutation testing follows the same idea of parametric significance tests, but uses resampling instead of theoretical results to estimate the sampling distribution. Since p-values are computed under the assumption that the null hypothesis is true, the resampling must be performed in a way that is consistent with the null hypothesis.

In practice, this means that the two populations must be identical when the null hypothesis is true or, equivalently, that they are exchangeable. This condition precludes the use of permutation testing in populations that differ even under the null hypothesis, because there is no way to sample that respects the null hypothesis.

Operationally, the main difference between bootstrapping and permutation testing is that the former samples *with* replacement, the latter *without*.

The general procedure for permutation testing is the following:

- Compute the test statistic for the original sample;
- Choose permutation resamples from the data, without resamples;
- Build the permutation distribution of the statistic with a large number of resamples;
- Find the p-value by comparing the original statistic with the permutation distribution.

Difference between the bootstrap and permutation testing As we have seen, the main practical difference between permutation testing (also called *randomization* and *bootstrapping* lies in the sampling procedure (with replacement vs. without replacement).

From a statistical standpoint, the permutation test is simpler to execute and provides exact probabilities, but is limited to null hypothesis testing under the assumptions we have described before. Bootstrapping is more general but the resulting probability estimates are only valid in the limit of very large samples.

Chapter 6

Results

This chapter presents the results of statistical analyses on the data obtained by behavioral testing and EEG recordings.

6.1 Sleep length and scoring

Sleep scoring is a manual process during which a skilled operator classifies different periods of sleep according to the most prevalent sleep patterns. The proportion of the different sleep stages during the napping day are summarized in figure 6.1.

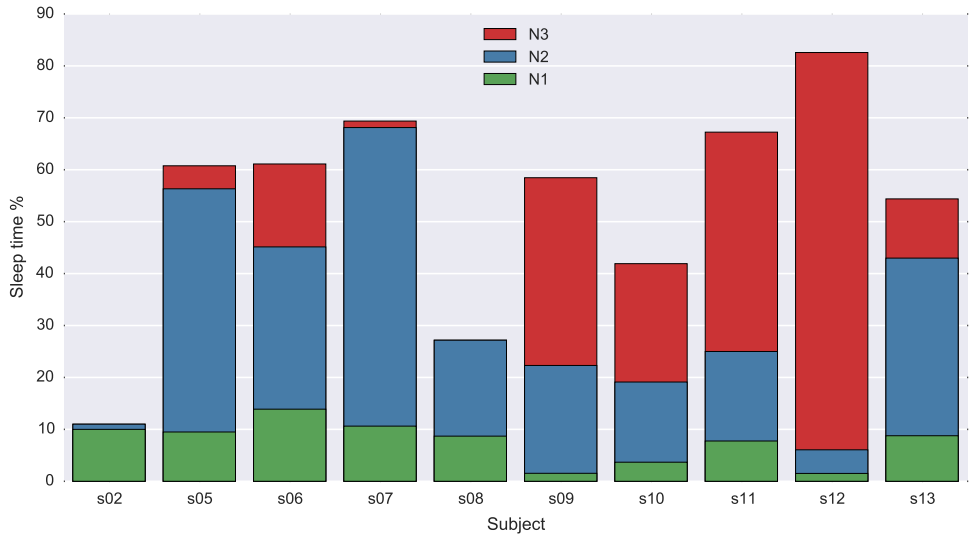


Figure 6.1: Sleep stages as identified for different subjects

6.2 Effect of napping on test performance

This section investigates the correlation between napping and performance in the inhibition test. Figure 6.2 summarizes the performance of subjects before and after the control day, and before and after the napping day.

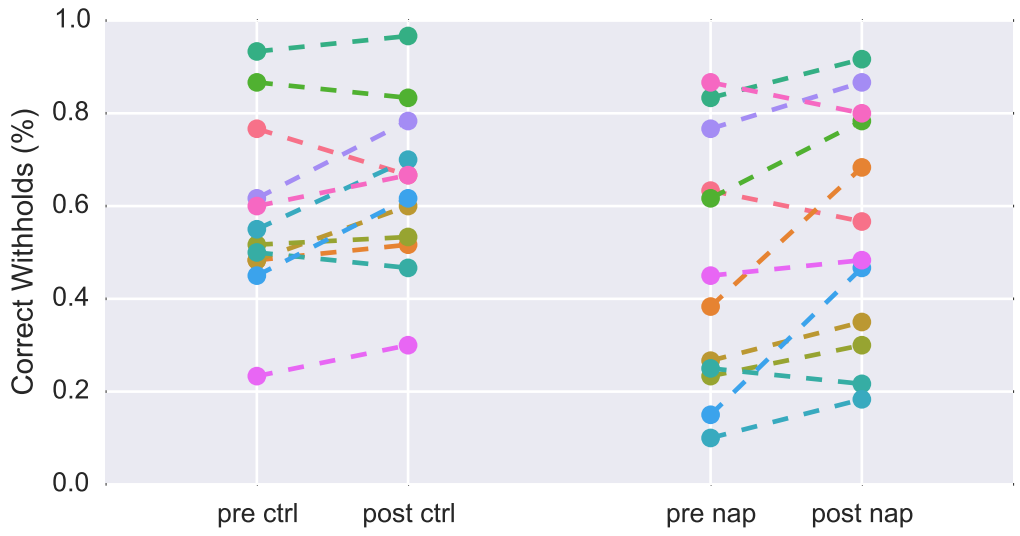


Figure 6.2: Effect of napping on inhibition test performance. Each color corresponds to a subject.

Performance generally had an higher baseline in the control day, probably due to training effects as napping was performed in the first session for most subjects.

To visualize the statistical significance of these results, we computed the probability density function for the changes (improvements) in inhibition performance before and after the napping (or control) period. The results are shown in figure 6.3.

Classical significance testing

Due to the presence of outliers, especially in the nap performance, using a *paired t-test* was deemed unreasonable since it requires Gaussian-distributed variables. We decided to use the Wilcoxon signed-rank test instead, which is non-parametric and makes no such assumption.

The computation was performed using the `scipy` Python package and its included `wilcoxon` function. We obtained a p-value of 0.0069 for the napping day, and 0.1872 for the control day, indicating that the improvement in the inhibition performance was indeed statistically significant.

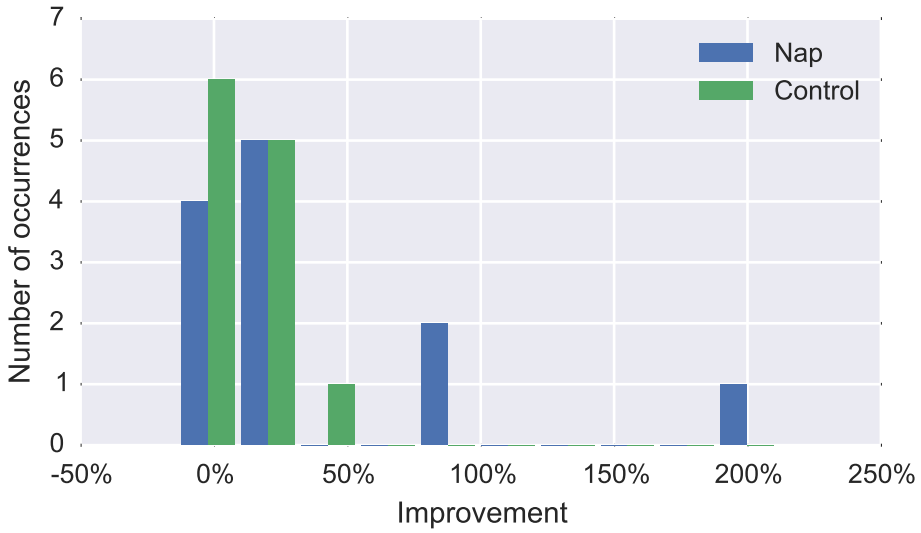


Figure 6.3: Probability density functions of performance improvements in the nap and control days.

Resampling-based significance testing

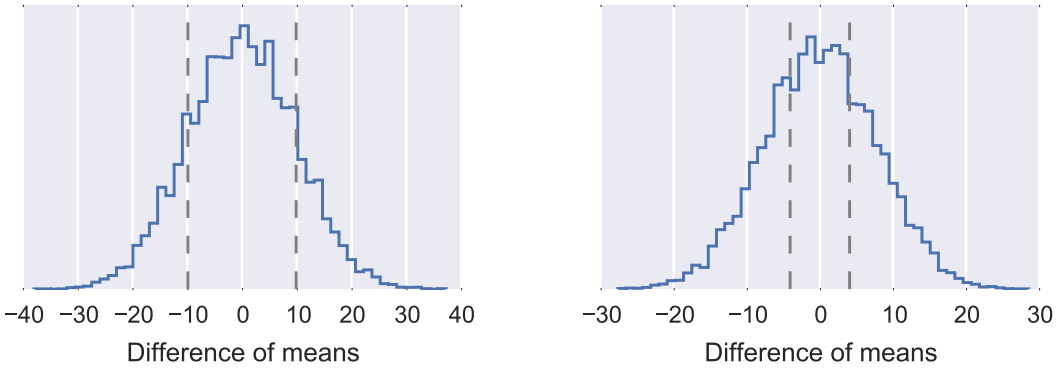
The permutation-based test presented in the previous chapter can be used to provide an additional point of view on the reliability of the data. The website in [Py Tut. 1, 2015] contains well-tested code for both bootstrap and permutation testing of datasets.

For each of the napping and control days, the permutation procedure mixes up the labels between the before and after condition, according to the null hypothesis. The results of the procedure are plotted in figure 6.4. The histogram represents the PDF of the difference of means before-after the napping (or control). The dashed vertical lines correspond to the observed value.

From the point of view of the permutation test, it would look like that the behavioral changes are not statistically significant (neither in the control day, nor in the napping day): the observed values are not really far enough into the tails of the sampling distribution. In the frequentist interpretation of statistic tests, this means that it would be fairly likely to observe these or larger values even if the null hypothesis was true.

The following table reports the p-values corresponding to figure 6.4.

condition	p-value
pre/post napping	0.313
pre/post control	0.600



(a) Sampling distribution for the null hypothesis in the napping case.

(b) Sampling distribution for the null hypothesis in the control case.

Figure 6.4: Sampling distributions for the mean of the performance in the control and napping cases.

Learning effect correction

One critique that could be levied against the previous approaches consists in the *learning* problem. Half of the subjects had the control day before the napping day, whereas the protocol for the other half was reversed. An argument could be made for performance improvements unrelated to the napping, but rather due to the subject's experience in the inhibition test.

To work around this problem, we carried out an analysis to find the optimal time-based *correction* to apply to the behavioral data. Inhibition performance before the first day's test and before the second day's test are presented in figure 6.5.

Test day	Correct responses rate
1st	41%
2nd	64%

To account for this difference, we introduce a *correction factor* in the ANOVA test:

$$0.636111/0.409722 = 1.55$$

6.3 Power spectrum analysis

As described in the previous chapter, the FFT is an useful tool to reduce the dimensionality of the investigation and focus the attention on the frequency bands that correspond to specific physiological activities.

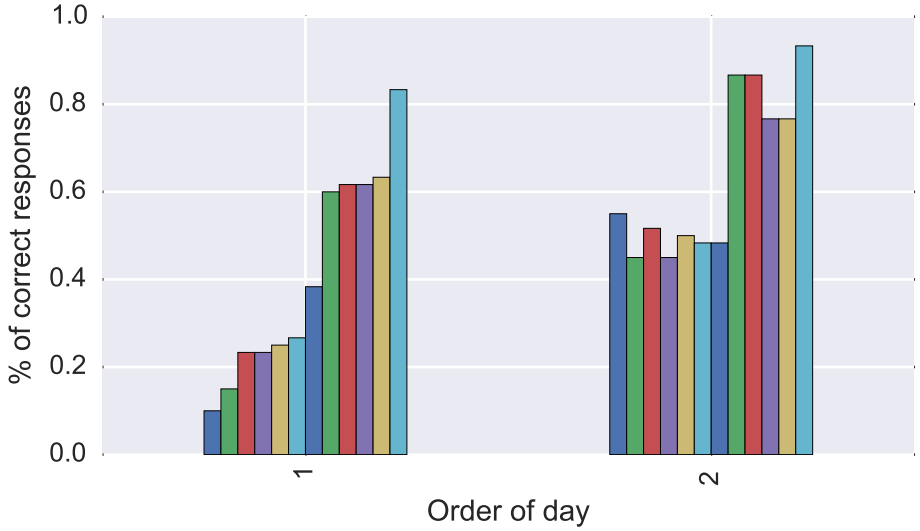


Figure 6.5: Inhibition performance in the first and second day of testing

In particular, given the sampling frequency and the number of samples, the EEG recording for each subject has the following format:

$$\sim 64 \text{ channels } \times 240 \text{ frequency bins} \times \sim 500 \text{ time periods}$$

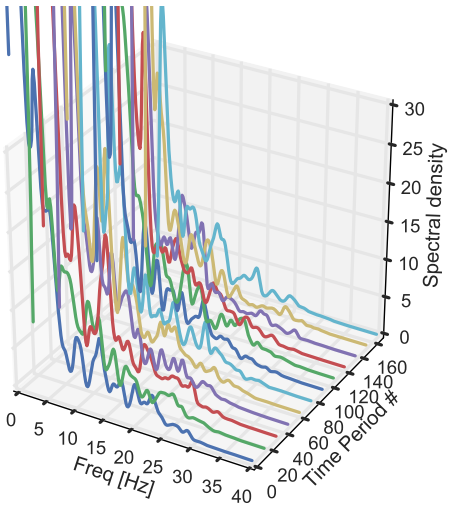
Each of the 240 frequency bins comprises a frequency range of around 0.16Hz, corresponding to the frequency range from 0 to 40Hz. As noted above, the number of channels in the original recording is 64, but the number can be reduced during the preprocessing stage as artifactual channels are removed. Likewise, individual recordings may have more or less periods depending on the duration of sleep.

This 3-dimensional data can be represented in a variety of ways. For example, figure 6.6 shows a sample of EEG data for one subject. Each colored line corresponds to the frequency spectrum in a single time period (the plot only shows 1 time period out of 10 for clarity).

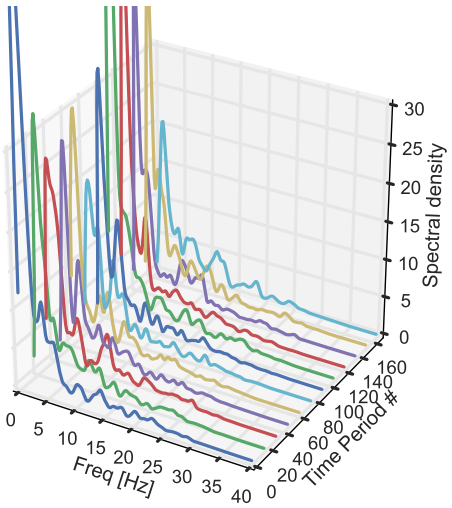
As can be seen in the leftmost image, the original data has very strong spikes due to physiological effects (electrode conduction, residual artifactual components, etc..). To reduce the magnitude of this problem, we have investigated algorithms for *normalization* of the data.

One promising solution, shown in the rightmost pane of figure 6.6 consisted in normalizing the *total power* in the band from 4-16Hz. The frequency band was chosen to minimize the effects of high-frequency artifactual components.

However, we did not proceed with normalization, and instead decided to carry out a preliminary investigation using the raw data.



(a) The non-normalized spectra shows huge spikes in the distribution.



(b) Normalization equalizes the different time periods.

Figure 6.6: Examples of normalized and normalized power spectra for one EEG channel.

6.4 Correlation with behavioral data

To proceed in the analysis, we tried to correlate activity in specific frequency bands to the performance in the inhibition test (as measured by the percentage of correct withdraws).

As a first step, we discarded information in the temporal dimension and averaged all time periods together. This process also had the convenient side-effect of eliminating differences in individual sleep length. Afterwards, it was necessary to create a single vector that would summarize the whole EEG information for a single subject. The most straightforward choice was to flatten the 2D structure in row-major order, ending up with a $240 \times 60 = 14400$ element 1D vector as follows:

$$[P_1 \; P_2 \; ... \; P_{60}]$$

with each $P_i \in \mathcal{R}^{240}$ containing the power densities of the i-th channel in each of the 240 frequency bins.

At this point, we can compute a 14400 element Person correlation vector by computing the subject-wise correlation between each scalar in the vector and its respective improvement in the inhibition test. The correlations can be easily visualized in a 2D heatmap, for example the one in figure 6.7.

The inspection of the heatmap suggests that the strongest correlations between EEG activity and test performance are concentrated around the 60th frequency bin

across most channels (around 10 Hz).

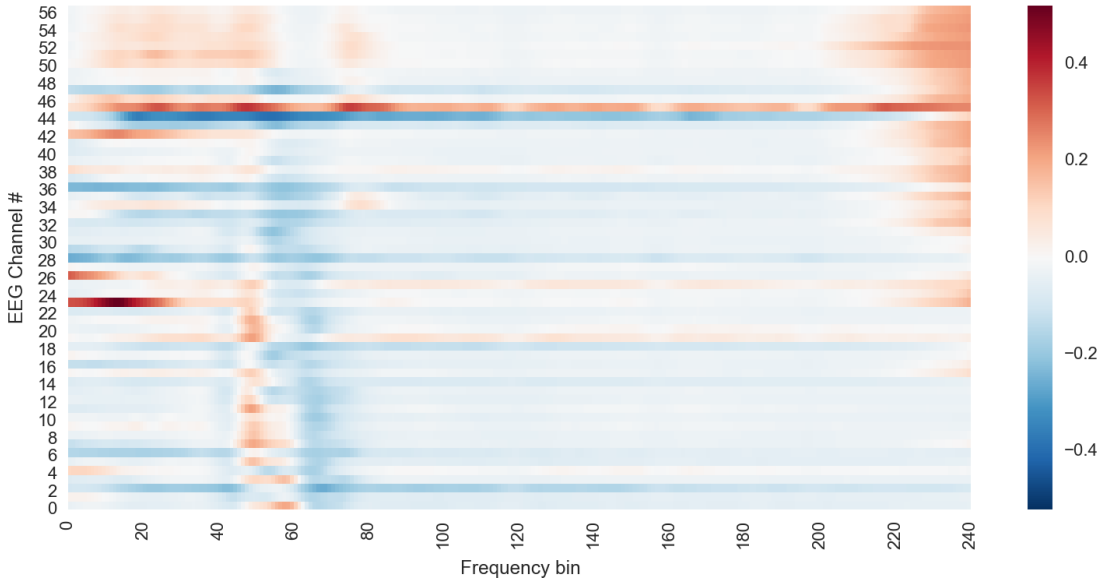


Figure 6.7: Example correlation map for one subject, averaged over the whole nap.

For significance testing, we followed the same permutation-based approaches outlined above. A random dataset was generated by taking all samples of the observed data and mixing them up across subjects, channels, and frequency ranges. This corresponds to the null hypothesis *a-priori* assumption. The procedure in the previous paragraph was repeated for each of these randomly-drawn samples. This allowed us to compute the *sampling distribution* of the correlation statistic and deduce the confidence level.

An example of the resulting sampling distribution is shown in figure 6.8, together with the observed value of the distribution. It should be noted that the resulting distribution is not Gaussian, further supporting our choice of using permutation based, rather than analytical, methods for significance testing.

Unfortunately, the significance analysis did not allow us to confidently rule out the null hypothesis, as the p-values were too high (only a few channel/frequency combinations with a p value between 0.1 and 0.2). Graphically, the situation was mostly similar to figure 6.8, with the observed correlation falling too far from the edges of the distribution to discard the null hypothesis.

6.5 Reducing dimensionality

Seeing as the previous approaches did not highlight appreciable correlations, we tried to reduce the dimensionality of the data to simplify statistical inference and

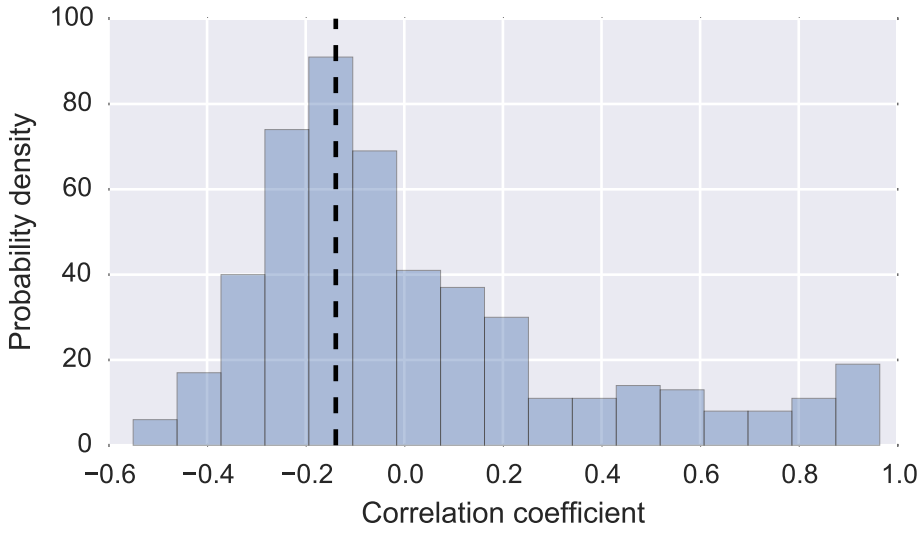


Figure 6.8: Example of the sampling distribution for the correlation of the 900th element of the FFT vector with test performance.

processing. Based on the results of the previous step, we limited our analysis to the lower frequency bands (from 0 to approximately 12Hz, corresponding to frequency bands 0-80).

The results of this restriction is shown in figure 6.9, similar to figure 6.7.

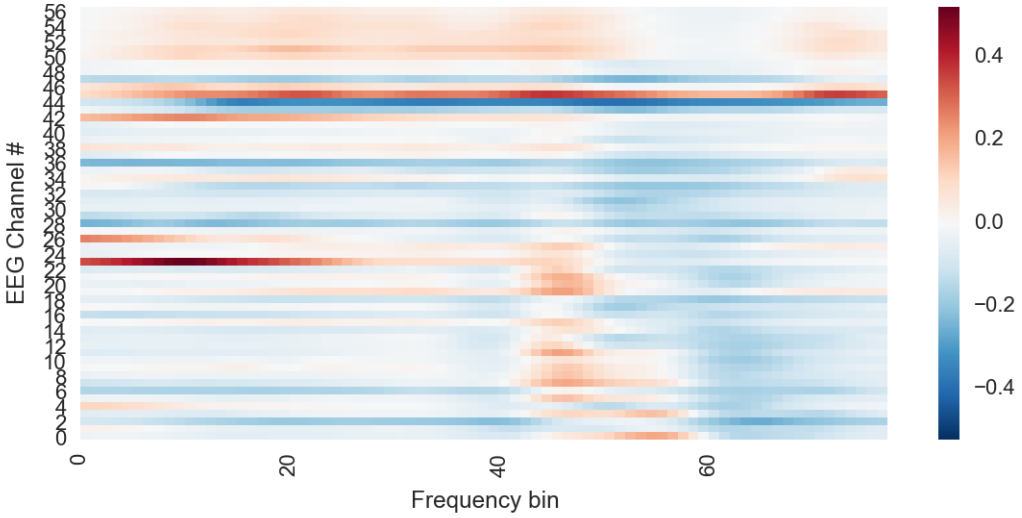


Figure 6.9: Correlation coefficient heatmap limited to the lower frequency bands.

Again, we repeated the same significance analysis and computed the p-value for the correlation statistic. Results are shown in the heatmap in figure 6.10.

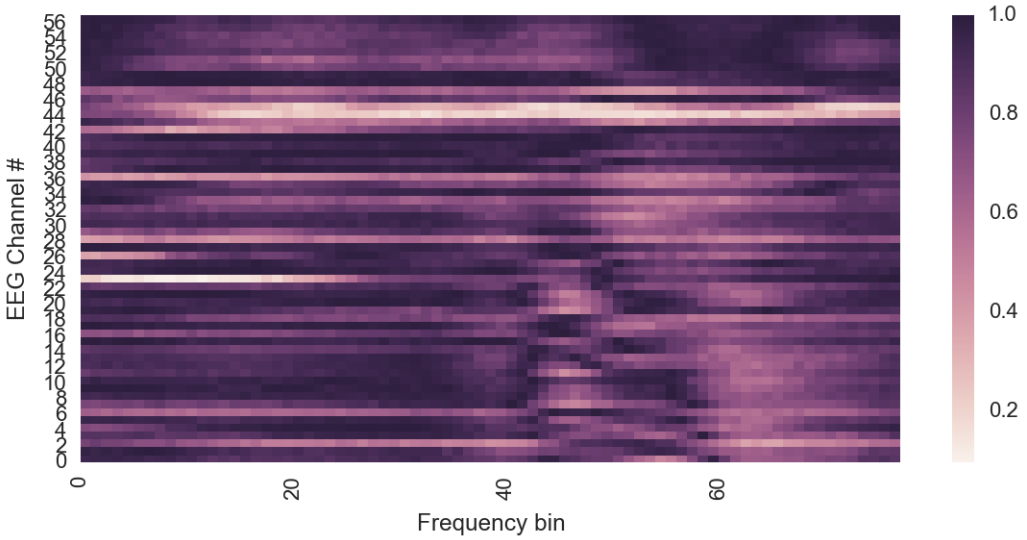


Figure 6.10: P-value heatmap limited to the lower frequency bands.

As can be seen using the scale on the right, the p-values are not low enough to discard the null hypothesis in any of the channel/frequency combinations. More specifically, none of the combinations had a p value below 0.09.

6.6 Limiting the scope

Given the preceding failures in detecting correlations, we decided to refocus our approach and drop the segmentation on frequency bands. Rather, we chose to compute a single scalar metric for each single channel, namely the sum of powers in the frequency bands up to 4 Hz. In essence, we decided to look for correlations in the aggregate, and not in each individual band.

Again, we repeated the random permutation to obtain the sampling distributions, obviously with much lower dimensionality (equal to the number of channels considered). An example is shown in figure 6.11.

Since the dimensionality of the data has been greatly reduced, it is now possible to plot all results in a scatter plot with low frequency energy (in arbitrary dimension) on the x axis and percentual improvement in the inhibition test on the y axis. Based on figure **??** and the corresponding numerical data, it is hard to discard the null hypothesis.

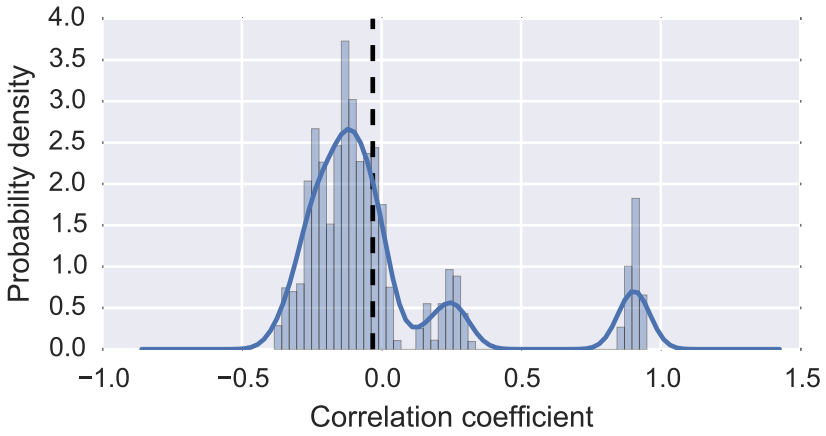


Figure 6.11: Sampling distribution for the correlation when considering energy at the lowest frequencies only.

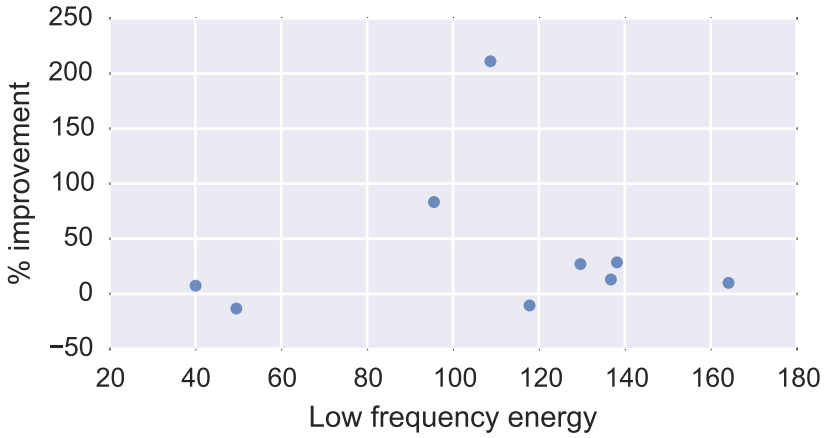


Figure 6.12: Correlation between test improvement and low frequency cumulative energy

6.7 Cluster-based correction

As discussed before, any statistical inference on the data must include a robust way to account for the multiple comparisons problem. We decided to follow the state of the art approach in [Bernardi et al., 2015] and use a *cluster-based correction*. Such correction pioneered by Forman in 1995, is based on the insight that activations in multiple neighboring channels make for a stronger statistical signal.

We used a Monte Carlo simulation with random permutations of the input data (in all channels, for each subject) to determine the shape of the null distribution.

Afterwards, we thresholded the activation of each individual channel at the 95% confidence level. We then looked at the maximum size of the remaining activated clusters. This procedure was repeated for 10 thousand times.

Thanks to this simulation, we obtained the sampling distribution of the maximum size of active clusters that one could expect, given that the null hypothesis was true. We could then establish a confidence interval of 95% on the shape of the distribution of the maximum cluster sizes themselves.

Chapter 7

Conclusions

We have presented several fundamental techniques that the scientific community uses to understand and quantify the complex signals recorded with an EEG system, including pre-conditiong and ICA analysis.

From the behavioral standpoint, we have analyzed the data to look for patterns of improvement correlated to afternoon napping.

Furthermore, we leveraged well-known signal processing techniques, such as the FFT, to draw inferences that would help predict test performance from the EEG recording.

Future opportunities for research

Even if the current analysis failed to turn out any evidence for correlation between test performance and EEG activity, we have explored but a very small corner of the huge number of metrics that can be extracted from EEG recordings. In particular, several refinements are in order:

- in the current analysis, we have discarded the effect of time and averaged together the EEG recordings across the whole duration of the nap. A more accurate analysis would account for different phases of sleep and for any time-dependent effects that may have gone unnoticed;
- in the absence of a sleeping control, we had to resort to cross-subject comparisons to assess any presence of correlation. Due to high individual variations, such comparisons are difficult to implement in practice. A new experiment design could introduce a control day in which subjects are asked to nap without having been subjected to the impulse inhibition test;
- some works in the literature have focused on other metrics of neuronal activity, for example θ wave activity. Since those can be automatically detected using

well-known algorithms, one could decide to look for effects on these higher-order effect, rather than at the intensity level.

Permutation testing code

```
import numpy as np
import numpy.random as npr
import pylab

def permutation_resampling(case, control, num_samples, statistic):
    """Returns p-value that statistic for case is different
    from statistc for control."""

    observed_diff = abs(statistic(case) - statistic(control))
    num_case = len(case)

    combined = np.concatenate([case, control])
    diffs = []
    for i in range(num_samples):
        xs = npr.permutation(combined)
        diff = np.mean(xs[:num_case]) - np.mean(xs[num_case:])
        diffs.append(diff)

    pval = (np.sum(diffs > observed_diff) +
            np.sum(diffs < -observed_diff))/float(num_samples)
    return pval, observed_diff, diffs
```

Bibliography

[Akerstedt and Folkard, 1996] Akerstedt, T. and Folkard, S. (1996). Predicting duration of sleep from the three process model of regulation of alertness. *Occupational and environmental medicine*, 53(2):136–141.

[Bernardi et al., 2015] Bernardi, G., Siclari, F., Yu, X., Zennig, C., Bellesi, M., Ricciardi, E., Cirelli, C., Ghilardi, M. F., Pietrini, P., and Tononi, G. (2015). Neural and Behavioral Correlates of Extended Training during Sleep Deprivation in Humans: Evidence for Local, Task-Specific Effects. *The Journal of Neuroscience*, 35(11):4487–4500.

[Comon, 1994] Comon, P. (1994). Independent component analysis, A new concept? *Signal Processing*, 36(3):287–314.

[FT Tut. 1, 2015] FT Tut. 1 (2015). Cluster-based permutation tests on time-frequency data. `http://www.fieldtriptoolbox.org/tutorial/cluster_permutation_freq`. Accessed: 2015-11-27.

[Garavan et al., 1999] Garavan, H., Ross, T. J., and Stein, E. a. (1999). Right hemispheric dominance of inhibitory control: an event-related functional MRI study. *Proceedings of the National Academy of Sciences of the United States of America*, 96(14):8301–8306.

[Garbarino et al., 2004] Garbarino, S., Mascialino, B., Penco, M. A., Squarcia, S., De Carli, F., Nobili, L., Beelke, M., Cuomo, G., and Ferrillo, F. (2004). Professional shift-work drivers who adopt prophylactic naps can reduce the risk of car accidents during night work. *Sleep*, 27(7):1295–1302.

[Hayashi et al., 2005] Hayashi, M., Motoyoshi, N., and Hori, T. (2005). Recuperative power of a short daytime nap with or without stage 2 sleep. *Sleep*, 28(1):829–836.

[Huber et al., 2004] Huber, R., Felice Ghilardi, M., Massimini, M., and Tononi, G. (2004). Local sleep and learning. *Nature*, 430(6995):78–81.

[Hyvärinen et al., 2002] Hyvärinen, A., Karhunen, J., and Oja, E. (2002). *Independent component analysis*.

[Khitrov et al., 2014] Khitrov, M. Y., Laxminarayan, S., Thorsley, D., Ramakrishnan, S., Rajaraman, S., Wesensten, N. J., and Reifman, J. (2014). PC-PVT: a platform for psychomotor vigilance task testing, analysis, and prediction. *Behavior research methods*, 46(1):140–7.

[Massimini, 2004] Massimini, M. (2004). The Sleep Slow Oscillation as a Traveling Wave. *Journal of Neuroscience*, 24(31):6862–6870.

[Metcalfe and Mischel, 1999] Metcalfe, J. and Mischel, W. (1999). A Hot/Cool-System Analysis of Delay of Gratification: Dynamics of Willpower.

[Miyake et al., 2000] Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, a. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognitive psychology*, 41(1):49–100.

[Moore et al., 2010] Moore, D., McCabe, G., and Craig, B. (2010). *Introduction to the Practice of Statistics: w/Student CD*. W. H. Freeman.

[NBT Tut. 1, 2015] NBT Tut. 1 (2015). Computing independent component analysis. `https://www.nbtwiki.net/doku.php?id=tutorial:compute_independent_component_analysis`. Accessed: 2015-11-02.

[NBT Tut. 2, 2015] NBT Tut. 2 (2015). Automatic and semi-automatic methods for eeg pre-processing. `https://www.nbtwiki.net/doku.php?id=tutorial:automatic_and_semi-automatic_methods_for_eeg_pre-processing`. Accessed: 2015-11-02.

[Py Tut. 1, 2015] Py Tut. 1 (2015). Data analysis with python. `http://people.duke.edu/~ccc14/pcfb/analysis.html`. Accessed: 2015-11-29.

[Rechtschaffen and Bergmann, 2002] Rechtschaffen, A. and Bergmann, B. M. (2002). Sleep deprivation in the rat: an update of the 1989 paper. *Sleep*, 25(1):18–24.

[Rubia et al., 2003] Rubia, K., Smith, A., Brammer, M., and Taylor, E. (2003). Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. *NeuroImage*, 20(1):351–358.