Universitat
de les Illes Balears

Improved detection of collective rhythms in multi-channel electroencephalography signals

Alejandro Morán Costoya

Master's Thesis

Master's degree in Physics of Complex Systems
at the
UNIVERSITAT DE LES ILLES BALEARS

Academic year 2016 - 2017

Date *18/09/2017*

UIB Master's Thesis Supervisor: Miguel Cornelles Soriano

Abstract

This work focuses on the experimental data analysis of phase-synchronized oscillators and, more specifically, electroencephalography data, in which multiple sensors are recording oscillatory voltage time series. The electroencephalography data analyzed in this dissertation were recorded by us through a commercial headset. Our goal is to optimally estimate the phase of phase-synchronized oscillators from noisy, phase lagged multivariate time series, which may be non-stationary. In other words, we want to recover the dynamics encoded in the noisy data of a system that behaves as a "common oscillator", which we cannot measure directly. To this end, we review some concepts and methods of signal processing, linear algebra and statistics, which we found necessary for the subject to be discussed. Traditional methods like principal and independent component analysis are compared to more recent approaches to extract a collective rhythm from phase-synchronized data. Furthermore, we reproduce and extend the work by Schwabedal and Kantz (PRL 116, 104101 (2016)) evaluating the performance of the Kosambi-Hilbert torsion method to extract a collective rhythm from multivariate oscillatory time series and comparing it to results obtained from principal component analysis. Their method generalizes singular value decomposition to account for possible phase lags among different time series and allows to focus the analysis on a specific spectral band, optimally amplifying the signal-to-noise ratio of the common rhythm. We found an improvement in the signal-to-noise ratio with respect to principal component analysis when the Kosambi-Hilbert torsion is applied to both synthetic and experimental data, namely, that the phase estimation is also improved.

Acknowledgements

En especial, me gustaría dar las gracias a mi tutor, Miguel Cornelles, por todo lo aprendido durante estos dos últimos años, sus ideas y su paciencia. Así como al resto de profesorado del Máster en Física de Sistemas Complejos por su labor.

Además, seguramente realizar este Máster habría sido más difícil sin los buenos ratos con mis compañeros y amigos, y el apoyo de mi familia y de Tania. Gracias a todos.

Contents

Chapter 1

Introduction

1.1 Motivation

Interpreting complex experimental data still poses a major challenge today and there are many tools developed for this purpose [1], but there is still much to learn and improve. In particular, the extraction of collective oscillations in coupled nonlinear systems from empirical data is typically unreliable [2]. In most cases, experimental data is contaminated by noise and other external sources, which may affect the performance of some studies and applications. We explore and test here recent improvements on the phase inference from such collective oscillatory signals in electroencephalography (EEG). In addition, such improvements are expected to work also for "low-cost" EEG headsets, which generally record lower SNR signals and have less measurement channels than professional devices.

In this introduction, we present a brief summary of several concepts that will be relevant for this dissertation.

1.2 Electroencephalography

EEG is a noninvasive method to record electrical signals coming from the brain activity. Neurons in our brain communicate with each other through ionic currents, which give raise to the voltage fluctuations measured in an EEG. These measurements, in most cases, take place in different places of the scalp through electrodes, so an EEG is an indirect measure of the brain activity. These measurements are superficial (2-dimensional instead of 3-dimensional) and the electrodes are placed above the muscles and skin, causing possible differences with intracranial measurements. Therefore, an EEG does not provide enough information to reconstruct what really happens in the brain. However, it is not all bad news and there is a lot to learn with EEGs. There are several advantages in comparison with other techniques such as magnetic resonance imaging (MRI) or functional MRI (fMRI): portability, cheaper hardware, high temporal resolution and real-time monitoring while any task is performed.

The EEG is the voltage difference over time between two different recording locations on the scalp. Actually, this is the same principle used in other bioelectrical recording techniques, such as electrocardiography (ECG), electromiography (EMG) or electroneurography (ENG). The rhythmical cortical oscillatory EEG activity emerges from an interaction between the thalamus and cortex[1] [3]. Scalp electrodes record mainly a combination of the inhibitory and excitatory postsynaptic potentials of neurons in the underlying cortex [4], which favours the slow, simultaneous potential charges generated in large cortical areas of pyramidal cells oriented in parallel at 90° to the plane of the scalp surface, facilitating spatial summation of the contribution to the current from each neuron towards the surface [5, 6]. However, in addition to neuronal activity in the surface of the cortex, scalp electrodes are also frequently affected by muscle fiber activity, which are referred to as artifacts.

The raw data measured in EEG is oscillatory, and it is common to examine the data for different frequency bands. A commonly studied frequency band is the alpha band, which corresponds to neural oscillations in

[1] The cortex is the outer layer of cerebral hemispheres composed of gray matter.

the frequency range of 8–13 Hz. In adults, the activity in this band is present being awake with the eyes open, and is strongly amplified when our eyes are closed, which we have studied here. Alpha brain waves are also present in some kinds of sleep and reversible coma. There are other frequency bands of interest, and some of them will be present in our data, e.g. delta (0.2–3 Hz) and theta (4–7 Hz) brain waves. We have recorded and analyzed both eyes closed and eyes open EEG activity (chapter 4). In the case of open eyes, we added a contribution at 15 Hz by watching a flickering screen that alternates black and white at this frequency.

1.3 Phase Synchronization

The concept of synchronization lacks a unique interpretation. Typically, synchronization refers to complete or partial coincidences among time-dependent states of systems or subsystems [7]. However, phase-synchronized systems do not necessarily exhibit typical synchronization because the amplitude dynamics may be not correlated. A dynamical system is phase-locked when all its degrees of freedom share the same phase, or more generally, phase-synchronized if phase differences among the degrees of freedom are bounded [8].

In general, any kind of synchronization is driven by interactions among systems or subsystems, which is known as *cooperative phenomena*. Also, cooperative behaviour can lead to *collective phenomena*, in which all types of synchronization can be included. In the case of phase synchronization, the time-dependent states of the system can be decomposed into several amplitude variables and one phase variable, i.e. there is a collective rhythm. Collective rhythms and phase synchronization were originally investigated to characterize chaotic oscillators [8]. However, here we will focus on collective oscillations from experimental data of spatially extended or networked systems.

Spatially extended or networked systems are systems measured from several locations through C channels to enhance our knowledge about the dynamics. Each channel records d-dimensional data, which oscillates in time. In the case of phase-locked systems, these measurements share phase with the other channel recordings. However, in practice, there exist measurement noise and phase lags among multivariate recordings, hindering the inference of the (possible) underlying collective rhythm. Some examples of spatially extended systems are: EEG data [9], locomotion data in the body frame of reference [10], ECG data [11] and volcanic tremor measurements [12]. In these systems the measurements can be simultaneous, e.g. EEG data, or not, e.g. locomotion data. As we will see in chapter 4, we will focus on EEG data.

Finally, we would like to highlight that the phase dynamics from unreliable multivariate time series, whose phases are not strictly the same due to measurement noise and phase lags, can indeed be used to reconstruct or estimate the actual, noiseless phase dynamics. One can interpret this as a procedure of information compression from the actual measurements, which are a set of C d-dimensional time series. As a result we can obtain a collective rhythm, which is a single k-dimensional time series, with $1 \leq k \leq d$, as we shall see in PCA (section 2.1) applied to multidimensional data.

1.4 The Analytic Signal

The analytic signal is a complex-valued function, $y_a(t)$, constructed from a real-valued function, an observable $y(t)$. The real part of $y_a(t)$ is the original real-valued function and the imaginary part is the Hilbert transform [13]:

$$y_a(t) = y(t) + i\mathcal{H}[y(t)]. \tag{1.1}$$

We are interested in the analytic representation of an arbitrary signal $y(t)$ since its representation in polar coordinates yields an amplitude envelope and an instantaneous phase representation, which can be useful to compare multivariate measurements of phase-locked systems. In this section, we introduce the Hilbert transform and indicate how to obtain the envelope and the phase.

1.4.1 Hilbert Transform

The Hilbert transform is a linear operator $\mathcal{H} : \mathbb{R} \to \mathbb{R}$, and it is defined as the convolution of the observable with the function $1/\pi t$:

$$\mathcal{H}[y(t)] = y(t) * \frac{1}{\pi t} = \frac{1}{\pi} \text{ P.V.} \int_{-\infty}^{\infty} \frac{y(\tau)}{t-\tau} d\tau. \tag{1.2}$$

Where P.V. indicates that the integral is taken as the Cauchy principal value [14]. Convolution in time by $1/\pi t$ is equivalent to a constant phase shift of $\pi/2$ for all frequencies and does not affect the amplitude.

Since the Hilbert transform is defined as the convolution indicated in equation 1.2, the Fourier transform, $\mathcal{F} : \mathbb{R} \to \mathbb{C}$, of $\mathcal{H}[y(t)]$ is the product of Fourier transforms

$$\mathcal{F}\left[\mathcal{H}[y(t)]\right](\omega) = \frac{1}{\pi} \mathcal{F}[y(t)](\omega) \mathcal{F}[t^{-1}](\omega). \tag{1.3}$$

Therefore, for physically relevant angular frequencies $\omega > 0$,

$$\mathcal{F}\left[\mathcal{H}[y(t)]\right](\omega) = -i\mathcal{F}[y(t)](\omega), \tag{1.4}$$

which means that, ideally, the Hilbert transform of the observable can be computed from a filter with unity amplitude response and constant $(\pi/2)$ phase response for all ω [15]. In the case of discrete signals defined in a bounded interval, the Hilbert transform can be computed approximately by means of a digital filter [16].

1.4.2 Amplitude Envelope and Instantaneous Phase

Relation (1.1) can be written as

$$y_a(t) = A(t) e^{i\phi(t)}, \tag{1.5}$$

where $A(t) = |y_a(t)|$ is the amplitude envelope and $\phi(t) = \arg\left[y_a(t)\right]$ is the instantaneous phase, which we illustrate in figure 1.1. Therefore, one can also reconstruct a real signal from these two magnitudes as $y(t) = A(t)\cos\left(\phi(t)\right)$. In the bottom pannel of figure 1.1, we show the instantaneous phase extracted from the amplitude modulated signal shown in the top pannel of this figure, whose frequency increases with time.

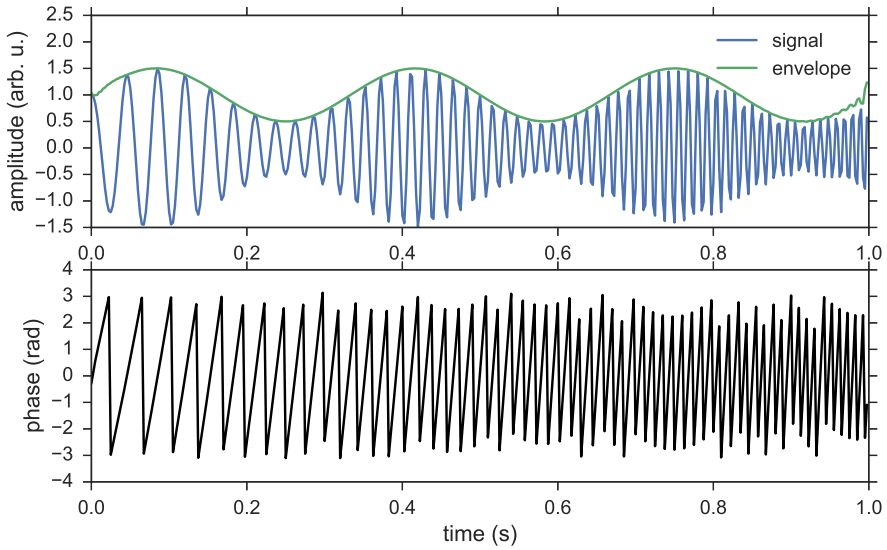

Fig. 1.1: Example of amplitude envelope (green) and instantaneous phase (black) extracted from an amplitude modulated signal with linearly increasing frequency (blue).

1.5 Signal-to-Noise Ratio

The signal-to-noise ratio (SNR) is a measure of the level of signal compared to the level of background noise of a time series, which in our case is evaluated in terms of signal and noise variance, although there are alternative definitions. Given a time series, the corresponding SNR is computed as the signal variance divided by the noise variance,

$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}}. \tag{1.6}$$

A high SNR indicates high precision data. The noise variance depends on the definition of noise, which in our case and for an arbitrary signal we define using bandpass filters at different frequencies and a given bandwidth. The level of noise then corresponds to the out-of-band variance, following the recommendations in [9]. The procedure to compute the SNR is the following:

1. Given a time series $x(t)$, select the desired center frequency and bandwidth to apply a bandpass filter to $x(t)$ and obtain $x^f(t)$, which is the filtered signal.
2. Compute the signal variance at the given center frequency as $\sigma^2_{signal} = \text{var}[x^f(t)]$.
3. Compute the noise variance at the given center frequency as $\sigma^2_{noise} = \text{var}[x(t) - x^f(t)]$.
4. Compute the SNR, see equation (1.6).

Here, everything that is not the signal within a given frequency range is considered to be noise. This procedure can be repeated for several center frequencies in the desired range to obtain the spectrum $SNR(f_c)$. As an example, in figure 1.2 we show the SNR for a superposition of three sinusoidal signals with amplitude modulation, with and without noise in the amplitude. This figure illustrates how the SNR decreases in presence of measurement noise. In this case, the SNR of the clean signal is approximately twice of the noisy one. The small amplitude discrepancies between the frequency peaks at 19, 20 and 21 Hz are due to finite sampling rate and the small side-peaks are due to the amplitude modulation.

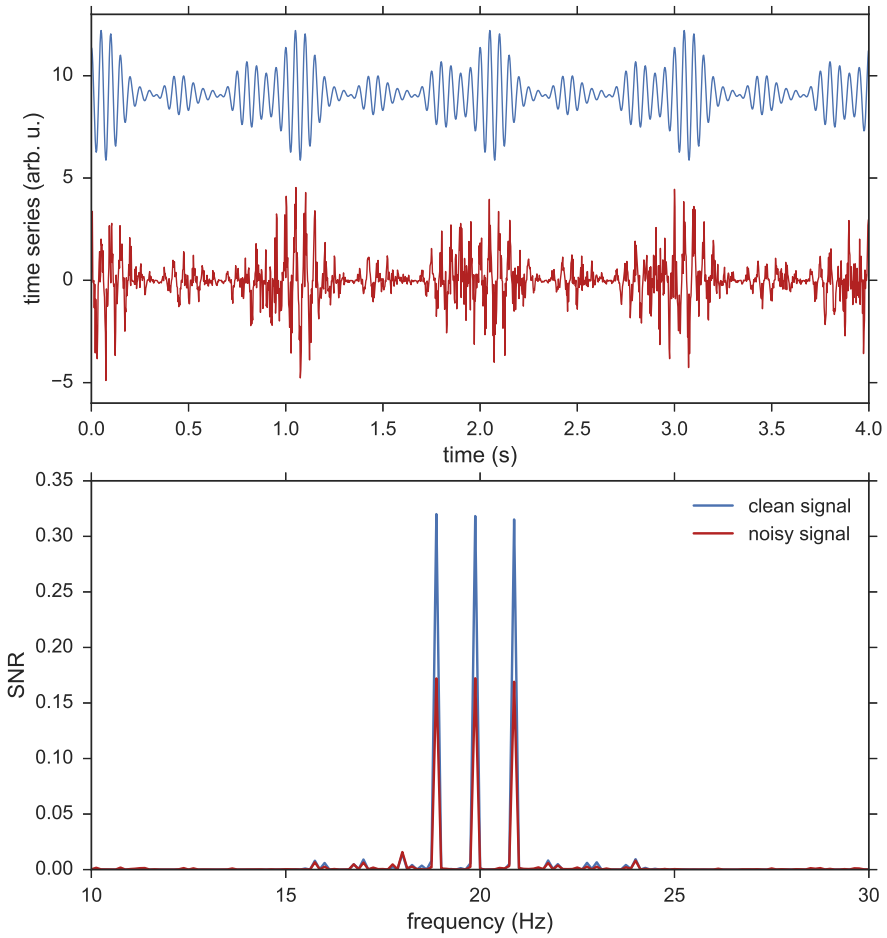

Fig. 1.2: (Top) The signals are a superposition of sinusoidal signals at 19, 20 and 21 Hz, and amplitude modulated with (red) and without (blue) noise. (Bottom) Example of SNR espectrum for a time series with (red) and without (blue) noise in the amplitude modulation.

1.6 Outline of this M.S. Thesis

This thesis is divided in four chapters. In this first chapter, we already introduced EEG, phase synchronization and basic theoretical concepts to understand the next chapter. Chapter 2 is devoted to review traditional and more recent methods for the extraction of collective rhythms from multivariate time series. In chapter 3, these methods are illustrated with 2-dimensional synthetic data to give a graphical intuition of how they work. Moreover, the synthetic data is used to show numerically the relation between the SNR and the phase estimation quality. Finally, in chapter 4, we show, explain and analyze EEG time series for two different experiments: brain activity with eyes closed and watching a flickering screen, which were recorded by us using a commercial EEG headset.

Chapter 2

Methods for the Extraction of Collective Rythms

The methods described in this chapter are:

1. Principal Component Analysis (1901) [17, 18].
2. Phaser Algorithm (2008) [10].
3. Kosambi-Hilbert Torsion (2016) [9].

2.1 Principal Component Analysis

Principal component analysis (PCA) is a non-parametric method to simplify multidimensional (noisy) experimental data. This type of data is often unclear and redundant. PCA is widely used in diverse fields such as taxonomy, biology, pharmacy, finance, agriculture, ecology, health and architecture [19]. In general, there are two main reasons to use PCA on real-world data:

- If we do not know which measurements reflect best the dynamics of the system or we do not have access to them, then we make arbitrary measurements on available states of the system.
- If we make arbitrary measurements, there may be more recorded dimensions than needed (redundancy).

2.1.1 Motivation

The goal of PCA is to determine which dynamics are important, which are redundant and which are just noise. We suggest the following toy example to understand the effect of applying PCA [20]. Imagine a ball of certain mass attached to a *massless*, *frictionless* spring, which oscillates along the x-axis about its equilibrium at a certain frequency. If we observe the motion of the ball using three cameras, as it is depicted in figure 2.1, we are measuring 6-dimensional data (there are 3 cameras, each recording 2-dimensional data) placed in the $\{\vec{a}, \vec{b}, \vec{c}\}$ arbitrary basis, while we only need to record the x-axis to capture the dynamics of the system, but the x-axis is in principle unknown. Applying PCA will provide us the orthonormal basis $\{\vec{x}, \vec{y}, \vec{z}\}$, in which the contributions in the directions $\vec{y}$ and $\vec{z}$ would be just noise or zero, revealing that the dynamics take place along the x-axis.

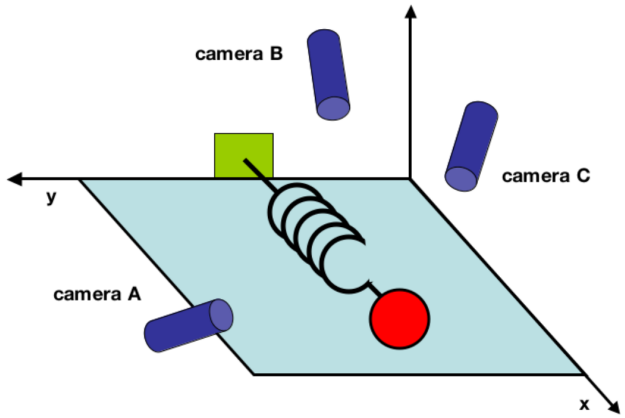

Fig. 2.1: Figure taken from [20]. Toy model example to motivate PCA.

More generally, we deal with measurements from certain number of detectors or channels, and each detector produces data in an arbitrary number of dimensions (e.g. position, velocity, voltage, etc.) at certain sampling rate, which is relevant in the case of temporal data.

Each sample corresponds to a column vector $\boldsymbol{x}_i$, in the case of figure 2.1 there are 3 channels recording 2-dimensional data, so $m = 3 \times 2 = 6$ and each sample corresponds to the column vector

$$\boldsymbol{x}_i^T = \begin{pmatrix} x_A & y_A & x_B & y_B & x_C & y_C \end{pmatrix}_i, \quad i = 1, 2, \dots n, \tag{2.1}$$

where n is the number of samples[1]. In the general case

$$\boldsymbol{x}_i^T = \begin{pmatrix} x_1 & x_2 & \dots & x_m \end{pmatrix}_i, \quad i = 1, 2, \dots n. \tag{2.2}$$

These column vectors are the columns of the data matrix $\boldsymbol{X}$, in which we collect the time evolution. In addition, the general case is neither restricted to a single oscillator, nor a fixed frequency or dimension, which means that the dynamics can be distinguished from the noise along more than one axis if it is required. In this work we will deal with examples that contain collective rhythms, so we link the common oscillator behaviour of the system to the principal component. The axis of the new basis which provides more information is the main principal component.

2.1.2 Change of Basis

The way to obtain the principal component(s) is via a change of basis, which in the case of PCA is a linear combination of the original basis. So, we define the linear transformation $\boldsymbol{P}$, such that

$$\boldsymbol{Y} = \boldsymbol{P}\boldsymbol{X}, \tag{2.3}$$

where $\boldsymbol{Y}$ is the convenient way to represent the data matrix $\boldsymbol{X}$. Equivalently,

$$\boldsymbol{y}_i = \begin{pmatrix} \boldsymbol{p}_1 \cdot \boldsymbol{x}_i \\ \vdots \\ \boldsymbol{p}_m \cdot \boldsymbol{x}_i \end{pmatrix}, \tag{2.4}$$

where $\boldsymbol{p}_i$ are the rows of $\boldsymbol{P}$ and $\boldsymbol{y}_i$ are the columns of $\boldsymbol{Y}$. In the following we will explain which is the best choice for $\boldsymbol{P}$.

2.1.3 Variance and Redundancy

Returning to the toy example in figure 2.1, each camera should show motion along a straight line, and deviations from the straight line correspond to noise. Assuming that the deviations are sufficiently small,

[1] e.g. if record data at 128 samples per second during 10 seconds, then $n = 1280$.

the dynamics take place in the direction with the largest variance, i.e. the ideal straight line, which we aim to recover from experimental data. Therefore, choosing a basis that maximizes the signal measured by the variance allows to distinguish the dynamics from the noise. Similarly, in the general case we can isolate the different contributions to the dynamics and the noise or less relevant dynamics.

Recall that the rows of $\boldsymbol{X}$ are measurements of a particular type, and the columns of $\boldsymbol{X}$ are all the measurements recorded in a given sample[2]. Therefore, the covariance matrix, which is defined as

$$\boldsymbol{C_X} = \frac{1}{n-1}\boldsymbol{X}\boldsymbol{X}^T, \tag{2.5}$$

is a symmetric $m \times m$ matrix in which the diagonal terms are the variance of particular measurement types and the off-diagonal terms represent the covariance between measurement types. Nonzero covariance indicate correlations between measurement types, which means that there is redundancy.

The goal will be to obtain an optimized covariance matrix $\boldsymbol{C_Y}$. This matrix should have minimum covariance (redundancy) and maximum variance (signal). Since $\boldsymbol{C_X}$ is symmetric and positive, diagonalization is always possible and results in null covariance and maximum variance. This result can be related to a more general formulation: singular value decomposition (SVD).

2.1.4 Singular Value Decomposition

SVD [21] is introduced in this section since it can be faster and is numerically more stable than direct diagonalization of the covariance matrix. We take as starting point the eigenvalue problem

$$(\boldsymbol{\chi}^T\boldsymbol{\chi})\hat{\boldsymbol{v}}_i = \sigma_i^2\hat{\boldsymbol{v}}_i, \tag{2.6}$$

where $\boldsymbol{\chi}$ is an $n \times m$ matrix, $\{\hat{\boldsymbol{v}}_1, \hat{\boldsymbol{v}}_2, \ldots, \hat{\boldsymbol{v}}_r\}$ is the set of orthonormal $m \times 1$ eigenvectors, σ_i are called singular values and $\boldsymbol{\chi}^T\boldsymbol{\chi}$ is a rank r, symmetric $m \times m$ matrix. We also define another set of orthonormal vectors $\{\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2, \ldots, \hat{\boldsymbol{u}}_r\}$, which are such that

$$\boldsymbol{\chi}\hat{\boldsymbol{v}}_i = \sigma_i\hat{\boldsymbol{u}}_i. \tag{2.7}$$

It is convenient to construct the $n \times m$ matrix $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Sigma} \equiv \begin{pmatrix} \sigma_{\tilde{1}} & & & & & \\ & \sigma_{\tilde{2}} & & & 0 & \\ & & \ddots & & & \\ & & & \sigma_{\tilde{r}} & & \\ & & & & 0 & \\ & 0 & & & & \ddots \\ & & & & & & 0 \end{pmatrix}, \tag{2.8}$$

so that the rank-ordered eigenvalues of $\boldsymbol{\Sigma}^2$ are $\sigma_{\tilde{1}}^2 \geq \sigma_{\tilde{2}}^2 \geq \ldots \geq \sigma_{\tilde{r}}^2 \geq 0$, and the corresponding orthogonal matrices $\boldsymbol{V}$ and $\boldsymbol{U}$ are constructed as

$$\boldsymbol{V} = \begin{pmatrix} \hat{\boldsymbol{v}}_{\tilde{1}} & \hat{\boldsymbol{v}}_{\tilde{2}} & \ldots & \hat{\boldsymbol{v}}_{\tilde{m}} \end{pmatrix}, \tag{2.9}$$

$$\boldsymbol{U} = \begin{pmatrix} \hat{\boldsymbol{u}}_{\tilde{1}} & \hat{\boldsymbol{u}}_{\tilde{2}} & \ldots & \hat{\boldsymbol{u}}_{\tilde{n}} \end{pmatrix}, \tag{2.10}$$

where $(m-r)$ and $(n-r)$ orthonormal vectors have been appended correspondingly to be consistent with the matrix notation

$$\boldsymbol{\chi}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{\Sigma}, \tag{2.11}$$

or equivalently,

$$\boldsymbol{\chi} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T, \tag{2.12}$$

[2]Here we assume that the samples are collected simultaneously, e.g. in the case of the toy example all cameras capture the motion at the same frequency.

which can be interpreted as a rotation, a stretch and a second rotation.

This abstract formalism has not been presented with a particular purpose and might not make sense or not seem to be related to PCA. However, multiplying equation (2.12) by $\boldsymbol{U}^T$ and defining $\boldsymbol{Z} \equiv \boldsymbol{\Sigma}\boldsymbol{V}^T$, we get the expression

$$\boldsymbol{U}^T\boldsymbol{\chi} = \boldsymbol{Z}. \tag{2.13}$$

Comparing to equation (2.3), $\boldsymbol{U}^T$ plays the same role than $\boldsymbol{P}$ because $\boldsymbol{U}^T$ describes a change of basis from $\boldsymbol{\chi}$ to $\boldsymbol{Z}$. Equivalently, if we define $\boldsymbol{Z} = \boldsymbol{U}^T\boldsymbol{\Sigma}$ instead, we get

$$\boldsymbol{V}^T\boldsymbol{\chi}^T = \boldsymbol{Z}. \tag{2.14}$$

The matrix $\boldsymbol{\chi}$ can be related to the data matrix $\boldsymbol{X}$ as

$$\boldsymbol{\chi} \equiv \frac{1}{\sqrt{n-1}}\boldsymbol{X}^T, \tag{2.15}$$

so that,

$$\boldsymbol{C}_{\boldsymbol{X}} = \boldsymbol{\chi}^T\boldsymbol{\chi}. \tag{2.16}$$

We already know that the principal components of $\boldsymbol{X}$ are the eigenvectors of $\boldsymbol{C}_{\boldsymbol{X}}$. We note that SVD (equation (2.12)) is equivalent to solving the eigenvalue problem that diagonalizes $\boldsymbol{C}_{\boldsymbol{X}}$:

$$\boldsymbol{C}_{\boldsymbol{X}} = \left(\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^T\right)\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\right) = \boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^T. \tag{2.17}$$

Therefore, the SVD of $\boldsymbol{\chi}$ tells us that the columns of $\boldsymbol{V}$ are the rank-ordered principal components of $\boldsymbol{X}$.

For the computation of $\boldsymbol{U}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{V}$, we refer to e.g., the reduction to the bidiagonal form [22, 23]. Exact (machine precision) solutions like this are sometimes called *full* SVD solutions, while solutions from approximate methods like randomized SVD [24], which is faster, are called *truncated* SVD solutions.

2.1.5 Dimensional Reduction

Using SVD as it is shown above provides the principal components, which are the columns of $\boldsymbol{V}$ and are sorted from highest to lowest relevance (variance). Often, one finds that the first $k < m$ principal components correspond to large variances in $\boldsymbol{C}_{\boldsymbol{Y}}$ compared to the rest. So, the most interesting behaviour occur only in the first k dimensions, e.g. in the case of the spring shown in figure 2.1, $k = 1$.

From the point of view of phase-locked systems exhibiting collective behaviour, the system can be regarded as a single "common oscillator". We expect to summarize the dynamics of an arbitrary number of d-dimensional measurements distributed in C channels in a single d-dimensional *eigensignal*, or even a signal in $k < d$ dimensions. An example is EEG, multivariate measurements in which each channel is recording voltage (1-dimensional data), therefore $m = C$ (number of channels) and the collective rhythm can be represented on $k = d = 1$ dimension. To obtain such a signal, we simply take the first column of $\boldsymbol{Y}$.

2.1.6 Algorithmic Description

The PCA algorithm through SVD consists of the following steps:

1. Given the data matrix $\boldsymbol{X}$, subtract off the mean for each column.
2. Construct $\boldsymbol{\chi} = \frac{1}{\sqrt{n-1}}\boldsymbol{X}^T$.
3. Apply SVD to $\boldsymbol{\chi}$ and get $\boldsymbol{U}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{V}$.
4. Project the original data: $\boldsymbol{Y} = \boldsymbol{V}^T\boldsymbol{X}$.
5. Use the desired number of dimensions (columns) of $\boldsymbol{Y}$.

2.1.7 Limits and Extensions of PCA

PCA is a very common and powerful option for dimensional reduction. However, there are two hypotheses that must always be kept in mind when applying PCA: *linearity* and *Gaussian distributed data*. The first constraint is not usually a problem because input data may be conveniently pre-processed using linear or nonlinear transformations, or extending the phase space as in [10] and [9]. Nevertheless, when using non-Gaussian data, the largest variances do not correspond to the meaningful axes, i.e. PCA does not work in this case. Instead, one should use Independent Component Analysis (ICA), which is the most general formulation of linear redundancy reduction [25]. It consists on finding the (not necessarily orthogonal) base on which the joint probability distribution can be factorized:

$$P(\boldsymbol{y}_i, \boldsymbol{y}_j) = P(\boldsymbol{y}_i)P(\boldsymbol{y}_j) \;\; \forall i, j. \tag{2.18}$$

So, $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ are statistically independent. Equivalently, from the point of view of information theory [26], the goal is to find a basis in which the mutual information between different sources is zero, i.e.

$$I(\boldsymbol{y}_i, \boldsymbol{y}_j) \equiv \log \frac{P(\boldsymbol{y}_i, \boldsymbol{y}_j)}{P(\boldsymbol{y}_i)P(\boldsymbol{y}_j)} = 0 \;\; \forall i \neq j. \tag{2.19}$$

We illustrate in figure 2.2 a 2-dimensional example with non-Gaussian data, in which PCA fails. In this case the axes with the largest variance do not correspond to the underlying basis, denoted by a_1, a_2 and successfully inferred using ICA. The only drawback of ICA is that it is a nonlinear optimization technique, making the solution more difficult to calculate because simple, direct convex optimization may not work. Therefore, whenever dealing with Gaussian distributed data, it is convenient to use PCA, which is much more efficient and equivalent to ICA in this case[3]. In our case, we will deal with EEG data, which is Gaussian for small enough time intervals if artifacts are not present, and without any further modification, see chapter 4.

Moreover, while PCA is mostly used for data compression neglecting the less important (lower variance) components, ICA splits the channel activity into independent components corresponding to the simplest possible (independent) physiological components in the case of EEG, e.g. ocular artifacts or electrode noise.

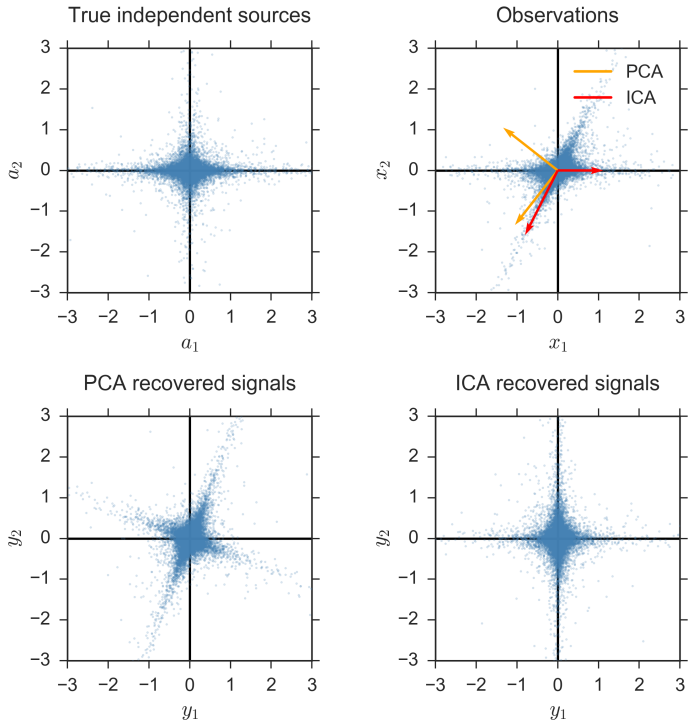

Fig. 2.2: Example of synthetic, non-Gaussian distributed data taken from the *scikit-learn* library examples [27]. PCA fails and ICA finds the base in which the new variables are statistically independent.

[3]For multivariate Gaussian distributions, zero covariance or correlation between different sources implies independence.

2.2 Phaser Algorithm

The Phaser algorithm was proposed by Revzen and Guckenheimer [10] and applies to the estimation of the global phase from multidimensional data produced by a locked system of coupled oscillators. This phase estimation algorithm uses similar mathematical concepts to a previous work carried out by Kraleman *et al* [11], however the targets are completely different.

For our purpose, the only drawback is that this algorithm is useful for generating a global phase, but not an amplitude envelope. However, it is possible to use the amplitude envelope given by PCA (section 2.1) and combine this amplitude with the phase extracted using the Phaser algorithm. For the sake of completeness, we will show that it makes little sense to use the Phaser algorithm since the phase produced by a direct PCA is already consistently a better estimation of the global phase in EEG data. A similar but more advanced method will be discussed in section 2.3.

2.2.1 Motivation

The Phaser algorithm was originally applied to construct a global phase from the Hopf oscillator model and cockroach locomotion synthetic and empirical data in the body frame of reference [10]. While this method turned out to be successful for the example of the cockreach, it is less reliable in the case of EEG data. The main difference between the cockroach locomotion and EEG measurements is the measurement noise, which is higher in EEG data and different channels can have very different SNR and estimated number of cycles[4], making the phase inference more difficult.

Although the algorithm can be also pre-trained and used with novel data, we only consider training data to fit a phase estimator. Now we will show the steps required to implement the Phaser algorithm and then give details and explain some steps.

2.2.2 Algorithmic Description

The Phaser algorithm consists of the following steps:

1. (*Metrization*) Transform measurements to independent scores, similar to z-scores[5] [28]. See equation (2.20).
2. (*Protophases*) Compute the individual instantaneous phases for each score time series using Hilbert transform.
3. (*Series correction operator*) Apply Fourier series based correction to the individual phases. This step is more robust to measurement noise in [10] than in [11].
4. (*Combining multiple estimates of phase*) Combine the individual phase estimates into a single global phase using PCA and apply Fourier series based correction again to compensate for remaining systematic errors.

2.2.3 Metrization

The scores $\boldsymbol{z}_j$ are defined such that $||\boldsymbol{z}_j||$ is the Mahalanobis distance [29] of the (Gaussian) covariance matrix, $\boldsymbol{C_X}$, of measurements $\boldsymbol{x}_j - \langle \boldsymbol{x}_j \rangle$:

$$\boldsymbol{z}_j \equiv \boldsymbol{C_X}^{-1/2} \left(\boldsymbol{x}_j - \langle \boldsymbol{x}_j \rangle \right) = \boldsymbol{U}^T \boldsymbol{\Lambda}^{-1/2} \boldsymbol{U} \left(\boldsymbol{x}_j - \langle \boldsymbol{x}_j \rangle \right), \tag{2.20}$$

where $\boldsymbol{C_X} = \boldsymbol{U}^T \boldsymbol{\Lambda} \boldsymbol{U}$ and small bold letters denote time series represented as column vectors. The scores are linearly uncorrelated because its covariance matrix is the identity. Notice that diagonalization of $\boldsymbol{C_X}$ can be done thorugh SVD of the centered original data, and normalization by $\boldsymbol{C_X}^{-1/2}$ transforms the original data into time series with uncorrelated measurement noise and similar variance.

[4] In the case of the cockroach, every 6 steps the insect has moved its 6 legs.

[5] Transformation to z-scores or standard scores refers to a particular normalization of the data: zero mean and variance equal to 1.

2.2.4 Series Correction Operator

The series correction operator, $\mathcal{P}$, applies to a single phase $\theta(t)$ and corrects systematic errors in $\theta(t)$. For this phase variable, the series correction operator has to provide ϕ such that

$$\frac{d}{dt}\phi[\theta(t)] = \omega, \tag{2.21}$$

so that,

$$\frac{d\phi}{d\theta}[\theta(t)] = \omega\frac{dt}{d\theta}(\theta), \tag{2.22}$$

$$\phi(\theta) = \omega\int_{\alpha=0}^{\theta}\frac{dt}{d\theta}(\alpha)d\alpha. \tag{2.23}$$

In Kraleman *et al.* [11], they follow a similar procedure, but equation (2.23) is solved by direct numerical integration, i.e. they substitute the derivative inside the integral by

$$\frac{dt}{d\theta}\left(\frac{\theta_j+\theta_{j+1}}{2}\right) \cong \frac{\Delta t}{\Delta\theta_j} = \frac{\Delta t}{\theta_{j+1}-\theta_j}, \tag{2.24}$$

where we use the notation $\theta_j \equiv \theta(t_j)$ and $\Delta t \equiv t_{j+1}-t_j$. However, Revzen *et al.* [10] argue that in most cases measurement noise is comparable to $\Delta\theta$, and numerical integration of (2.23) may not converge or converges slowly because $\Delta\theta$ may be distributed in a ratio distribution with heavy tails [30]. The alternative proposed in [10] is to estimate and smooth $\frac{d\theta}{dt}(\theta)$ instead of $\frac{dt}{d\theta}(\theta)$. To do this, we define the truncated Fourier series of order N of $f(\theta) \equiv \frac{d\theta}{dt}(\theta)$ as

$$\hat{f}(\theta) \equiv \mathcal{F}_N\left[\frac{d\theta}{dt}(\theta)\right], \tag{2.25}$$

with coefficients estimated sorting all $\langle s, f(s)\rangle$ pairs and integrating using the trapezoidal rule.

The Fourier series is computed again on the inverse relation to obtain the coefficients $C_k \in \mathbb{C}$, such that we obtain the following approximation for $\frac{dt}{d\theta}(\theta)$:

$$\mathcal{F}_N\left[\frac{1}{\hat{f}(\theta)}\right] = \frac{\hat{dt}}{d\theta}(\theta) = \sum_{k=-N}^{N} C_k e^{ik\theta}. \tag{2.26}$$

Now, the measurement noise has been reduced because of the averaging used in the calculation of the Fourier coefficients of $\mathcal{F}_N\left[\frac{d\theta}{dt}(\theta)\right]$. Then, we can substitute equation (2.26) into the integral (2.23), which gives the approximation

$$\hat{\phi}(\theta) = \theta - \frac{1}{C_0}\sum_{k=1}^{N}\frac{2}{k}\mathrm{Im}\left[C_k e^{-ik\alpha}\right]\Big|_{\alpha=0}^{\theta}, \tag{2.27}$$

which is valid if the phase θ does not lose or gain cycles relative to the actual phase ϕ, also, since time series are real, $C_{-k} = C_k^*$ has been used to obtain equation (2.27). However, in general, low enough SNR will avoid that the number of cycles of the individual phases match the cycles of the exact phase. In addition, the Phaser algorithm has been developed to deal with stationary data. The entire process of approximating ϕ from θ is written as $\hat{\phi} = \mathcal{P}[\theta]$.

This method does not work with non-stationary data because the Fourier series approximate periodic and piece-wise continuum functions. The Fourier coefficients should also be averaged over several periods to remove a significant amount of noise, which may not be possible using windowing techniques (see 4.3.1) to deal with non-stationary data.

2.2.5 Combining Multiple Estimates of Phase

Once the individual phases have been estimated using the series correction operator, they are combined into a single, improved global phase, $\hat{\phi} \approx \phi$, of the phase-locked system with actual phase ϕ. The combination has

the purpose of improving the SNR. First, an analytic signal with constant amplitude envelope is reconstructed for each coordinate as $\hat{q}_{2j}(t) + i\hat{q}_{2j+1}(\theta)$, so that

$$\begin{aligned} \hat{q}_{2j}(t) &\equiv \rho_j \sin\left(\hat{\phi}_j(t)\right), \\ \hat{q}_{2j+1}(t) &\equiv \rho_j \cos\left(\hat{\phi}_j(t)\right), \end{aligned} \tag{2.28}$$

where ρ_j has been previously obtained from the time averaged amplitude envelope of the corresponding time series, i.e.

$$\rho_j = \langle |z_j(t) + i\mathcal{H}[z_j(t)]| \rangle, \tag{2.29}$$

where $z_j(t)$ are the z-scores of the original time series. The magnitudes ρ_j are expected to be higher when $\hat{\phi}_j$ are closer to the actual phase ϕ.

Therefore, since $\hat{q}_{2j}(t)$ and $\hat{q}_{2j+1}(t)$ are orthogonal, we fill a data matrix $\hat{\boldsymbol{Q}}$ with the time series $\hat{q}_j(t)$ organized in columns to perform PCA accounting for small phase shifts, which will be further discussed in section 2.3.4. The first two principal components, $\hat{\boldsymbol{v}}_{\tilde{1}}, \hat{\boldsymbol{v}}_{\tilde{2}}$ of $\hat{\boldsymbol{Q}}$ are used to obtain two orthogonal projections, which provide a phase estimation that is also series-corrected with the operator $\mathcal{P}$:

$$\hat{\boldsymbol{\phi}} = \mathcal{P}\left[\arg\left(\hat{\boldsymbol{Q}}\left(\hat{\boldsymbol{v}}_{\tilde{1}} + i\hat{\boldsymbol{v}}_{\tilde{2}}\right)\right)\right], \tag{2.30}$$

where $\hat{\boldsymbol{\phi}}$ contains $\hat{\phi}(t)$ for the recorded times.

2.3 Kosambi-Hilbert Torsion

Schwabedal and Kantz [9] discussed the possible benefits of improved phase inference and proposed a method called Kosambi-Hilbert torsion (KHT), which optimally infer the phase dynamics of a collective rhythm. KHT has the same target than the Phaser and PCA algorithms for $k = 1$ applied to collective rhythms. KHT is a transformation based on methods proposed by Kosambi [31] and Hilbert, hence its name. It maximally amplifies the SNR of an oscillatory signal which is supposed to be common in all channels.

2.3.1 Motivation

Depending on the level of noise present in a channel, it can happen that the oscillation amplitude falls below the noise level during a small period of time. This effect can produce phase fluctuations up to 2π, and corresponds to a reduction of the SNR in this interval. Therefore, maximizing the SNR instead of the raw signal variance, as in PCA, should avoid these phase fluctuations as much as possible in the resulting global signal.

The problem with standard PCA and ICA relies on the fact that these two methods compute optimal linear combinations of the properly normalized raw time series $x_j(t)$. Therefore, small phase lags among time series coming from different channels and low SNR can affect its performance. The problem of phase lags could seem easily solvable since it seems sufficient to choose a reference channel and maximize the cross-correlation channel-wise before applying PCA. Nevertheless, by doing this we are assuming that all frequency contributions are equally phase-lagged, which is in general not true. Instead, KHT addresses the general problem extending the phase space and with a few additional steps and the same compact matrix formulation of SVD/PCA.

In the next subsections we indicate the steps of the KHT algorithm (2.3.2). Moreover, we address the theoretical explanation of how KHT works in three steps: normalization (2.3.3), phase adjustment (2.3.4) and optimal combination of phase-locked channels (2.3.5).

2.3.2 Algorithmic Description

The KHT algorithm consists of the following steps:

1. (*Reference phase*) Choose a reference channel, $\boldsymbol{x}_1$.

2. (*Normalization*) Normalize each channel signal $\boldsymbol{x}_j$ by dividing through $\text{std}[\boldsymbol{x}_j - \boldsymbol{x}_j^f]$, which is the noise standard deviation of $\boldsymbol{x}_j$.

3. (*Extended phase space*) Construct the $n \times m$ extended data matrix[6]

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x_1} & \boldsymbol{x_2} & \mathcal{H}(\boldsymbol{x_2}) & \dots & \boldsymbol{x_C} & \mathcal{H}(\boldsymbol{x_C}) \end{pmatrix},$$

where $\mathcal{H}$ is the Hilbert transform and $m = C$ is the number of channels[7]. Notice that x_1 is not Hilbert-transformed.

4. (*Band-pass filter*) Compute the filtered data matrix $\boldsymbol{X}^f$ at the desired frequency and bandwidth.

5. (*SVD*) Apply SVD to the filtered extended data matrix:

$$\boldsymbol{X}^f = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T.$$

6. (*Collective rhythm estimation*) Apply the rotation $\boldsymbol{V}$ to the unfiltered extended data matrix, $\boldsymbol{X}$, keeping only the first column of the result:

$$\boldsymbol{y} \equiv \boldsymbol{y}_1 = (\boldsymbol{X}\boldsymbol{V})_1 = \boldsymbol{X}\hat{\boldsymbol{v}}_{\tilde{1}},$$

where $\boldsymbol{y}$ can also be interpreted as a time series $y(t)$ for $t = 0, \frac{1}{S}, \frac{2}{S}, \dots, \frac{n}{S}$, being S the sampling rate.

2.3.3 Normalization

Usually, data normalization is carried out by setting zero mean and variance equal to one. This procedure is also known as conversion to z-scores or standarization, which is a particular method of feature scaling widely used in machine learning and data science [32, 33], e.g. PCA and Phaser algorithm. However, in this case the normalization is different because the KHT algorithm aims to maximize the SNR for a certain frequency range instead of the resulting signal variance. We do normalize to zero mean but do not scale the time series with its standard deviations, we use the noise standard deviation.

To show this, let us suppose that all time series are phase-locked. In that case, introducing or not Hilbert transforms in the extended data matrix does not make any effect because $\mathcal{H}[x_j(t)]$ are orthogonal to $x_{i=j}(t)$ and $x_{i\neq j}(t)$. Also, recall that when a collective rhythm, $y(t)$, is estimated using PCA, the algorithm is maximizing $\text{var}[y(t)]$[8], while KHT should maximize the SNR of $y(t)$, SNR_y, for the desired frequency band. According to the definition of SNR (section 1.5),

$$SNR_y = \frac{\text{var}[y^f(t)]}{\text{var}[y(t) - y^f(t)]}. \tag{2.31}$$

Since KHT is applied to the filtered data matrix, SVD would yield the rotation that maximizes $\text{var}[y^f(t)]$ if we were normalizing each channel dividing by $\text{std}[x^f(t)]$. Therefore, if instead the time series are normalized by dividing by $\text{std}[x_j(t) - x_j^f(t)]$, each filtered channel variance becomes $SNR_j \equiv \frac{\text{var}[x_j^f(t)]}{\text{var}[x_j(t) - x_j^f(t)]}$, which is the definition of the SNR of $x_j(t)$. Thus, using this normalization, as it is indicated in the second step of the algorithm (*Normalization*), SVD provides the rotation that maximizes SNR_y since the diagonal elements of the covariance matrix are SNR_j. This refinement introduced by the KHT algorithm is expected to improve the quality of the extracted common rhythm.

[6] We have reversed the convention used in 2.1 for $\boldsymbol{X}$, so that the columns of $\boldsymbol{X}$ are time series expressed here as $n \times 1$ column vectors $\boldsymbol{x}_j$ or $\mathcal{H}(\boldsymbol{x}_j)$.

[7] Note that in general $m \neq C$, but it holds for one-dimensional individual recordings, e.g. EEG and seismogram data.

[8] PCA maximizes $\text{var}[y(t)]$ properly when it is applied to phase-locked time series.

2.3.4 Phase Adjustment

In general, KHT addresses the complete phase-locking problem, i.e. frequency-wise and channel-wise phase-locking with respect $x_1(t)$ (*Reference phase*). Recall the Hilbert transform (1.4.1) and the notion of the analytic signal (1.4.2), which play an important role in performing this task. The analytic signal is a complex function of time, which can be decomposed in its real and imaginary parts or in amplitude envelope and instantaneous phase. A different weighting in the real and imaginary parts allows to modify the instantaneous phase. Therefore, extending the phase space by using the Hilbert transform channel-wise (*Extended phase space*) and filtering (*Band-pass filter*) will allow the principal component to be twisted in this phase space without constraints to optimally correct the possible phase lags for the selected frequency.

As regards to the reference channel, let us explicitly demonstrate that avoiding the term $\mathcal{H}[x_1(t)]$ in the extended phase space locks the phase of $y(t)$ to the phase of $x_1(t)$. According to the last step of the algorithm (*Collective rhythm estimation*), one can also write $y(t)$ as a weighted sum:

$$y(t) = \alpha_1 x_1(t) + \sum_{j=2}^{C} \left(\alpha_j x_j(t) + \beta_j \mathcal{H}[x_j(t)]\right), \tag{2.32}$$

where we have defined the coefficients

$$\alpha_1 \equiv v_{\tilde{1},1}, \quad \beta_1 \equiv 0, \quad \alpha_{j>1} \equiv v_{\tilde{1},2(j-1)}, \quad \beta_{j>1} \equiv v_{\tilde{1},2j-1}. \tag{2.33}$$

In order to introduce the instantaneous phases, we extend the real space of $y(t)$ to the complex plane to obtain the analytic signal, i.e.

$$\begin{aligned} y(t) + i\mathcal{H}[y(t)] =& \alpha_1 x_1(t) + \sum_{j=2}^{C} \left(\alpha_j x_j(t) + \beta_j \mathcal{H}[x_j(t)]\right) \\ &+ i\alpha_1 \mathcal{H}[x_1(t)] + i \sum_{j=2}^{C} \left(\alpha_j \mathcal{H}[x_j(t)] - \beta_j x_j(t)\right), \end{aligned} \tag{2.34}$$

which is equivalent to the same expression in polar coordinates

$$\begin{aligned} A(t)e^{i\phi(t)} &= \alpha_1 A_1(t) e^{i\phi_1(t)} + \sum_{j=2}^{C} \left((\alpha_j - i\beta_j)x_j(t) + (i\alpha_j + \beta_j)\mathcal{H}[x_j(t)]\right) \\ &= \alpha_1 A_1(t) e^{i\phi_1(t)} + \sum_{j=2}^{C} \gamma_j A_j(t) e^{i(\phi_j(t) + \Delta\phi_j)}, \end{aligned} \tag{2.35}$$

where we have defined the parameters γ_j and $\Delta\phi_j$ such that $\gamma_j e^{i\Delta\phi_j} \equiv \alpha_j - i\beta_j$. Finally, returning to the real space, we get

$$y(t) = A(t)\cos\left(\phi(t)\right) = \alpha_1 A_1(t)\cos\left(\phi_1(t)\right) + \sum_{j=2}^{C} \gamma_j A_j(t)\cos\left(\phi_j(t) + \Delta\phi_j\right). \tag{2.36}$$

From expression (2.36), we note that applying SVD to the filtered extended data matrix (step 5 of the algorithm) will provide us the parameters γ_j and $\Delta\phi_j$ that maximize the SNR of $y(t)$ for the desired frequency. Moreover, maximum SNR should occur when the time series are phase-locked. This procedure is similar to the strategy followed in the Phaser algorithm to account for phase lags (equation (2.28)) since the Hilbert transform of $\cos\left(f(t)\right)$ is $\mathcal{H}[\cos\left(f(t)\right)] = \cos\left(f(t) + \pi/2\right) = \sin\left(f(t)\right)$.

In addition, we note that by avoiding the Hilbert transform of the reference channel we do not allow its phase to be modified (i.e. $\Delta\phi_1$ does not appear in equation (2.36)) and ideal optimization adjusts parameters $\Delta\phi_j$ such that $\phi_1(t) = \phi_j(t) + \Delta\phi_j^{opt}$, where $\Delta\phi_j^{opt}$ are the optimal parameters $\Delta\phi_j$. Therefore, KHT locks the phase of the channels $x_{j>1}(t)$ to the phase of the reference channel $x_1(t)$.

2.3.5 Optimal Combination

Assuming all phase lags have been successfully locked to $\phi(t) = \phi_1(t)$, let us adjust the amplitudes separately. Namely, we suppose optimal parameters $\Delta\phi_j = \Delta\phi_j^{opt}$ and define the phase-locked time series

$$\begin{aligned} x_j^{lock}(t) &= A_j(t)\cos\left(\phi_j(t) + \Delta\phi_j^{opt}\right) \\ &= A_j(t)\cos\left(\phi(t)\right) \end{aligned} \tag{2.37}$$

Accordingly, the collective rhythm would be a simple weighted sum

$$y(t) = \sum_{j=1}^{C} \gamma_j x_j^{lock}(t), \tag{2.38}$$

where we have defined $\gamma_1 \equiv \alpha_1$ to simplify notation, and now define SNR_Σ as the SNR of this weighted sum, which yields the upper bound[9]

$$SNR_\Sigma = \max_{f_c, \hat{\boldsymbol{v}}_{\tilde{1}}, \boldsymbol{y}} \left(SNR_y\Big(f_c, \hat{\boldsymbol{v}}_{\tilde{1}}\Big)\right) = \sum_{j=1}^{C} SNR_j(f_c), \tag{2.39}$$

where $SNR_j(f_c)$ are the SNR of the individual channels at the center frequency f_c. Then, one can define the SNR enchancement with respect to the maximum as

$$\Delta SNR \equiv \frac{SNR_y}{SNR_\Sigma}, \tag{2.40}$$

whose maximum value is 1.

We will now test these methods on an idealized data model (chapter 3) and on the EEG data (chapter 4).

[9] Recall we have demonstrated that optimization with respect to $\hat{\boldsymbol{v}}_{\tilde{1}}$ is equivalent to optimization with respect to $\{\gamma_j\}, \{\delta_j\}$.

Chapter 3

Idealized Data Model

In this chapter, we analyze synthetic, noisy oscillatory multivariate data model hiding a collective rhythm. This data model is defined in section 3.1 and used to visualize the PCA and KHT algorithms in section 3.2 for the simplest case: 2-dimensional data. We emphasize the comparison of performance of the KHT and PCA algorithms because the Phaser algorithm is not expected to work properly for non-stationary data, such as EEG data. This will be further discussed in chapter 4.

Finally, the signal-to-noise ratio (SNR) is shown to be proportional to the phase estimation quality, ΦEQ, for the synthetic data (section 3.3). This phase estimation quality is essentially a measure of the goodness of the phase estimation, which is possible to compute only if the actual phase of the "common oscillator" is known.

3.1 Data Model Formulation

An idealized data model suggested in [9] is used to discuss the performance of the different methods introduced in chapter 2. It is constructed such that each (noiseless) signal representation s_j maps the collective phase $\phi(t)$ to a quasisinusoidal signal, where $j = 1, 2, \ldots, C$, being C the number of channels. In general, each channel can produce a d-dimensional signal, however we consider $d = 1$, analogously to EEG data. When measurement noise, $\xi_j(t)$, is included to the ideal signals s_j, we emulate the channel recordings in EEG as

$$x_j(t) = s_j(\phi(t), t) + \xi_j(t), \tag{3.1}$$

where the measurement noise shows fluctuations of total power σ_j^2, and autocorrelations c_j[1], i.e.

$$\langle \xi_i(t)\xi_j(t+\tau)\rangle = \sigma_j^2 \delta_{ij} c_j(\tau), \tag{3.2}$$

which means that correlations of the measurement noise among different channels are not considered, i.e. the source of noise is independent for each channel.

This data model can also account for time dependent amplitude modulations and arbitrary phase lags among different signals s_j. Equations (3.1) and (3.2) are quite general and we consider delta correlated measurement noise, $c_j(\tau) = \delta(\tau)$, and amplitude modulated sinusoidal signals with uniformly distributed random phase lags $\Delta\phi_j$

$$s_j(\phi, t) = A_j(t)\sin(\phi + \Delta\phi_j), \quad A_j(t) \geq 0 \ \forall t, \tag{3.3}$$

where we consider stochastic phase dynamics with a mean angular frequency ω:

$$\dot{\phi} = \omega + \rho\zeta(t); \quad \langle \zeta(t)\zeta(t')\rangle = \delta(t - t'). \tag{3.4}$$

In summary, the model is considering measurement noise added to phase lagged quasisinusoidal signals that have intrinsic stochastic phase dynamics.

[1] The autocorrelation must fulfill $\int_{\mathbb{R}} c_j(\tau)d\tau = 1$.

For practical purposes we will choose

$$A_j(t) = \frac{1}{2}\left(1 + \sin\left(\frac{\omega t}{15} + \Delta\phi_j^A\right)\right), \tag{3.5}$$

which means that the signal amplitude varies more slowly than the collective oscillations. In particular we choose 1/15 of the mean frequency of the collective oscillations, which we fix to $\omega/2\pi = 10$ Hz. Moreover, we choose the default parameters $\sigma_j^2 = 1$ and the noises

$$\xi_j \in (0, 0.4); \quad \Delta\phi_j^A, \Delta\phi_j \in (0, 2\pi). \tag{3.6}$$

An example of the synthetic data, the signal $\sin(\phi(t))$ and the estimation of such signal using PCA and KHT is shown in figure 3.1. The top panel of this figure shows 5 out of 64 channel signals with $\rho = 5$, used to estimate the collective rhythm using PCA and KHT, which are shown in the bottom panel. We recall that the role of ρ, introduced in equation (3.4), is to set the noise intensity of the angular frequency, which can be also interpreted as the intensity of small variations in the phase ϕ. If we look closely to the bottom panel, we can see that the KHT estimation (red) is slightly more similar to the actual collective rhythm (black) than the PCA one (blue). However, this is not easy to see with the naked eye and we will quantify this resemblance numerically later.

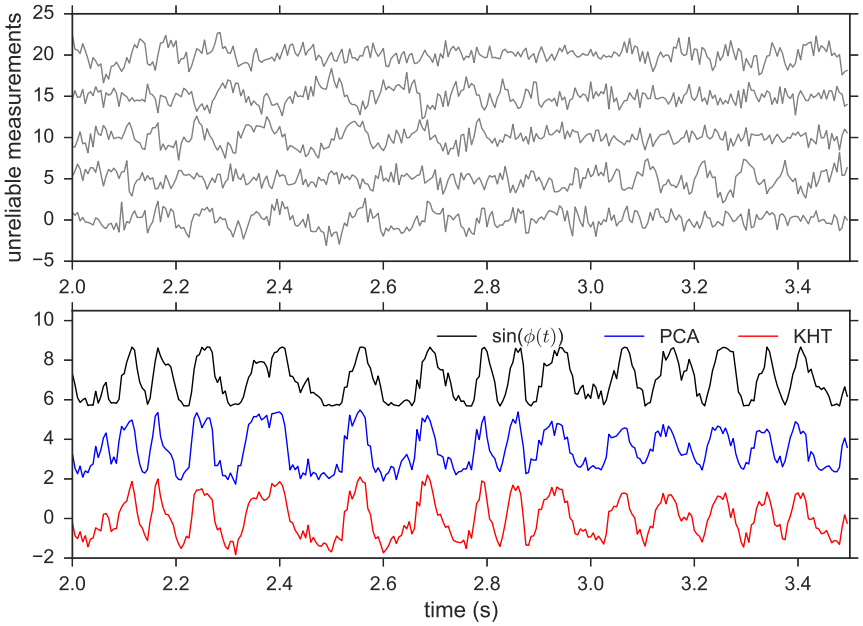

Fig. 3.1: Example of collective signal extraction from unreliable synthetic data. (Top) 5 out of 64 channel signals. (Bottom) Collective oscillation estimation using PCA and KHT compared to the actual signal $\sin(\phi(t))$ using $\sigma_j = 1\ \forall j$, $\rho = 5$, $C = 64$.

3.2 Visualization and Analysis of 2-dimensional Phase Spaces

We will show 2-dimensional examples ($C = 2$) of PCA applied to unfiltered and filtered data of the idealized model, and compare to the results obtained using the KHT. The same intuition developed in this section can be extended to higher dimensions.

On the one hand, since the emulated noisy quasisinusoidal data (see time series in the top panel of figure 3.1) is Gaussian distributed data, one can apply PCA directly. This is expected to work when there is only one dominant frequency range or, if there are more than one, phase lags corresponding to the different dominant frequency bands are similar.

On the other hand, filtering the data setting a narrow frequency range removes almost all the noise contributions. This allows us to specify the frequencies of interest of the raw signals, so that there is always only one dominant frequency range. Moreover, the phase space trajectory is expected to exhibit a limit cycle for a stationary trajectory and, more generally, bounded. Since time series are phase-synchronized, limit cycles will describe elliptical trajectories, for which PCA also works. However, measurement noise typically introduces spurious phase slips, which may affect the performance of PCA applied to filtered data.

3.2.1 Phase Space Topology

Although it is usually said that PCA works perfectly only for Gaussian distributed data, this is not exactly true. PCA works for data distributed according to other distributions, such as ring-like elliptic or uniform elliptic topology, typical from stationary, filtered, phase shifted data, as in the first panel of figure 3.3. The topology of this data is due to phase lags between $x_1(t)$ and $x_2(t)$. Perfectly circular ring-like topology implies a phase lag of $\pi/2$ (or $\pi/2+2k\pi$), while linear topology with small variations due to finite bandwidth implies zero phase lag (or $2k\pi$).

The advantage of applying PCA to filtered data instead of raw data is that this method works for small enough bandwidth even if the raw data is not Gaussian, e.g. electrocardiography data. Nevertheless, if phase lags among time series are present, then PCA may fail. Applying PCA or ICA to $x_1(t) = \cos(\omega t)$ and $x_2(t) = \sin(\omega t)$, x_1 and x_2 would appear as statistically independent components, i.e. their trajectory in the phase space has circular topology because the functions are orthogonal. Besides, we must not forget the problem of the low SNR, if the amplitude of a signal and the noise level σ_j, are of the same order of magnitude small fluctuations can cause spurious phase slips up to 2π.

In addition, we emphasize that the time ordering is not relevant when applying PCA. ICA and PCA also work for time independent data, e.g. there may be a relationship between adult foot and hand sizes, which are also random variables but time ordering does not play any role. Therefore, we do not specify the time ordering in the phase spaces of the figures that will come next in future sections.

3.2.2 Principal Component Analysis ($C = 2$)

In the case of PCA, it is convenient to check that the data is Gaussian, so that PCA and ICA are equivalent. Otherwise it would be better to use ICA, as in figure 2.2. We illustrate that the synthetic data is Gaussian in figure 3.2. In this figure, the left panel shows the normalized raw data represented in its phase space. Also, middle and right panels show the normalized data represented in the basis provided by PCA and ICA, which are indicated in the left panel. For Gaussian distributed data principal and independent components are proportional as it is shown in this figure. We note that in this case we have sufficient statistics and greater differences between PCA and ICA components may arise using less data points.

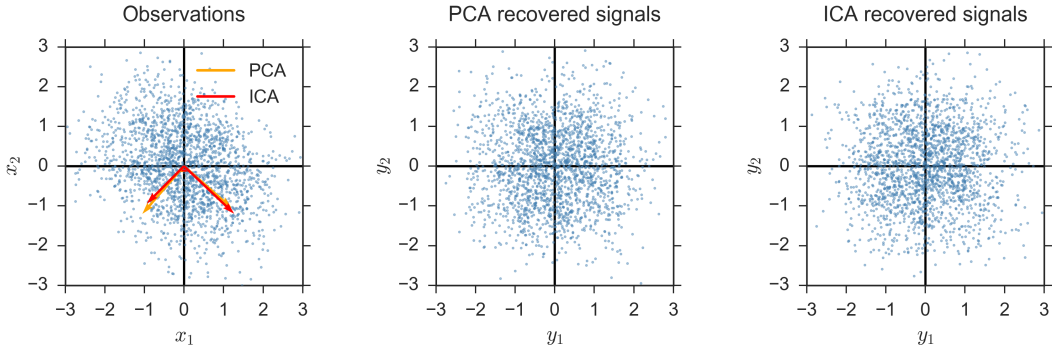

Fig. 3.2: Example of 2-dimensional phase space in which 16 seconds of the synthetic data are represented (blue points). Also, principal and independent components are shown in orange and red, respectively. Note that x_1, x_2 are normalized by its standard deviation. Parameters: $\sigma_j = 1\ \forall j$, $\rho = 5$, $C = 2$ and the sampling rate is $S = 128$ samples per second.

As regards to the filtered (noiseless) data, principal and independent components are also expected to match even if raw data is not Gaussian. We illustrate this in figure 3.3. In this figure, the left panel shows

the normalized filtered data. Also, middle and right panels show the normalized data represented in the basis provided by PCA and ICA, which are indicated in the left panel. The 2-dimensional filtered data shows an elliptic ring-like topology and noise is almost negligible due to the narrow band filtering between 9.5 and 10.5 Hz around the mean frequency of the collective oscillation. As expected, principal and independent components are proportional.

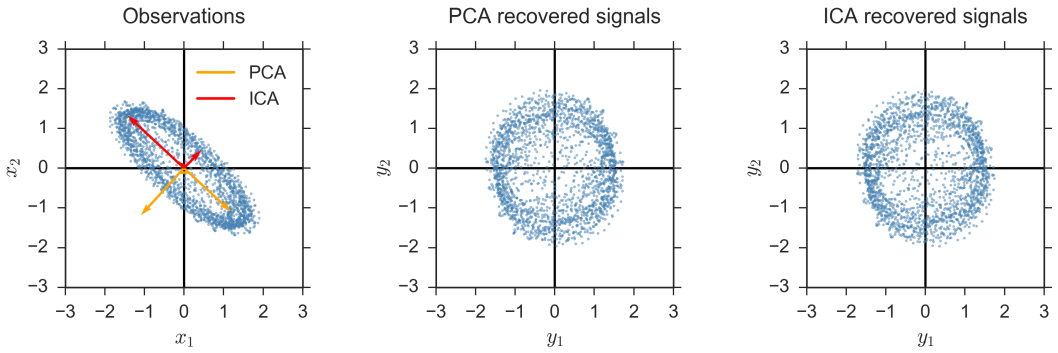

Fig. 3.3: Example of 2-dimensional phase space in which 16 seconds of the filtered synthetic data are represented (blue points), with center frequency of 10 Hz and bandwidth of 1 Hz. Also, principal and independent components are shown in orange and red, respectively. Note that x_1, x_2 are normalized by its standard deviation. Parameters: $\sigma_j = 1\ \forall j$, $\rho = 5$, $C = 2$ and the sampling rate is $S = 128$ samples per second.

3.2.3 Kosambi-Hilbert Torsion ($C = 2$)

In the first panel of figure 3.3, the principal component points in the highest variance direction, which coincides with the major axis of the ellipse. Thus, the phase lag does not pose a problem in this case. This is probably because we use a sufficient amount of data. However, sometimes there is not so much data available, or data do not remain stationary for long time intervals.

We expect PCA to fail when the highest variance direction does not coincide with the major axis of the ellipse, which is more likely to happen for short time intervals due to spurious phase slips and stochastic phase dynamics. This is illustrated for two time series in figure 3.4, although it may occur in an arbitrary number of dimensions. This figure shows filtered data around 10 Hz (blue points) and the same data with an additional phase shift of ~ 0.6 rad in x_2 (black points), so that channels are phase-locked as much as possible. In addition, the highest variance principal components are shown for PCA computed from raw data (blue arrow), PCA computed from filtered data (orange arrow) and a projection of the KHT component in the plane x_1–x_2 (red arrow). In other words, in figure 3.4, the projection of the KHT component is obtained from the actual component shown in figure 3.5.

We note that this procedure is equivalent to phase shift $x_2(t)$ so that it is locked to $x_1(t)$ as much as possible and compute the highest variance principal component using $x_1(t)$ and the shifted version of $x_2(t)$. Exactly this is what is intended to illustrate in figure 3.4. In this figure, the KHT component is a projection in the plane x_1–x_2 and coincides with the highest variance principal component that we would get if we phase shift x_2 by ~ 0.6 rad and left x_1 unchanged (black points).

Obtaining a collective rhythm from this projection would require raw time series to be equally phase shifted than the filtered ones, so that the collective rhythm would be obtained by projecting the phase shifted unfiltered data onto the first principal component obtained from filtered, phase shifted data. Therefore, applying the KHT directly, the result is equivalent to first optimally phase shift the filtered and raw data, then apply SVD to the shifted filtered data and finally project the shifted unfiltered data onto the first principal component.

Actually, if one applies SVD/PCA to filtered data to obtain a collective rhythm, most likely, the first intuition to improve the results is to manually phase shift time series to a reference channel and then apply SVD to get the highest variance principal component. We note here that the KHT algorithm automatizes and optimizes this process by adding the Hilbert transforms and also properly normalizing all time series by its out-of-band standard deviation, so that the optimization objective becomes the SNR of the collective rhythm.

Let us recapitulate illustrating each method for $C = 2$. To obtain a collective rhythm, in each case it is done differently:

- **PCA**: the first principal component computed from the raw data (data in figure 3.2) points in the direction indicated by the blue arrow in figure 3.4. To obtain the collective rhythm, the raw data is projected onto this direction.
- **Filtered PCA**: the first principal component computed from the filtered data (data in figure 3.3) points in the direction indicated by the orange arrow in figure 3.4. To obtain the collective rhythm, the raw data is projected onto this direction.
- **KHT**: using this algorithm the phase space is extended to $2C - 1$ dimensions, as it is shown in figure 3.5 for $C = 2$. There, the $(2C-1)$-dimensional data is shown and the KHT component is indicated (red arrow). Figure 3.5 represents a 3-dimensional phase space in which the axis are the normalized by the out-of-band standard deviation time series $x_1^f(t)$, $x_2^f(t)$ and $\mathcal{H}[x_2^f(t)]$, as it is indicated in section 2.3. To obtain the collective rhythm, the unfiltered extended phase space trajectory is projected onto the KHT component, which is the first principal component of the filtered extended phase space trajectory.

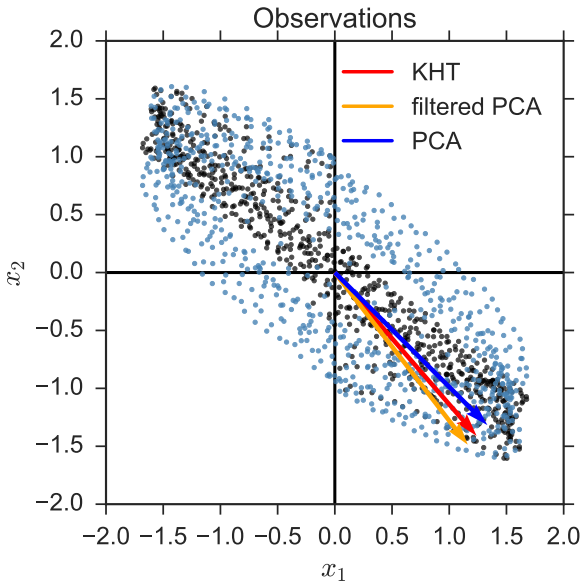

Fig. 3.4: Example of 2-dimensional phase space in which 6 seconds of the filtered and normalized synthetic data are represented (blue points), with center frequency of 10 Hz and bandwidth of 1 Hz. This data is normalized by the out-of-band standard deviation channel-wise. Black points represent the same data phase shifted ~ 0.6 rad and the projection of the KHT component is shown in red, which is also shown in the extended phase space in figure 3.5. Also, principal components computed from filtered and unfiltered data are shown in orange and blue, respectively. Note that x_1, x_2 are normalized by its standard deviation. Parameters: $\sigma_j = 1\ \forall j$, $\rho = 5$, $C = 2$ and the sampling rate is $S = 128$ samples per second.

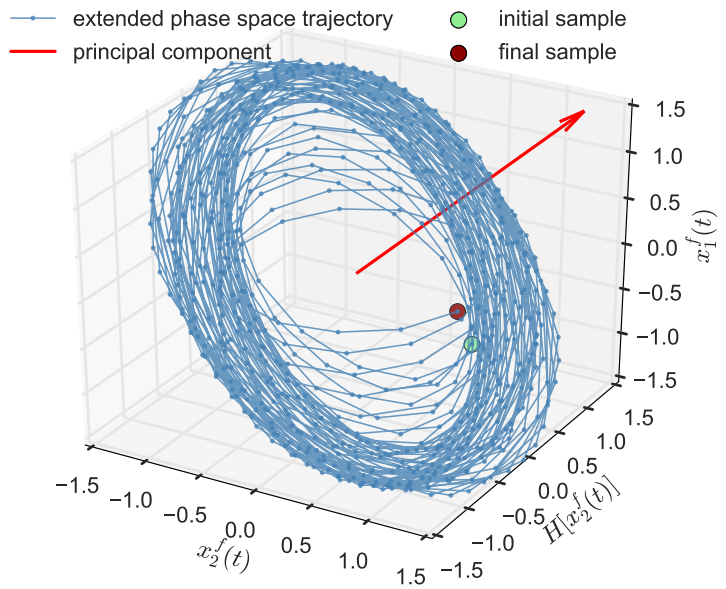

Fig. 3.5: Example of 3-dimensional extended phase space in which 6 seconds of the filtered synthetic data are represented (blue points), with center frequency of 10 Hz and bandwidth of 1 Hz. Also, the KHT component is shown in red. Note that x_1, x_2 are normalized by its standard deviation. Parameters: $\sigma_j = 1\ \forall j$, $\rho = 5$, $C = 2$ and the sampling rate is $S = 128$ samples per second.

Here, we do not intend to justify which method is better for the extraction of collective rhythms. We simply highlight the differences of the three algorithms and illustrate in figure 3.4 that the results may be different. However, we advance that the KHT yields better signal quality in most cases, which makes sense when it is compared to filtered PCA because we avoid spurious phase slips. Comparing the results for the PCA on filtered and unfiltered data, we get better results using PCA applied to unfiltered data most of the times. Thus, filtered PCA is not included in comparisons between PCA and the KHT.

3.3 Phase Estimation Quality

Here, as in [9], we define the phase estimation quality as the inverse of the variance of the difference between the original and estimated signals,

$$\Phi\text{EQ} = \frac{1}{\text{var}\big[\sin\left(\phi(t)\right) - y(t)\big]}, \tag{3.7}$$

This quantity evaluates the goodness of the phase estimations computed from PCA and KHT methods. The utilized signals are similar to the black and blue/red signals for PCA and KHT in figure 3.1, respectively. Note that the quantity ΦEQ is higher when the fluctuations of the difference between the original and estimated signals are small. We show numerically that ΦEQ is proportional to the SNR as it is defined in section 1.5 (filtered signal variance divided by out-of-band signal variance) when the number of channels is high enough (see figure 3.6). The results for a low number of channels are actually not shown in figure 3 from reference [9] and it is relevant for the improvement of low cost EEG headsets, which usually have few channels. However, even when the number of channels is low, better SNR is equivalent to a better phase estimation on average.

In figure 3.6, it is shown how random normal deviations of amplitudes $\rho = 0, 1, 2, 3, 4, 5$ affect the estimated phase performance using both PCA (dashed lines) and KHT (solid lines). Comparing the mean SNR (left panel) and ΦEQ (right panel), we can observe that both quantities are proportional for high enough number of channels. Also, we note that for $\rho = 0$ the relation between both quantities and the number of channels is a power law, while for $\rho > 0$ the relationship is sublinear in logarithmic scale. In [9], they propose as a likely reason that "the Hilbert transformed channels with dynamic noise cannot span a proper embedding space". Actually, the same behaviour is observed for PCA, in which the Hilbert transform does not play any role. We attribute the sublinear behaviour to the fact that high enough deviations from the actual phase

produce phase lags among time series greater than 2π, so that amplitude dynamics are much less correlated in PCA. In the case of KHT, we can only correct phase lags up to 2π, so that amplitude dynamics are also not correlated for high enough deviations from the actual phase.

In the right panel of figure 3.6, it is worth mentioning the similar quality of the phase up to $\rho = 3$. If the intrinsic stochastic phase dynamics have large deviations from the mean angular frequency, then the quality of the estimated signals falls dramatically. Moreover, notice that the quality of the signal estimated by the KHT is above the one estimated by the PCA in most scenarios.

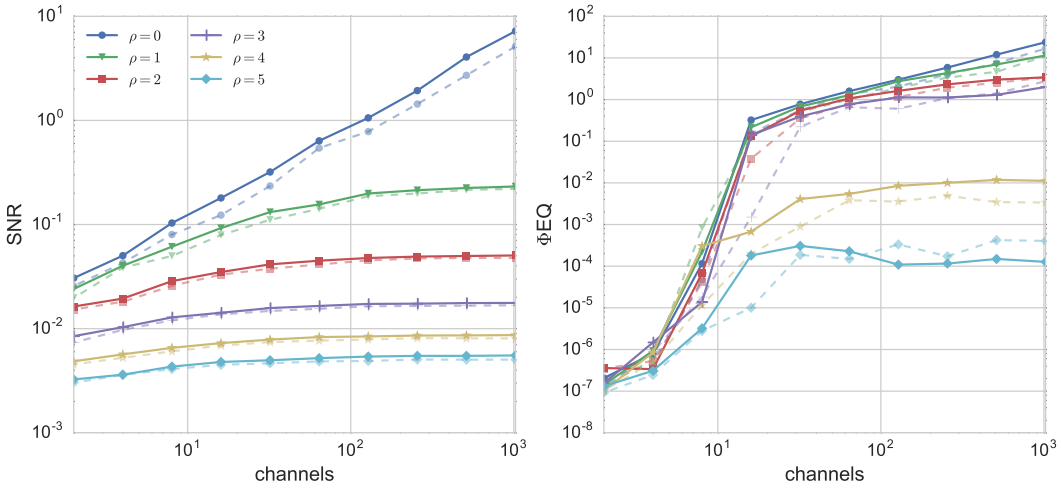

Fig. 3.6: SNR and phase estimation quality as functions of the number of channels and the deviation, ρ, from the mean angular frequency. Continuous and dashed lines correspond to KHT and PCA estimations of the underlying dynamics, respectively.

Therefore, the signals with better SNR are much more likely to provide a better phase estimation. Since in real data the original signal is always unknown, we will evaluate the phase estimation quality by computing the SNRs when dealing with EEG data in chapter 4.

Chapter 4

Collective Signals in EEG data

The data analyzed in this chapter was recorded by us. In particular, we utilized an Emotiv EPOC® headset as the EEG recording device. This commercial ("low-cost") headset has been shown to produce reliable measurements if special care is taken [34–36], while in other cases the low SNR makes the results not as good as those achieved by medical-grade devices [37, 38].

Since the SNR of consumer-grade EEG devices is significantly lower than that of EEG devices used for clinical or research purposes, improving the SNR of collective oscillations could make some consumer headsets match up to the medical ones for certain tasks involving this kind of analysis. We first describe the specifics of the EEG device and two different experimental settings. In particular, we recorded several examples of brain activity with the eyes closed and observing a screen flickering at 15 Hz.

Here, the methods introduced in chapter 2 are used to extract a collective signal from EEG data. We emphasize the comparison of performance of KHT and PCA because we will illustrate in section 4.3 that the phase estimation using the Phaser algorithm is not suitable for the analysis of EEG signals.

The extracted collective signal will be analyzed in terms of instantaneous phase and SNR, which is numerically shown to be proportional to the phase estimation quality for an idealized data model (section 3.3). To deal with EEG experimental data it is also necessary to address non-stationarity (section 4.3.1), which is something we have not considered so far. We show that the data is non-stationary and analyze it using a simple windowing technique together with the methods already introduced to compute the desired collective signal (PCA, Phaser and KHT).

4.1 Data Collection

A 14 channel wireless Emotiv EPOC® headset (figure 4.1) has been utilized to generate the data we analyze in the present work. In this device saline based wet sensors are used to register the signal of each channel. The raw data is collected at 128 samples per second simultaneously for each channel and sent to the computer in real time via wireless transmission. Each electrode has a resolution of 0.51 μV and a bandwidth of 43 Hz.

We have recorded our own data using the Emotiv device described above. Before placing the headset in the head, the electrodes are slightly wetted with a saline solution that improves skin contact (higher conductivity). Here, we have performed two types of measurements: brain activity with closed eyes and observing a screen flickering at 15 Hz.

In the first task, the subject closes his/her eyes and brain activity is measured during 30 seconds. A similar process is repeated for the flickering task, in which the subject looks at a flashing screen with alternating colors (black and white) at a 15 Hz frequency.

While using this device, we noticed that it was not trivial to get acceptable signal in all 14 electrodes. After collecting the data we checked whether the signal of each channel was valid for further analysis. In our case channels T7 and T8 have been discarded because the voltage signal is the lowest and the SNR is almost white noise. The corresponding positions and signal quality of all channels can be seen in figure 4.2. The raw data of the remaining channels is shown in figure 4.3, in which we can see that most EEG channels are well synchronized. In the next section (4.2) we will filter the data to show the target common oscillation more clearly and show the SNR spectrum of the raw data.

Fig. 4.1: 14 channel Emotiv EPOC® headset. Sampling rate: 128 samples/s. Electrode resolution: 0.51 μV. Bandwidth: 43 Hz.

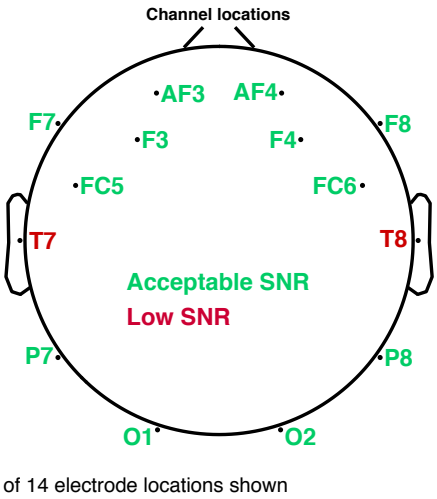

Fig. 4.2: Channel locations and labels of the Emotiv device electrodes. The signals corresponding to red labels have been removed due to their low signal. Here, we consider 12 channels (green labels) for the data analysis.

4.2 Raw Data and Phase-Locking

In this section, we show that the data from both experiments, eyes closed and flicker at 15 Hz, exhibit a collective rhythm. To do this, we rely on the SNR and filtered time series of the individual channels. Although we show 3 seconds of raw data in figure 4.3 for clarity, we actually analyze 30 seconds for each experiment.

Moreover, we have recorded several realizations for each experiment in order to average the SNR over 4 realizations for the eyes closed experiment and 6 realizations for the flicker at 15 Hz to achieve similar relative errors for both experiments.

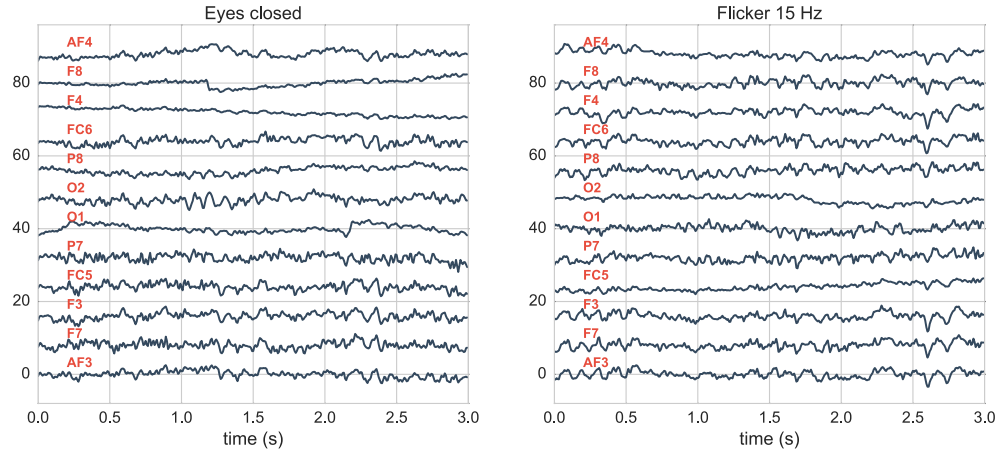

Fig. 4.3: 3 seconds of normalized EEG time series for two different experiments measured using the Emotiv device shown in figure 4.1. (Left) Eyes closed. (Right) Flicker 15 Hz. (Red) Channel locations shown in figure 4.2.

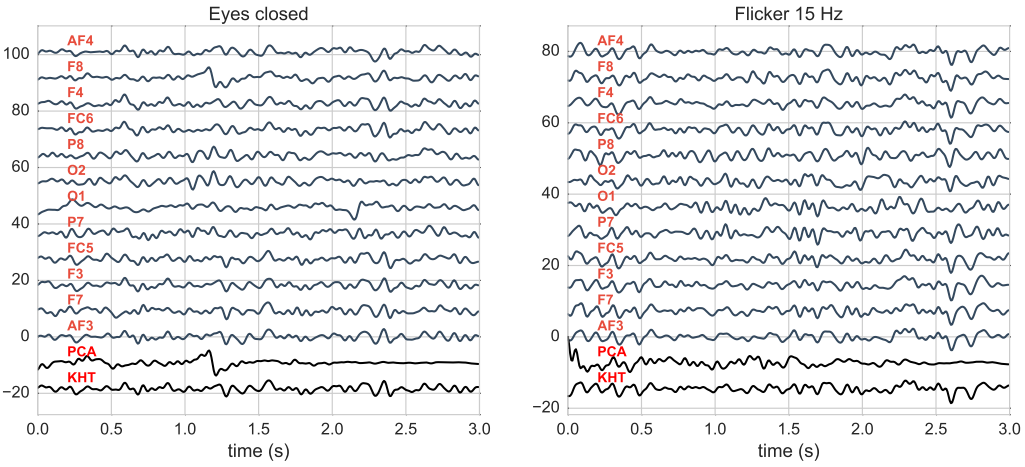

Fig. 4.4: Examples of synchronization in EEG filtered channels, PCA and KHT for 3 seconds of one realization of each experiment: eyes closed and flicker at 15 Hz. The filtered data has been obtained using a sharp band-pass filter between 1 and 20 Hz, and the resulting data has been normalized using standard normalization to plot the time series together. (Left) PCA and KHT, which have been computed for 8.5 Hz expected frequency and 2 Hz bandwidth. (Right) PCA and KHT, which have been computed for 15 Hz expected frequency and 2 Hz bandwidth.

Figures 4.5 and 4.6 show the average SNR of each channel for both experiments in logarithmic and linear scales. In the logarithmic plot we can better visualize the SNR of the channels within the same experiment, while in the linear representation we can also compare the SNR between both experiments. In the case of the eyes closed, one can observe common high activity between 6 and 11 Hz approximately. Such activity is greater in the occipital lobes and especially high in the O2 channel. These lobes are fundamentally dedicated to the vision. In the case of the flicker, we can observe a common rhythm oscillating at 15 Hz in almost all channels in addition to the activity between 6 and 11 Hz. The activity due to the flicker is especially high in the O1 channel (left occipital lobe) and especially low in the O2 channel (right occipital lobe). This asymmetry between left and right lobes observed in both experiments is called lateralization of brain function[1]. In the experiments with eyes closed and flicker, we will use channels O2 and O1 as a reference for the KHT respectively.

[1]The lateralization of brain function is the tendency of some neural processes to be more dominant in one hemisphere than the other.

However, common frequency bands does not necessarily imply synchronization. Although one can already observe the same sort of phase syncronization from raw data shown in figure 4.3, it becomes much more clear by looking at filtered signals. We show an example of filtered signals in figure 4.4, discarding high frequencies, which are not relevant in our case. Also, note that PCA and KHT filtered signals have been included at the bottom of figure 4.4, but will be discussed in the next section (data analysis). In this figure, the data is filtered using a sharp bandpass filter between 1 and 20 Hz. Note that neglecting higher frequencies is equivalent to smooth the data with respect to figure 4.3 for better comparison. Neglecting frequencies below 1 Hz allows to counter long lasting oscillations, which can make synchronization difficult to see with a naked eye in a 3 seconds range.

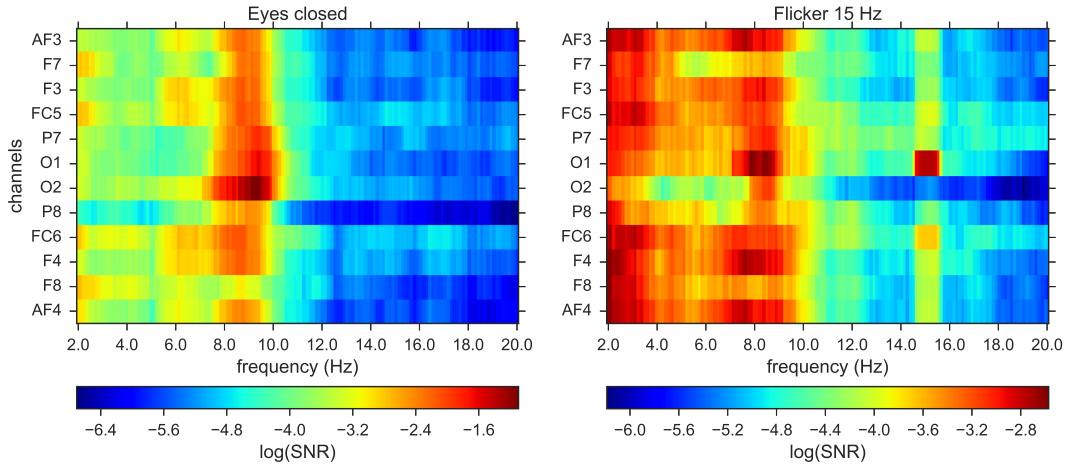

Fig. 4.5: Logarithm of the mean SNR of the data corresponding to two different experiments. (Left) Average over 3 realizations with eyes closed. (Right) Average over 6 realizations looking at a 15 Hz flicker light. Note the different color-codes for the two experiments.

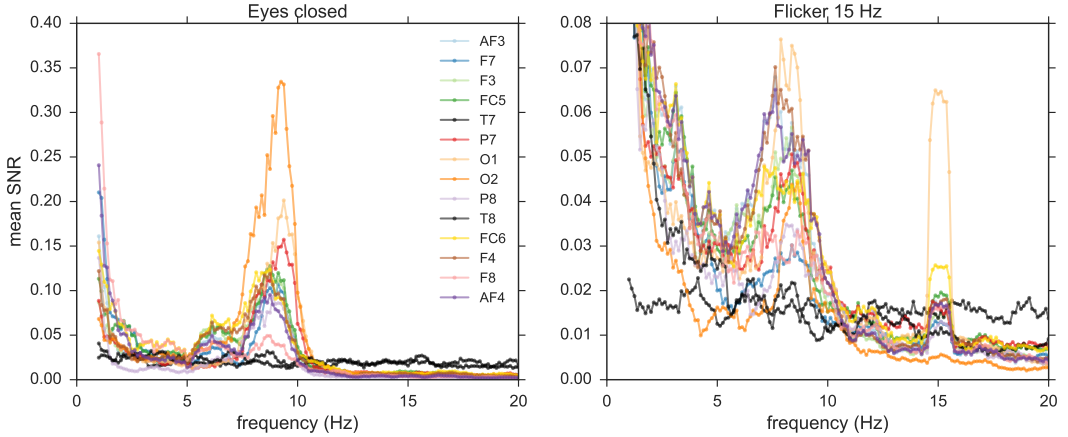

Fig. 4.6: Mean SNR of the data corresponding to two different experiments. (Left) Average over 3 realizations with eyes closed. (Right) Average over 6 realizations looking at a 15 Hz flicker light.(Black) The black lines correspond to discarded channels T7 and T8, see figure 4.2.

4.3 Data Analysis

The data analysis performed in this section relies on concepts already introduced in previous chapters, such as SNR and instantaneous phase. We aim to compare the performance of KHT and PCA methods to extract a collective rhythm. There is, however, a point that we have not covered yet, which is the fact that EEG

data is non-stationary. Thus, we first introduce a simple windowing technique to deal with non-stationary data and then estimate the phase and the SNR of the underlying collective rhythm.

4.3.1 Addressing Non-stationarity

Variations of the signal and noise amplitudes, artifacts, or even brief disconnections are not features of periodic or quasi-periodic data. These drawbacks make the mean and variance to be different at two different temporal windows, which affects the performance of SVD applied to either KHT or PCA. Typically, EEG data is non-stationary and the SNR changes in time. As an example we use one data set with the eyes closed, computing the SNR for 10 seconds windows, with temporal shifts of 0.25 seconds within consecutive windows. The result is shown in figure 4.7. In this figure we show the computed mean SNR of the alpha band (from 7.5 to 12.5 Hz) which actually becomes higher as time passes. We have observed that the SNR can also oscillate, or decrease after starting the task.

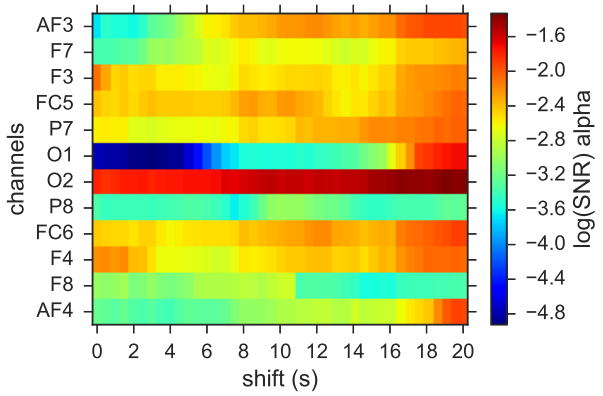

Fig. 4.7: Mean SNR logarithm of the alpha band. The SNR is computed from eyes closed EEG data set in 10 seconds windows, which are shifted in steps of 0.25 seconds for each channel.

A simple solution to deal with non-stationary data is to compute the collective rhythm on small enough temporal windows, with an overlap of a half of the window length. The length of these segments must be chosen so that all segments are stationary. Successive segments are concatenated such that there is a smooth transition between different segments. Given two segments, we define their concatenation as a weighted average over time at the overlap interval that gives more importance to the closest segment at each time. Let T_w be the window length, so that averages take place in regions of length $T_w/2$.

The simplest case is appending a segment $s(t)$, nonzero in the range $[T_0 - T_{overlap}, T_1]$, to a signal $y_{old}(t)$, nonzero in the range $[t_0, T_0]$. The goal is to obtain a new signal, $y_{new}(t)$, nonzero in the range $[t_0, T_1]$. For these times it must be fulfilled that

$$t_0 + T_{overlap} \;<\; T_0 \;<\; T_1. \tag{4.1}$$

One way to implement the concatenation is

$$\begin{aligned} y_{new}(t) \equiv\; & y_{old}(t)^\frown s(t) \\ \equiv\; & y_{old}(t)\Big(\Theta(t-t_0) - \Theta(t-\tau)\Big) \\ & + \Big(y_{old}(t-\tau)W(t-\tau) + s(t-\tau)(1-W(t-\tau))\Big)\Big(\Theta(t-\tau) - \Theta(t-T_0)\Big) \\ & + s(t)\Big(\Theta(t-T_0) - \Theta(t-T_1)\Big), \end{aligned} \tag{4.2}$$

where $\tau \equiv T_0 - T_{overlap}$ and $y_{old}(t)^\frown s(t)$ denotes the concatenation with an overlap time $T_{overlap}$ between the two time series and

$$W(t) = \frac{1}{2}\left(1 + \cos\left(\pi \frac{t}{T_{overlap}}\right)\right), \tag{4.3}$$

Also, in expression (4.2), Θ is the Heaviside step function

$$\Theta(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{2} & t = 0 \\ 1 & t > 0 \end{cases}. \tag{4.4}$$

We note that the term $\Big(y_{old}(t)W(t) + s(t)(1 - W(t))\Big)$ is a weighted average and illustrate in figure 4.8 the effect of this average at the overlap region, in which the frequency and SNR change smoothly.

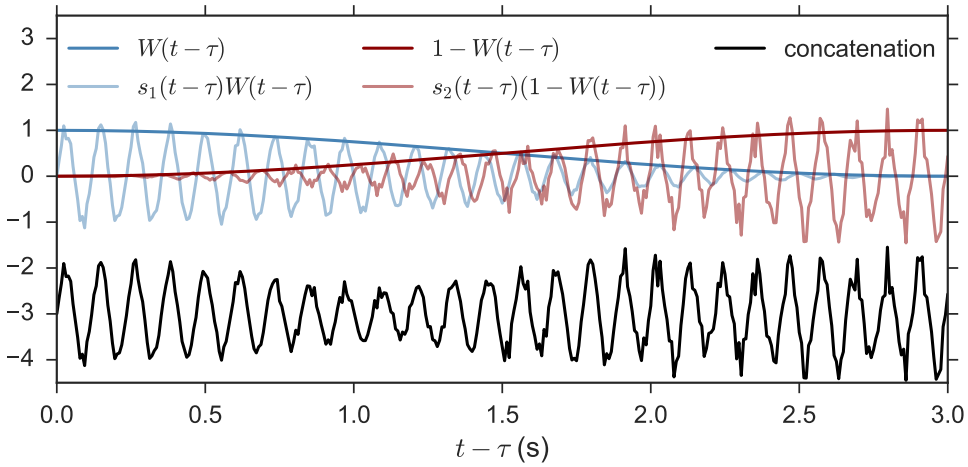

Fig. 4.8: Overlap region, $T_{overlap} = 3$ s, of two segments and their concatenation, which is shifted down for clarity. Blue and red segments have different noise amplitudes and frequencies. The blue segment is an 8.5 Hz sinusoidal signal, and the red segment is a 9 Hz sinusoidal signal with higher noise amplitude and shifted $\pi/8$ with respect the blue one. Also, these segments are weighted by $W(\tau)$ and $1 - W(\tau)$, respectively, so that the sum yields their concatenation.

Let us remember that the purpose of the procedure described above is to append a segment $s(t)$ to a signal $y_{old}(t)$. Since we actually have to append multiple segments of time series with $T_w/2$ overlap, we can do this sequentially by appending segments $s_i(t)$, $i = 1, 2, 3, \ldots, N$, one after the other, i.e.

$$\begin{aligned} y(t) &= s_1(t)^\frown s_2(t)^\frown s_3(t)^\frown \ldots^\frown s_N(t) \\ &= \big(\ldots\big(\big(s_1(t)^\frown s_2(t)\big)^\frown s_3\big)^\frown \ldots\big)^\frown s_N(t). \end{aligned} \tag{4.5}$$

In general, the number of oscillations per window is a free parameter and should be low enough to avoid non-stationary behaviour but high enough to have reasonable statistics when applying SVD. In our case, as in [9], equation (4.2) and 20 oscillations per window were used to perform these concatenations. Therefore, the window length is $T_w = 20/f_c$, where f_c is the center frequency.

4.3.2 Does it make sense to apply PCA to EEG data?

It makes sense to apply PCA to either time dependent and time independent raw data points as long as they are Gaussianly distributed. We observed that this holds approximately in our case since we have to apply PCA on temporal windows containing 20 oscillations, which may be not contain sufficient statistics. However, in most cases principal and independent components are quite similar.

We show an example of the phase space of 5 seconds eyes closed EEG data from channels P7 and O2 in figure 4.9 to illustrate that EEG data is Gaussian because principal and independent components are almost proportional despite using only a few data points. Thus, it makes sense to apply PCA. The same behaviour holds for ∼20 oscillations of filtered eyes closed EEG data (figure 4.10), with center frequency of 8.5 Hz and bandwidth of 0.5 Hz. However, it can sometimes occur that only the components with the largest variance are similar in PCA and ICA due to the lack of data when analyzing non-stationary data part by part or the possible contribution of artifacts, which would contribute to modify mostly components with lower variance

in ICA with respect to PCA. Nevertheless, this does not pose a problem for our purpose of extracting a collective rhythm, since we are only interested in the component with the largest variance.

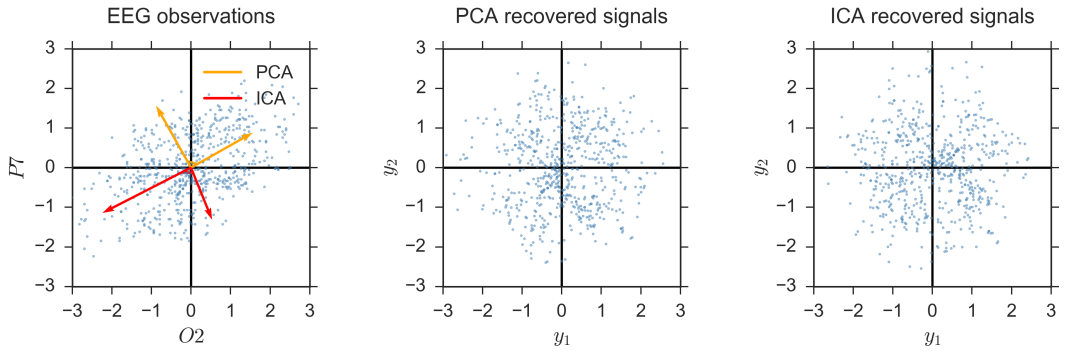

Fig. 4.9: Example of 5 seconds eyes closed EEG data from channels P7 and O2. Principal (orange) and independent (red) components are almost proportional.

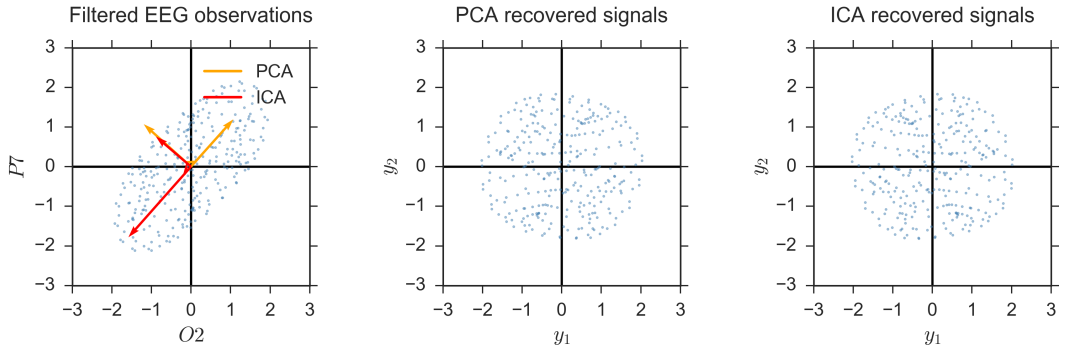

Fig. 4.10: Example of ∼20 oscillations of filtered eyes closed EEG data from channels P7 and O2, with center frequency of 8.5 Hz and bandwidth of 0.5 Hz. Principal (orange) and independent (red) components are almost proportional.

In general, the relationships between channels illustrated in figure 4.9 and 4.10 are similar for other pairs of channels. Thus, high dimensional raw data is also Gaussian, and filtered data can be considered *stationary* for practical purposes, using 20 oscillations per window.

4.3.3 Phase Estimation

This section is devoted to two different purposes. First of all, we will illustrate that the KHT tends to lock the global phase to the reference channel. Secondly, we will show that the Phaser algorithm does not work properly when it is applied to our EEG data.

As regards to the phase-locking when using the KHT, the reference channel is always chosen to be the one with higher SNR. In figure 4.11, we show the individual phases extracted for all EEG channels (gray) and the collective phases extracted from PCA (blue) and KHT (red). The phase of the channel used as a reference for the computation of the KHT is highlighted in black. In general, the KHT is phase-locked to the reference channel if such channel provides a good estimation of the phase. This is the case in the left panel of figure 4.11, in which the phase extracted by the KHT method follows the phase of the reference channel. In the right panel, one of the individual channel phases counts more cycles than both PCA and KHT, which provide a better phase estimation than this particular channel. So we also illustrate that the signal with the maximum number of cycles is not always the one with the best phase estimation, there may be additional spurious phase slips. Moreover, in this case, both the KHT and the PCA estimate the phase similarly to the channel with higher SNR.

After careful evaluation, we have discarded the Phaser algorithm to analyze EEG time series. Actually, our next purpose is to show that, in general, we do not obtain good phase estimations using this method.

Therefore, the results using this method are in most cases omitted for clarity.

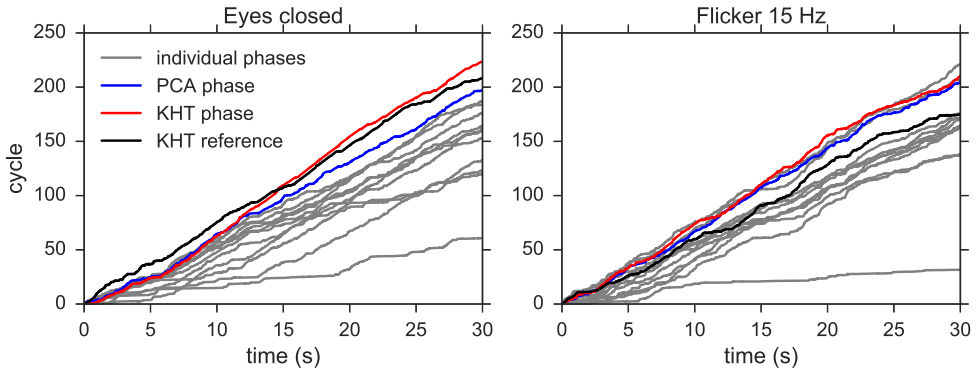

Fig. 4.11: Two examples of unwrapped phases: eyes closed and flicker at 15 Hz for 30 seconds data sets. The figure shows individual, PCA and KHT phases, using frequency bands of 5–11 Hz and 14–15 Hz for eyes closed and the flicker, respectively. The reference phase used for KHT is represented in black. The parameters are: 12 channels, 128 samples per second for each center frequency and 20 oscillations per window.

We observed that in general a simple PCA works better than the Phaser algorithm for this kind of data. The latter only works correctly when the number of cycles of the different channels is very similar, as in figures 4.12 and 4.13 top left panels, for which we choose time intervals manually so that all channels have almost the same number of cycles.

On the one hand, in the left panels of these two figures, we represent the extracted phases (solid lines) and the corresponding linear regression (dashed lines) as a function of time for the estimations obtained from the three methods: PCA (blue), Phaser (green) and KHT (red). The phase estimation has been obtained from 10 seconds data sets in both cases: eyes closed (figure 4.12) and flicker at 15 Hz (figure 4.13). Also, notice that top and bottom panels are different. In the bottom panels, the phase is approximated using 11 channels, while in the top panels, it is approximated using 5 channels (AF3, P7, O1, P8 an AF4) in the case of the eyes closed and only 2 channels (P7 and O2) in the case of the flicker. We have manually picked the number of channels in each scenario such that the differences between the methods are better captured.

On the other hand, in the right panels of figures 4.12 and 4.13, we plot the residual phases as a function of the cycles for each of the three methods, which have been shifted from each other for clarity. The residual phases are computed as the difference between each phase and the corresponding linear fit, i.e. it evaluates deviations from an ideal model with constant angular frequency. The residual phase allows us to compare the total number of cycles and similarities between different residual phases.

Comparing figures 4.12 and 4.13, we note that the KHT is the most consistent method when the number of channels is modified. Recall that the results shown in the bottom panels are for 11 channels, while the results for the top panels are for 5 channels in the case of the eyes closed and only 2 for the flicker. Thus, the KHT is more consistent because the slope of its phase is the most similar between top left and bottom left panels for both experiments. Moreover, the residual phases are also the most similar when comparing top right and bottom right panels for the KHT estimation. Therefore, this method is more robust when adding channels with lower SNR.

However, comparing PCA and Phaser does not seem so straightforward as in the case of the KHT just by looking at figures 4.12 and 4.13. In order to compare the performance of the three methods we will evaluate the corresponding SNRs. We observed that the signal recovered from the phase obtained using the Phaser algorithm does not provide good results for the SNR compared to PCA and KHT in most cases. The only case in which we obtained comparable SNR from the three methods is for the phases shown in the top panel of figure 4.13. This is because the main rhythm present in channels P7 and O1 oscillates at 15 Hz, both channels have high enough SNR and the detected number of cycles is very similar for each channel. Nevertheless, when data has not been selected manually, it would be by chance that these conditions hold. In fact, PCA generally works better than the Phaser algorithm. We show an example adding channel AF4 to the previous case (channels P7 and O1) in figure 4.14. In this figure it is shown that PCA and KHT algorithms reveal a clear signal at 15 Hz, which is not present in the collective rhythm extracted using the Phaser algorithm.

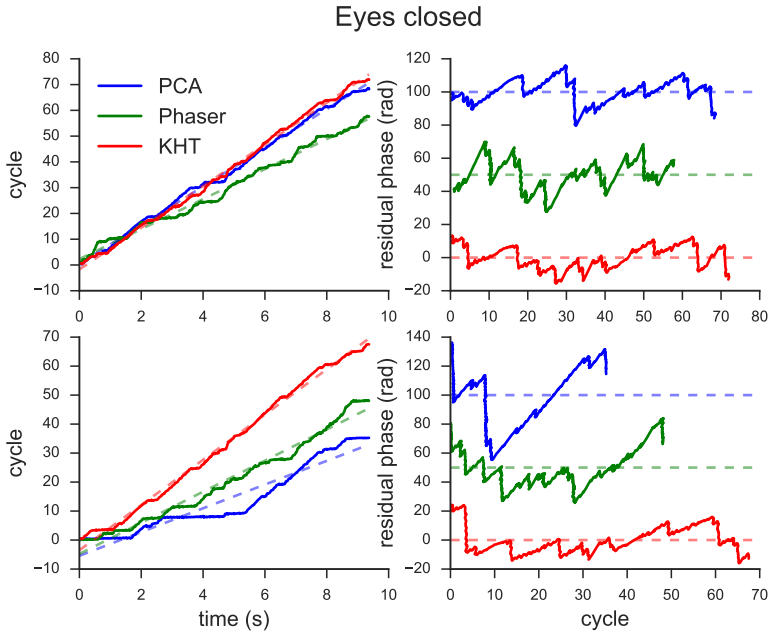

Fig. 4.12: (Left) Instantaneous phases and linear fits. (Right) Residual phases as functions of the corresponding cycles. (Bottom) 11 channels have been used. (Top) Only 5 channels have been used: AF3, P7, O1, P8 and AF4.

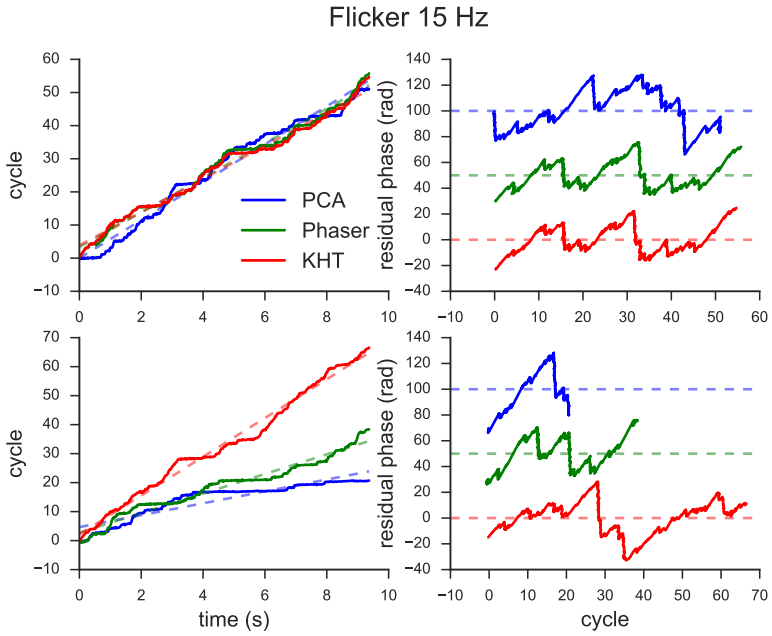

Fig. 4.13: (Left) Instantaneous phases and linear fits. (Right) Residual phases as functions of the corresponding cycles. (Bottom) 11 channels have been used. (Top) Only 2 channels have been used: P7 and O1.

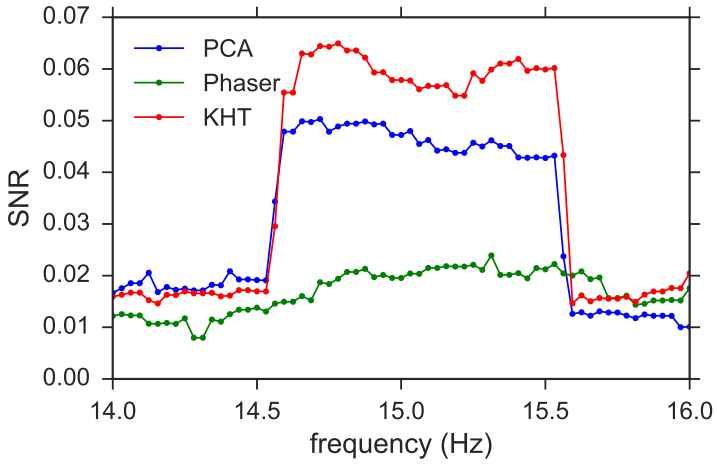

Fig. 4.14: SNR of a 10 seconds flickering collective signal extracted from channels O1, P7 and AF4. The SNR is computed for 1 Hz bandwidth and three different methods have been used for the extraction of the collective signal: PCA (blue), Phaser (green) and KHT (red).

4.3.4 Signal-to-Noise Ratio

Now we will use the SNR as a metric to compare the phase estimations from the PCA and KHT methods. Recall that we showed with numerically generated data that an improvement in the SNR corresponds to an improvement in the phase estimation quality (see chapter 3). Since the actual phase is here unknown, we rely on the SNR to estimate the quality of the different methods.

In figure 4.15 we show the SNR and its enhancement for eyes closed and flickering experiments. One can observe a general enhancement using KHT (red) in contrast to PCA (blue) with 12 channels, and specifically at the peaks, as it is the case of the eyes closed around 9 Hz, and the case of the flickering around 15 Hz. Recall that the SNR enhancement, denoted by ΔSNR, has been previously defined in section 2.3 and it is given by equation (2.40). The SNR enhancement is bounded between 0 ($SNR = 0$) and 1 (theoretical limit). It is shown in the inset axis of figure 4.15 for both eyes closed and flickering experiments. The ΔSNR reveals maximum enhancements at the peaks. In the case of the eyes closed the peak around 9 Hz is enhanced by 37% using PCA and up to 50% using the KHT with respect to the theoretical limit. While in the case of the flicker at 15 Hz, this frequency is enhanced by up to 28% using PCA and up to 39% using the KHT with respect to the theoretical limit.

In figure 4.15, we also note that KHT enhances other less relevant frequency bands which are not enhanced or are even lost using PCA. An example is the activity from 5 to 7 Hz for the eyes closed experiment. The raw data shows a small peak at this frequency band, and it is present in all channels. However, the SNR of the PCA estimation drops clearly below the mean SNR computed from the raw data (black), i.e. the activity is lost in this frequency band. In contrast, the SNR of the KHT estimation is enhanced, i.e. the activity from 5 to 7 Hz is present in the KHT estimation of the collective rhythm.

Interestingly, the projection to obtain the collective rhythm in the case of PCA does not correspond to the component with the largest variance. The highest variance projection does not necessarily correspond to the *eigensignal* with highest SNR. In fact, we used the second largest variance projection to plot the blue lines in figure 4.15, which turns out to correspond to the *eigensignal* with highest SNR. Actually, using the highest variance projection in PCA we obtained a SNR curve similar to that of the mean over channels.

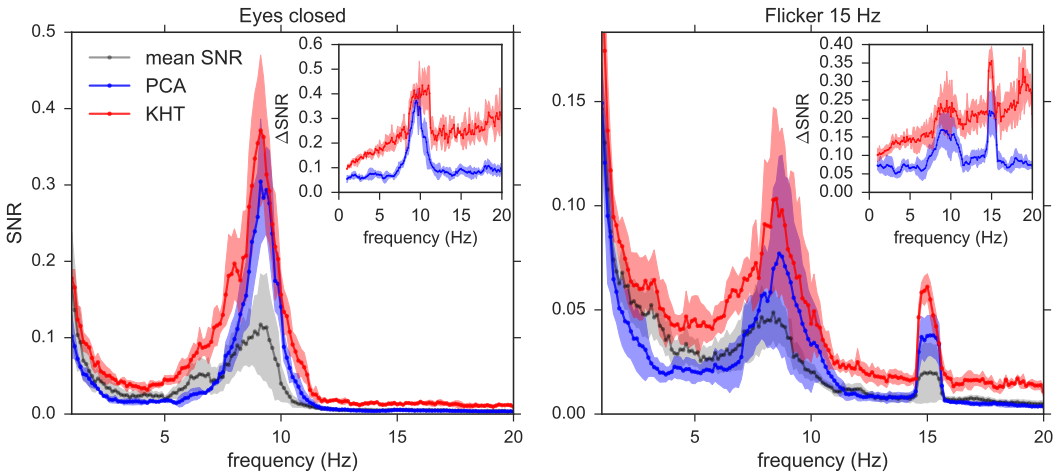

Fig. 4.15: Mean SNR and SNR enhancement (see equation (2.40)), and their standard deviations of the PCA (blue) and KHT (red) collective rhythms. Also, the mean SNR over channels and experiments and its standard deviation is shown in black/grey. PCA results correspond to the highest SNR *eigensignal*, which in this case does not coincide with the highest variance projection. The results are shown for both experiments: eyes closed (4 realizations) and flicker at 15 Hz (6 realizations), and each realization lasts 30 seconds. The parameters are: 12 channels, 128 samples per second, 1 Hz bandwidth for each center frequency (KHT) and 20 oscillations per window.

Figure 4.15 shows the results for 12 channels, which are all the available ones. These results depend on the number of channels used in the analysis. Thus, we show in figure 4.16 the SNR at the peaks for different number of channels, added in decreasing order of SNR. These peaks are at 9 Hz in the case of eyes closed and 15 Hz for the flickering. As expected after the analysis of the synthetic data (figure 3.6), KHT provides in general a better phase estimation than PCA, while the order of magnitude of the obtained SNR is the same for both methods. The main difference with the synthetic data is that the SNR is not the same for all time series. Therefore, adding very noisy channels may decrease the SNR of the extracted collective rhythm. In this regard, it seems that KHT is more robust to the addition of channels with lower SNR. As shown in figure 4.15, the SNR of the phase extracted using PCA indeed fluctuates a lot when using a different number of channels.

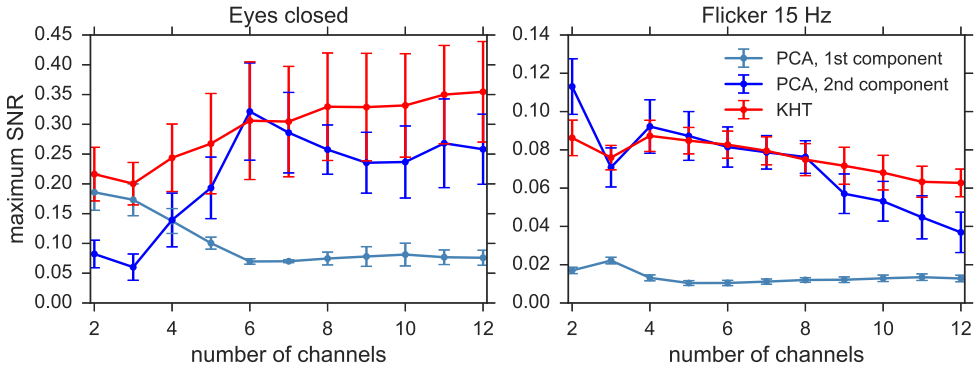

Fig. 4.16: Eyes closed and flickering experiments' SNR averaged over realizations. The parameters are the same as in figure 4.15, changing the number of channels and evaluating the SNR of the estimated collective rhythms only at the peaks. These collective rhythms have been estimated using PCA (using first and second largest variance projections) and KHT methods. The peaks are located at 9 Hz in the case of eyes closed and 15 Hz for the flickering.

Although we have seen an improvement in the SNR of the estimated collective rhythm using KHT compared to PCA, there is not a big difference because experimental data does not seem to contain major phase lags. Therefore, we wonder what happens when phase lags among time series are artificially added, and are not only due to low quality data. In advance, we already know that PCA will not work better in presence

of significant phase lags. But, how does higher phase lags affect the performance of KHT? To answer this question we shift in time all channels using random uniform shifts and analyze the SNR. Figure 4.17 shows the SNR at the peaks of interest for each experiment, varying the number of channels, added in decreasing order of SNR, and for different number of shifted samples. In the case of the eyes closed experiment, the analyzed peak is located at 9 Hz, while in the case of the flickering we analyze the 15 Hz peak. In this figure, we change the maximum number of shifted samples for each realization. Each channel is shifted by a random number of samples within the allowed range $\left[-\max(shift_j), \max(shift_j)\right]$ for $j = 1, 2, \ldots, C$. In figure 4.17, the horizontal axis is $\max(shift_j) = 1, 2, \ldots, 14$, which has the same value for all channels.

One can observe that for both experiments the results are very similar. In the case of PCA, the higher is the number of channels, the lower is the SNR and, the higher is the temporal shift, the lower is the SNR. In contrast, when using the KHT, the higher is the number of channels, the higher is the SNR and, the higher is the temporal shift, the higher is the SNR. In addition, we note that the KHT estimation reaches a maximum SNR as we increase the maximum temporal shift.

One could expect the SNR of the KHT to remain approximately constant when the maximum shift is changed, however the SNR increases. We have checked that the variance of the out-of-band signal decreases because the amplitudes lose correlation for high enough time shift, i.e. the time shift produces a phase shift and a change in amplitude. Thus, the KHT will align the phases by adding the optimal phase lags up to 2π for the frequency range of interest, while the out-of-band signal is not necessarily aligned. Using this procedure, the phase approximation is better than that without temporal shifts.

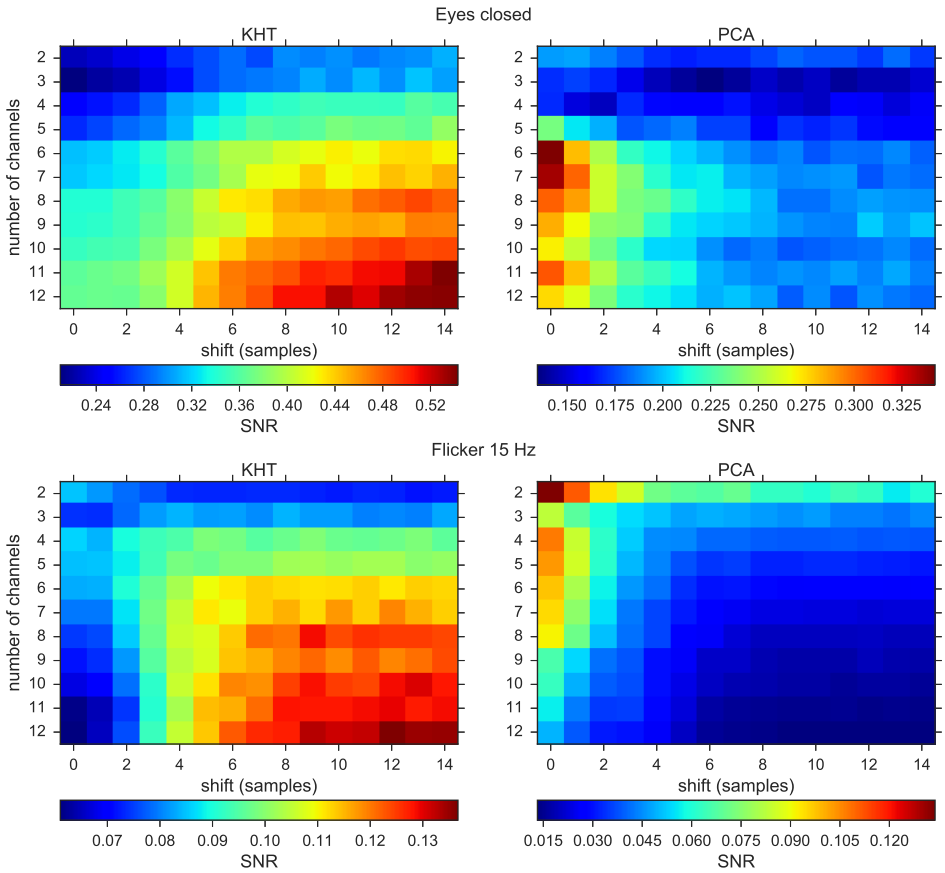

Fig. 4.17: Eyes closed (top) and flickering (bottom) mean SNR at the corresponding peaks at 9 and 15 Hz computed from KHT (left) and PCA (right) collective rhythm estimations. The horizontal axis indicates the maximum time shift. This time shift is random uniform among time series and we obtain the SNR averaged over 30 random realizations and the corresponding experiment realizations for each cell. The vertical axis indicates the number of channels added in decreasing order of SNR used for the computation of both quantities from the collective KHT collective rhythm estimation.

Additionally, we illustrate the differences between the time shifted KHT estimation and other signals in figure 4.18, which are shifted from each other for clarity. In this figure we use a 30 seconds eyes closed data set

and only 7 seconds are shown. The top time series is the reference signal (grey) chosen for the computation of the KHT estimation, computed from the channel with highest SNR. From top to bottom, the second and third time series are PCA estimations using the projection onto the highest variance principal component from filtered (light green) and unfiltered (dark green) data. The fourth time series is the PCA estimation using the projection onto the second principal component computed from raw data, which turns out to be the choice with highest SNR when using PCA. The fifth and sixth time series are KHT estimations centered at 9 Hz with 1 Hz of bandwidth; the fifth one is the KHT computed from raw data (red) and the sixth one is the KHT computed from time shifted raw data (dark red). Both PCA and KHT estimations have been computed by parts of 20 oscillations corresponding to 9 Hz and the SNR corresponding to each signal, computed at the frequency band 8.5–9.5 Hz.

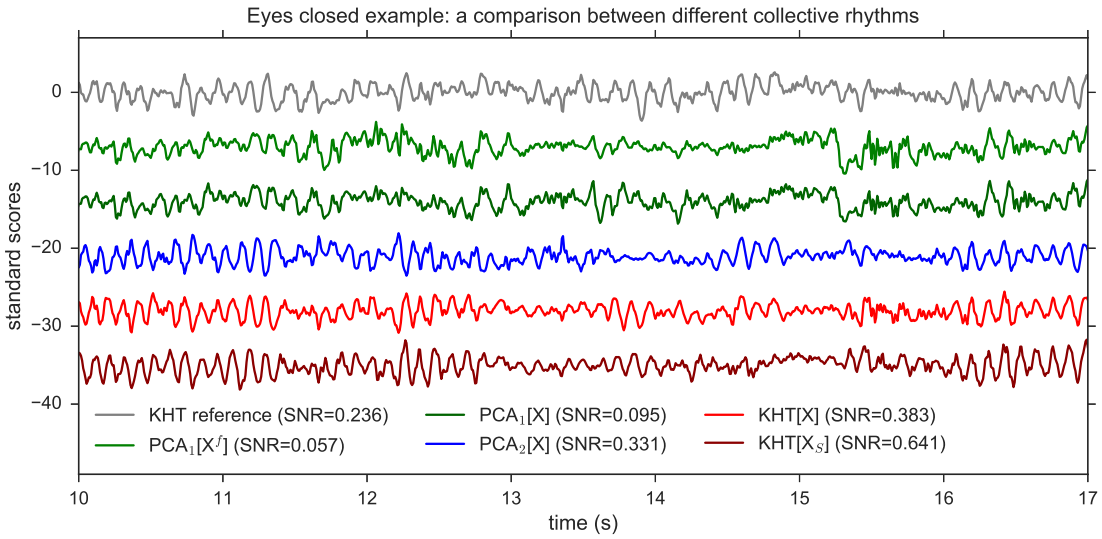

Fig. 4.18: Several examples of estimations of collective rhythms from a data set with the eyes closed is shown. For each signal, the corresponding SNR for the frequency band 8.5–9.5 Hz is indicated in the legend. The PCA signals are computed by parts of 20 oscillations at 9 Hz and KHT signals are also computed at this frequency band (8.5–9.5 Hz). (Grey) Reference channel for the KHT estimations, this channel is the one with highest SNR. (Green) PCA estimation using the highest variance principal component and filtering the data. (Dark green) PCA estimation using the highest variance principal component computed from the raw data. (Blue) PCA estimation using the second principal component computed from the raw data. (Red) KHT estimation. (Dark red) Example of a KHT estimation computed from time shifted data. We used 12 channels for the computation of all the collective rhythms, except for the top time series, which is raw data.

In figure 4.18, we note that the best PCA estimation of the collective rhythm already has a good SNR ($SNR = 0.331$), the original KHT yields a better estimation ($SNR = 0.383$) and the KHT computed from time shifted data yields an even better estimation ($SNR = 0.641$). The SNR of the better estimation is 2.7 times higher than the SNR of the reference signal and almost 2 times higher than the SNR of the best PCA estimation.

4.4 Discussion

Applying the KHT to our experimental data we observed a general improvement over PCA results for the phase estimation. Although this improvement may not seem very significant, the KHT algorithm also provides information about the phase lags among time series if the channel measurements are well calibrated in time. These phase lags can be easily computed from the SVD coefficients (section 2.3.4) and present oscillatory behaviour correlated among channels.

The KHT yields reliable results independently of the number of channels used in the analysis. In contrast, the optimum number of channels for the correct estimation of the global phase using PCA has to be manually picked. This makes the KHT method more appealing for an unsupervised use.

Finally, we propose two uses of the KHT method to estimate the collective rhythm from multivariate oscillatory signals and also get reliable information about phase lags as follows. First, use the original KHT

to estimate the evolution of the phase lags at every 20 oscillations. This will yield reliable estimations if channels are well calibrated in time, so that phase shifts will correspond to the actual phase shifts among different regions of the scalp. Second, use the KHT estimation from artificially time shifted data to get a better collective rhythm estimation.

Chapter 5

Conclusion

Throughout this work, some concepts about statistics and linear algebra have been useful to introduce and understand SVD/PCA, which is the core of the other two algorithms described in chapter 2: Phaser and KHT algorithms. These last two are extensions of PCA for the extractions of collective rhythms and apply to time dependent multivariate data.

Although applying simple SVD/PCA may not be optimal for the extraction of collective rhythms from multi-channel oscillatory data, this algorithm is still useful for our purpose. Although PCA is less powerful than ICA, it is optimal to be used for the topology observed in filtered, time dependent data. In turn, the KHT algorithm computes the optimum torsion to project the signal onto the highest SNR component of the extended phase space trajectory of the filtered signals. This procedure, however, does not improve much the results over standard PCA in our case, so we also add artificial phase lags among time series before using the KHT. Using this trick some improvement has been observed with respect to PCA results as regards the SNR, which we showed numerically that is proportional to the phase estimation quality using an idealized data model (chapter 3).

In order to compare the results provided by the different methods, it is important to choose tasks or experiments that really test their performance in a common framework. In our case, the experiments have been chosen to test the methods on fundamentally different signals. On the one hand, the brain activity with the eyes closed presents a globally distributed oscillation around 9 Hz and the measured signal is significantly higher than with the eyes opened. On the other hand, the brain activity of the flickering produces an EEG collective rhythm at 15 Hz within a narrow frequency range, i.e. the flicker produces a localized oscillation. We tested the methods on 2 different scenarios: high SNR with a broad spectrum and low SNR with a narrow spectrum. For the first case (eyes closed), when the channels with lower SNR are added to the analysis, the results still keep improving when using the KHT, while using PCA, we found that it is optimum to use only the 6 channels with higher SNR. For the second case (flicker at 15 Hz), when channels with lower SNR are added to the analysis, the results become slightly worse or constant using the KHT, while using PCA, we found that using only 2 channels is optimum and adding more channels with lower SNR makes the SNR of the collective rhythm to start decreasing significantly.

The content shown in this dissertation is in part the introduction of basic contents and the reproduction and explanation of previous works. In addition, we highlight our contributions to the topic, which are:

1. Direct numerical comparison between SNR and phase estimation quality (chapter 3), particularly for a low number of channels.

2. We have recorded our own EEG experimental results in order to characterize the signal quality of a commercial "low-cost" headset. These experimental measurements are explained, shown and analyzed in section 1.2 and chapter 4.

3. Since for our experimental data the KHT provides an improvement in the SNR and thus in the phase estimation quality with respect to PCA, we expect similar improvements in the SNR of other "low-cost" EEG devices.

4. We showed that applying the KHT to channels randomly shifted in time can enhance the phase estimation of the collective oscillation.

As further work in this line, it would be interesting to test the outcome of the KHT method in machine learning applications together with low-cost EEG devices. For instance, the SNR enchancement provided by the KHT method would more likely improve the performance of pattern recognition methods. In addition, using high-precision headsets, we could get very good estimations to the actual phase lags and its time evolution in different brain regions and compare the connectivity in terms of correlation and coherence (phase lags).

Bibliography

[1] R. Hegger, H. Kantz, and T. Schreiber, "Practical implementation of nonlinear time series methods: The TISEAN package," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 9, no. 2, pp. 413–435, 1999.

[2] J. T. Schwabedal, "Parametric phase diffusion analysis of irregular oscillations," *EPL (Europhysics Letters)*, vol. 107, no. 6, p. 68001, 2014.

[3] P. Andersen, S. Andersson, and T. Lømo, "Nature of thalamo-cortical relations during spontaneous barbiturate spindle activity," *The Journal of Physiology*, vol. 192, no. 2, pp. 283–307, 1967.

[4] R. Cooper, A. Winter, H. Crow, and W. G. Walter, "Comparison of subcortical, cortical and scalp activity using chronically indwelling electrodes in man," *Electroencephalography and clinical Neurophysiology*, vol. 18, no. 3, pp. 217–228, 1965.

[5] D. Amaral and P. Lavenex, "Hippocampal Neuroanatomy In: The Hippocampus Book: (Anderson, P., Morris, R., Amaral, D., Bliss, T., O" Keefe, J., Eds.)," 2007.

[6] G. Buzsáki, C. A. Anastassiou, and C. Koch, "The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes," *Nature Reviews. Neuroscience*, vol. 13, no. 6, pp. 407–420, 2012.

[7] S. Boccaletti, J. Kurths, G. Osipov, D. Valladares, and C. Zhou, "The synchronization of chaotic systems," *Physics Reports*, vol. 366, no. 1, pp. 1–101, 2002.

[8] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, "Phase synchronization of chaotic oscillators," *Physical Review Letters*, vol. 76, no. 11, p. 1804, 1996.

[9] J. T. Schwabedal and H. Kantz, "Optimal extraction of collective oscillations from unreliable measurements," *Physical Review Letters*, vol. 116, no. 10, p. 104101, 2016.

[10] S. Revzen and J. M. Guckenheimer, "Estimating the phase of synchronized oscillators," *Physical Review E*, vol. 78, no. 5, p. 051907, 2008.

[11] B. Kralemann, L. Cimponeriu, M. Rosenblum, A. Pikovsky, and R. Mrowka, "Phase dynamics of coupled oscillators reconstructed from data," *Physical Review E*, vol. 77, no. 6, p. 066205, 2008.

[12] D. Fee and R. S. Matoza, "An overview of volcano infrasound: From Hawaiian to Plinian, local to global," *Journal of Volcanology and Geothermal Research*, vol. 249, pp. 123–139, 2013.

[13] J. O. Smith, *Mathematics of the discrete Fourier transform (DFT): with audio applications*. Julius Smith, 2007.

[14] R. P. Kanwal, *Linear integral equations*. Springer Science & Business Media, 2013.

[15] P. F. Panter, *Modulation, Noise and Spectral Analysis: applied to information transmission*. McGraw-Hill, 1965.

[16] M. J. Smith and R. M. Mersereau, *Digital filtering: A computer laboratory textbook*. John Wiley & Sons, Inc., 1993.

[17] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[18] H. Hotelling, "Analysis of a complex of statistical variables into principal components.," *Journal of Educational Psychology*, vol. 24, no. 6, p. 417, 1933.

[19] P. Sanguansat, *Principal component analysis-multidisciplinary applications*. InTech, 2012.

[20] J. Shlens, "A tutorial on principal component analysis," *arXiv preprint arXiv:1404.1100*, 2014.

[21] V. Klema and A. Laub, "The singular value decomposition: Its computation and some applications," *IEEE Transactions on Automatic Control*, vol. 25, no. 2, pp. 164–176, 1980.

[22] G. Golub and W. Kahan, "Calculating the singular values and pseudo-inverse of a matrix," *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, vol. 2, no. 2, pp. 205–224, 1965.

[23] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische Mathematik*, vol. 14, no. 5, pp. 403–420, 1970.

[24] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Review*, vol. 53, no. 2, pp. 217–288, 2011.

[25] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, vol. 46. John Wiley & Sons, 2004.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[27] "FastICA on 2D point clouds." http://scikit-learn.org/stable/auto_examples/decomposition/plot_ica_vs_pca.html. Accessed: 2017-07-25.

[28] L. Crocker and J. Algina, *Introduction to classical and modern test theory*. ERIC, 1986.

[29] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The Mahalanobis distance," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, no. 1, pp. 1–18, 2000.

[30] M. D. Springer, *The algebra of random variables*. John Wiley & Sons, Inc., 1979.

[31] D. Kosambi, "Statistics in function space," in *DD Kosambi*, pp. 115–123, Springer, 2016.

[32] J. Grus, *Data science from scratch: First principles with Python*. " O'Reilly Media, Inc.", 2015.

[33] J. Quackenbush, "Microarray data normalization and transformation," *Nature Genetics*, vol. 32, no. Supp, pp. 496–501, 2002.

[34] N. A. Badcock, P. Mousikou, Y. Mahajan, P. de Lissa, J. Thie, and G. McArthur, "Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs," *PeerJ*, vol. 1, p. e38, 2013.

[35] N. A. Badcock, K. A. Preece, B. de Wit, K. Glenn, N. Fieder, J. Thie, and G. McArthur, "Validation of the Emotiv EPOC EEG system for research quality auditory event-related potentials in children," *PeerJ*, vol. 3, p. e907, 2015.

[36] J.-A. Martinez-Leon, J.-M. Cano-Izquierdo, and J. Ibarrola, "Are low cost Brain Computer Interface headsets ready for motor imagery applications?," *Expert Systems with Applications*, vol. 49, pp. 136–144, 2016.

[37] M. Duvinage, T. Castermans, M. Petieau, T. Hoellinger, G. Cheron, and T. Dutoit, "Performance of the Emotiv Epoc headset for P300-based applications," *Biomedical engineering online*, vol. 12, no. 1, p. 56, 2013.

[38] F. Nijboer, B. van de Laar, S. Gerritsen, A. Nijholt, and M. Poel, "Usability of three electroencephalogram headsets for brain–computer interfaces: a within subject comparison," *Interacting with Computers*, vol. 27, no. 5, pp. 500–511, 2015.