



# UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Fisica

Tesi di Laurea

Speech envelope and EEG signal analysis as a tool to  
study language development.

Relatore

Prof. Samir Suweis

Correlatore

Dr. Ramón Guevara

Laureando

Sebastiano Monti

Anno Accademico 2020/2021



## **Abstract**

When humans listen to speech, their neural activity tracks the slow amplitude fluctuations of the speech signal over time, known as the speech envelope. Recently, using electroencephalography, the neural activity of newborns and 6-month-old have been investigated through such a technique. In this thesis, we first study the related theoretical framework of signal analysis. Then we re-analyse both acoustic and EEG data of a group of subjects and investigate correlations between the amplitude and phase of the speech envelope and EEG signals, together with scaling analysis of the latter taken from the theory of critical phenomena.



# Contents

<b>Introduction</b>	<b>iii</b>
<b>1 Neural processes and electroencephalography</b>	<b>1</b>
1.1 Physiological bases of neural signal transmission . . . . .	1
1.2 Neural oscillations . . . . .	2
1.3 Overall aspects on EEG . . . . .	3
1.4 EEG strengths and limitations compared to other techniques . . . . .	6
1.5 Types of feasible analysis from EEG recorded signal . . . . .	7
1.5.1 Event related potentials (ERPs) . . . . .	8
1.5.2 Time-frequency analysis . . . . .	8
1.6 Causality relation between electric fields and cognition . . . . .	9
1.7 Brief insight on critical phenomena in the brain . . . . .	10
<b>2 Experiment</b>	<b>13</b>
2.1 Subjects and data acquisitions . . . . .	13
2.1.1 Newborns . . . . .	13
2.1.2 Six-month-olds . . . . .	14
2.1.3 Stimuli and data acquisition . . . . .	14
2.2 Hilbert transform . . . . .	14
2.3 Methods for amplitude tracking analysis . . . . .	15
2.4 Methods for phase tracking analysis . . . . .	16
2.5 Mean phase coherence . . . . .	18
<b>3 Results</b>	<b>19</b>
3.1 Speech envelope tracking at birth . . . . .	19
3.1.1 Amplitude tracking . . . . .	19
3.1.2 Phase tracking . . . . .	19
3.2 Speech envelope tracking at 6 months . . . . .	20
3.2.1 Amplitude tracking . . . . .	20
3.2.2 Phase tracking . . . . .	21
3.3 Phase synchronization between different channels . . . . .	22
<b>4 Discussion and conclusions</b>	<b>25</b>
<b>Bibliography</b>	<b>29</b>



# Introduction

Speech envelope is defined as slow amplitude fluctuations in function of time of a speech signal, which can be seen as a particular case of an acoustic signal characterized by peaks at the syllabic rate.

Studies suggest that human brain is able to synchronize with some features of the speech envelope while listening to a speech and this phenomenon is known as speech envelope tracking [1, 2, 3, 4]. This mechanism is thought to play a role in processing the speech signal and it is believed that the quality of synchronization is linked to a better comprehension of the speech [1, 2].

Two types of synchronization mainly occur. One is called amplitude synchronization and takes place when the amplitude of the activity coming from the auditory cortex matches with the speech envelope. The other is called phase synchronization, or phase-locking, and happens when the phase of the oscillations in the auditory cortex matches with the phase of the speech envelope.

Despite the apparent connection between quality of envelope tracking and comprehension of speech, some studies seem to address something different. In particular, two of them, found phase synchronization in  $\theta$  band frequency (4-8 Hz) to be independent of comprehension [5, 6]. The importance of these results comes from the fact that theta band frequency overlaps with the frequency band associated with the syllabic rate of a common speech (4-5 Hz). Moreover, other studies found envelope tracking to be present even for time compressed and time reversed speech [7, 8, 9].

These contradictions put into question whether speech envelope tracking and comprehension are really causally connected and what are the roles of amplitude and phase tracking in speech processing. Are these the only mechanisms that contribute on speech processing? Are these somehow related to each other?

Since it is known that speech perception is already present in newborns from prenatal experiences, while comprehension arises only later in the growth of a child, it may be interesting to study speech envelope tracking during development, to clarify if this mechanism is present even in absence of comprehension. Moreover, it would be useful to understand if familiarity with a certain language may affect speech envelope tracking. These are the main purposes of the the study that represent the central object of this work of thesis [10].

Before introducing the various hypothesis that are expected to be the possible results of this study, it is useful to briefly describe the experiment. Two groups of babies have been studied: one formed by an ensemble of newborns, while the other by 6 months old babies. Both the groups have been exposed to three sets of repeated sentences in three different languages (one set per language). The languages that have been chosen for the experiment were: the native language of the babies (French), a rhythmically similar unfamiliar language (Spanish) and a rhythmically different unfamiliar language (English). The listening of the sets of sentences took place compatibly with the attention threshold of the babies and the volume was set to a normal conversational level (65-70 dB). During the listening, the neural activity of the babies has been recorded using electroencephalography, to which from now on we will refer as EEG. Doing some operations on the recorded sentences (auditory stimuli) and on the EEG raw data, it is possible to extract respectively what are called the speech envelopes and the evoked responses. Finally, analyzing those signals, it is possible to assess if and in what measure amplitude and phase tracking have taken place. Therefore, further on in this thesis, will be presented more on specific the basic principles of EEG, the details about the methodology and the materials used to record and analyze the data and some theoretical tools of signal analysis.

We now list some a priori hypotheses on what we can expect as results. The obtained results will be presented later in the thesis. For the group of newborns we can state that, if envelope tracking is associated with comprehension we shouldn't find it at this age. On the contrary, if envelope tracking

consists in a basic auditory ability, newborns may show it and, if so, differences between the languages should be observed, based on prenatal experiences. Behavioural studies shows, in fact, that newborns are capable to recognize their native language [11, 12] and to distinguish rhythmically different but not rhythmically similar languages [13, 14].

On the other hand, studies suggest that 6-month-old babies acquire the capability to distinguish even rhythmically similar languages [15, 16], so, many possibilities can be valid. If the development and the growing experience with speeches increases the ability of envelope tracking, we should find an enhanced synchronization for all the three languages. While, if the native language especially influence the ability of envelope tracking, we should find some differences in synchronization between the unfamiliar languages and the native language. However, many experiments investigated the neural processes at the base of the emerging of linguistic knowledge in babies of several months of life, for their native language. What was found is that, a shift in the attention from the syllabic characteristics of a speech to the phonetic ones occurs [17, 18]. Therefore, the most valued hypothesis for 6-months-olds is that development temporarily attenuate envelope tracking for the native language.

In this experiment's framework it becomes also interesting to investigate whether some evidences of criticality may be present. Critical phenomena describe the behaviour of a dynamics system poised in proximity of a special (critical) point of the system phase space, separating different phases of the system. In such condition, system microscopic fluctuations are not damped, but instead propagate at every scale. A prominent feature that characterize a critical system is the presence of scaling laws: universal functions that describe the shape of apparently erratic fluctuations at every scale of the system. Particularly, for critical phenomena, these laws have the analytic form of power laws:  $y \propto c/x^k$  where  $x$ ,  $y$  are two variables,  $c$  is a constant and  $k$  is called critical exponent. The obedience to power laws reflects a feature of universality, characteristic of critical phenomena. In other words, the fact that different critical systems have in common the same critical exponents may highlights the possibility that the same fundamental physical principle underlie the studied phenomena.

A classical example of system displaying criticality is the Ising model, in which the magnetic spins moments inside a material interact, giving birth to a macroscopic phenomenon of magnetization under certain temperature conditions. This macroscopic phenomenon is called phase transition and consists in a manifestation of a magnetic field in a ferromagnetic material when it is cooled below a critical temperature called Curie temperature. Power law behaviour, in this phase transition of a material from paramagnetic to ferromagnetic (when cooling its temperature near the critical point), is observed in the dimension of coherent spin domains measured over time.

Critical like phenomena can be encountered in a broad variety of different research fields. From birds' flocking [19] to earthquakes [20], to traffic jams [21] and other more phenomena. The main difference between these latter examples and the Ising model, is that they do not require an external parameter (like temperature) to be tuned in order to induce a phase transition. Instead, in these systems, criticality seems to emerge from within the system itself, without any need of tuning external parameters. This phenomenon is called self organized criticality [22, 23] and the mechanisms at its basis are mainly two: one is related to the accumulation and subsequent release of a certain amount of energy, while the other refers to a concept of "memory" of the system. In analogy with earthquakes dynamics: a great amount of energy accumulates over time because of increasing tension between tectonic plates, then, at a certain point, it is violently released in an earthquake. Immediately after this huge release of energy, it is unlikely that other violent earthquakes follow. From recent evidences that will be listed further on, self organized criticality is thought to play an important role also in some neural and cognitive processes.

In this work of thesis, because of how EEG data were recorded in experiment [10], we did not explicitly assess criticality, but only moved the first steps in such direction, basing on another evidence of critical phenomena in a different biological function: the human heartbeat [24].

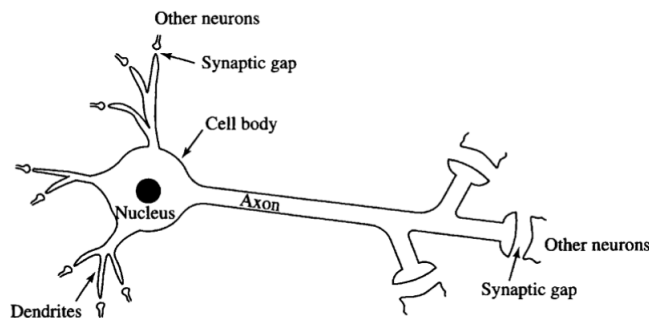


# Section 1

## Neural processes and electroencephalography

### 1.1 Physiological bases of neural signal transmission

Before going through the details of the study [10], object of this thesis, it is useful to first introduce some basic concepts underlying the functioning of electroencephalography. Most of the contents are taken from [25].



**Figure 1.1:** Structure of a typical neuron. Picture taken from [26].

Examining the physiology at the basis of neural processes is helpful to better understand which neural events EEG is able to measure and what are his strengths and limitations. First of all, it is necessary to see the structure of a neuron and explain how the electric signal is generated at its level.

The cell body of the neuron contains the nucleus and the organelles that are necessary for the cell to fulfill all its functions. From the cell body depart expansions named dendrites and a long tubular fiber called axon, which ramifies, at its terminal part, in synaptic boutons. Generally, the axons of the neurons that belong to the central nervous system, are wrapped with a myelin sheath in almost all its length, with the exception of some regions, called nodes of Ranvier. For contrast, at the peripheral level, myelin sheath is absent in most of the sympathetic neural fibers. In figure 1.1 it is possible to see a rough picture of a typical neuron.

There are two types of synapses. One is chemical, while the other is electrical and, in some cases, both can take place at the same time. Both are fundamental for signal transmission, but the most common is the chemical signal transmission. For this reason we only provide to describe this main process.

The signal received from the dendrites of the target postsynaptic neuron is of chemical type and it is mediated by neurotransmitters. These chemical substances diffuse through the synaptic gaps from the presynaptic neuron, activating ion channels on the cell membrane of the postsynaptic target neuron. As the ion channels are activated, a flux of ions enter and exit the membrane and perturb the  $-70\text{ mV}$  membrane potential at rest. To maintain a homeostatic environment, neuronal cells work to keep constant the potential at its rest value at the expense of a great amount of energy in terms of adenosine triphosphate (ATP).

Depending on whether the neurotransmitters act on closing or opening the ion channels and on the ions involved, two different processes may occur. In some cases, the continuous incoming of neurotransmitters, induces the ions' concentration to vary in such a way to cause an exceeding of the threshold value

of  $-55\text{ mV}$  in the membrane potential. A series of successive processes then generate an electric signal, called action potential (or spike), that travels through the axon in direction of the synaptic boutons. In these cases, the shift in membrane potential is called an excitatory postsynaptic potential (EPSP). In particular an EPSP is depolarizing, so, intuitively, it makes the inside of the cell more positive, increasing the probability of shooting an action potential. Sometimes, a single EPSP isn't large enough to bring the membrane potential to its threshold value, but it can sum together with other EPSPs to trigger an action potential. In other cases, the neurotransmitters act in such a way to maintain the membrane potential below its threshold value, avoiding the shooting of an action potential. In these cases, the shift in the membrane potential is called an inhibitory postsynaptic potential (IPSP). IPSPs hyperpolarize the inside environment of the neuronal cell, making it more negative. Their importance consists in the fact that they can counteract or cancel the excitatory effects of EPSPs.

In the light of these competing mechanisms, the electric signal that crosses a neuron is hence given by the sum of all excitatory and inhibitory postsynaptic potentials.

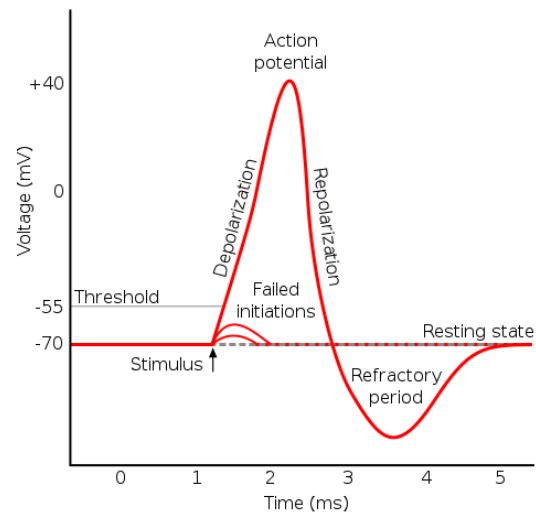
The myelin sheath plays a crucial role in the transmission of this signal along the axon, because allows it to "jump" across neighbours nodes of Ranvier, boosting its speed from  $0,5 - 2\text{ m/s}$  (without myelin sheath) up to about  $20 - 100\text{ m/s}$  (with sheath). This process is called saltatory conduction. Other secondary functions of the myelin sheath are of mechanical protection and nourishment of the axon, but, most importantly, of electrical insulation from outside extraneous signals. The electric signal, is then selectively transmitted to other target neurons through the gaps between the synaptic boutons of the presynaptic cell and the dendrites of the target postsynaptic cells via chemical synapses. At this point, the whole process is repeated equally as described above for all the new target neurons and spread across entire populations of neuronal cells.

Despite the fact that diffusion processes are relatively slow compared to the propagation of electric signals, they are anyway very fast within the domain of transmission of chemical neural signals. This happens because the synaptic gaps are very short and measure approximately  $20\text{ nm}$ . For various reasons that will not be specified in this work, the combination of electrical signals transmission along neurons' axons and chemical processes in the synaptic gaps, represents the most efficient way of transmitting neural signals in humans.

It is important to specify that, while the action potential is a type of impulse that either happens at its full strength or doesn't happen at all, the postsynaptic signal is more gradual. Presynaptic and postsynaptic cells can dynamically change their signaling behaviour based on their internal state or on the number of signals they receive from other cells. Neurons can in fact generate multiple action potentials in sequences, called spike trains, that can form different kinds of patterns, like, for example: rhythmic spiking, bursting and oscillatory activity. In particular, this latter pattern could be observed also in sub-threshold fluctuations of the membrane potential. While for rhythmic spiking and oscillatory activity is intuitive to understand what kind of pattern they describe, bursting is a phenomenon in which rapid sequences of action potential spiking are followed by quiescent periods much longer than the average inter-spike intervals.

## 1.2 Neural oscillations

This section is dedicated to describe the basic mechanisms that produce neural fluctuations on a larger scale with respect to the single neuron. Neural oscillations are generally defined as fluctuations in the



**Figure 1.2:** Plot of a typical action potential. Picture taken from [27].

electric dendritic activity of populations of neurons. Firing patterns coming from single neurons can propagate to entire populations of neuronal cells. If this activity become synchronous between the cells of the population, rhythmic changes in the electric potentials will sum up by constructive interference and result in large amplitude oscillations of the local field potential.

In neuronal ensembles, rarely neurons shoot at the same time with full synchronization and, in some cases, they shoot at random rates. However, the firing probability is rhythmically modulated in such a way that neurons are more likely to shoot at the same time. This gives rise to oscillations in their mean activity. And, particularly, the frequencies of these large scale oscillations, in general, do not necessarily match with the frequencies of the firing patterns of single neurons.

An important mechanism that produces neural oscillations in cortical regions of the brain, fundamental for EEG detection, is based on a shifting balance between excitation and inhibition processes in charge of excitatory pyramidal cells and inhibitory interneurons. When a population of pyramidal cells activate from inputs coming from some brain cortical areas, their excitation increases as they continue exciting each other. At a certain point, interneurons within this population also activate and, as their activity increases, excitatory cells become inhibited. Finally, interneurons' activity decreases, allowing again the excitation of pyramidal cells to increase. Many additional factors contribute then in modulating the frequency, phase and amplitude of the oscillations.

Frequency band's name	Boundary values [ $Hz$ ]
Delta	2 – 4
Theta	4 – 8
Alpha	8 – 12
Beta	15 – 30
Lower gamma	30 – 80
Upper gamma	80 – 150

**Table 1.1:** Summary of the brain rhythm frequency bands.

Oscillatory brain activities associated with cognitive processes occur simultaneously within a broad spectrum of frequencies, spanning from about 2  $Hz$  up to 150  $Hz$ . All these frequencies can be isolated using signal processing techniques and they can be divided into different frequency bands. This distinction between bands is made because different cognitive processes and neural functions seem to occur at different frequency ranges, however the classification is not universally established and can slightly vary depending on the context of study. For example, differences across subjects can be observed in EEG peak frequencies, possibly depending on individual brain structure, age and brain chemistry. A possible classification is described in the table 1.1.

Another crucial fact about EEG that will be deeper analyzed later in the thesis is that measured signals from large scale oscillatory activities show a broad spectral content similar to pink noise. In other terms, EEG displays a frequency scaling close to  $1/f$ . This leads to important paths of research in interpreting some brain processes in terms of critical phenomena [28].

In an even more macroscopic view, neural oscillations can also arise from interactions between different areas of the brain, coupled through the structural connectome. One of the major applications in the study of these phenomena is, for example, the construction an efficient whole brain model.

### 1.3 Overall aspects on EEG

EEG is a high temporal resolution technique, used to study neurocognitive processes by measuring electric potentials recorded from different electrodes with respect to a special electrode taken as a reference. The typical units that are used to express the measured potentials from neural activity are the  $\mu V$ . For various reasons EEG is an optimal method for doing analysis and research to investigate neural processes on humans and animals. In particular, patterns from EEG recordings, seem to reflect neurophysiological models of cognitive functions. This represent an important aspect for our purposes, given the fact that we aim to study processes that occur in the auditory cortex of the babies.

EEG measurements can be made both in vitro or in vivo. In particular, the latter method can be carried out both invasively or not. An interesting application of invasive method is used, for example, to study neural processes on EEG epilepsy patients. This, applied with mean phase coherence analysis,

that will be later examined, would may help to better understand the mechanisms that cause a seizure and eventually predict an incoming seizure hours before it actually takes place [29]. For the particular experiment on which we concentrate, for obvious reasons, a noninvasive method has been chosen.

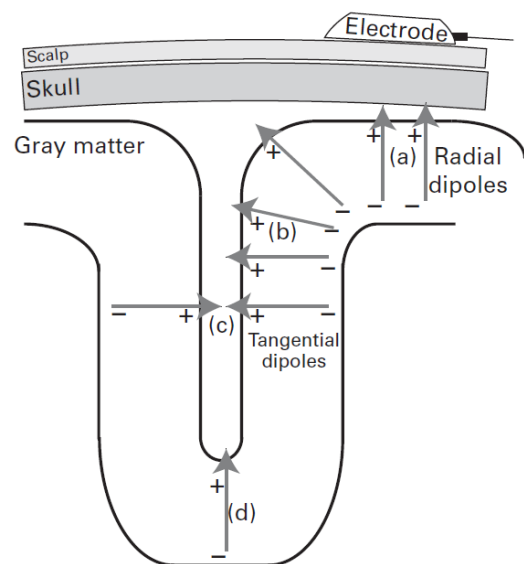
Noninvasive technique consists in measuring the signal coming from cortex's neural activity by putting some electrodes (also referred as channels) on a subjects' scalp. Typically the number of electrodes that are used is of the order of tens (in average, 64 channels are used), but depending on the spatial accuracy that we want to achieve in the measurement, the number can increase up to about 256 electrodes. For costs reasons and other more important issues that we will later specify, it is inconvenient to use a great number of electrodes for EEG measurements.

In paragraph 1.2 we have seen that signals generated by single neuronal cells can propagate across entire populations of neurons and their activity can become synchronous. This means that in particular circumstances and areas of the brain, electric fields generated by these populations are more likely to sum without canceling each other and become enough strong to be measured by EEG electrodes posed on a subject's scalp. In particular, these areas are the ones belonging to the brain's cortex. Here are located particular types of cells called cortical pyramidal neurons, which are of fundamental importance for EEG recordings, because they are oriented perpendicularly with respect to the cortical surface and almost parallel to each other. In such a way they are capable to send electric currents in the same parallel directions, that can sum and increase the magnitude of the associated electric fields.

Despite the fact that brain tissue, skull and scalp have different and non homogeneous electric conductivities, these are anyway sufficient to guarantee the propagation of electric signals up to the most superficial layer of the skin. Then, since air's electrical conductivity is almost zero, a physical electrical bridge must be interposed between the skin of the scalp and the EEG electrode. For this purpose, an electroconductive gel or salt-water-soaked sponges are usually utilized.

At this point, is probably clear that EEG technique is not able to measure all classes of events that occur in brain dynamics. First of all, noninvasive EEG is not enough sensitive to detect molecular or synaptic events nor action potentials coming from single neurons. EEG cannot even measure local field potentials generated by small ensembles of neurons. These type of events generate either no electrical fields or weak electrical fields that are not enough strong to propagate on top of the scalp and be measured by external electrodes. Instead, These events can be measured only using invasive or in vitro EEG techniques. Even some electrical fields that should be powerful enough to be measured on top of the scalp, may not be detected by noninvasive EEG. For example, field potentials on opposing sides of a sulcus would cancel if they had similar strengths.

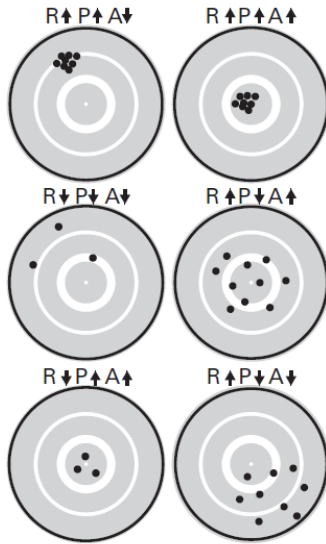
Another class of events that are not measurable by noninvasive EEG are electric signals coming from deep brain sources such as thalamus, basal ganglia, hippocampus and brainstem. Two main reasons are at the basis of this difficulty. The first one is that the field strength decreases as an the inverse of the squared distance [30], so even powerful fields generated by deep brain areas have only a small impact on the measured signal. To detect deep brain activity from the scalp, deep brain sources must produce very powerful fields and it is necessary to take many trials for averaging, to guarantee a sufficient signal to noise ratio. In second instance, activity from deep brain structures is difficult to measure from the scalp because populations of neurons in subcortical structures are not often arranged in a



**Figure 1.3:** Illustration of local field potentials represented as electric dipoles in different orientations with respect to the skull. Dipoles in (a) will contribute majorly to EEG signal. Dipoles in (c) are unlikely to be measured because they cancel each other. Dipole in (d) will make a smaller contribution to EEG than dipole (a) because it is far away from the electrode. Picture taken from [25].

geometrically parallel orientation. For this reason it is more likely that electrical fields generated by synchronous activity of single neurons cancel each other out at the macroscopic scale rather than summing. Although it is possible to measure deep brain activity using scalp EEG, it remains anyway inconvenient to use this type of technique for such measurements.

At this point it is important to introduce some terminology and other features that characterize EEG technique. In particular we want to clarify the meaning of three terms that, despite the apparent similarity, refer to distinct concepts. These are spatial and temporal resolution, precision and accuracy.



**Figure 1.4:** Intuitive illustration of the meaning of resolution (R), precision (P) and accuracy (A). Up and down arrows indicate high or low level. Resolution is the number of dots, precision is the spread of the dots and accuracy is the distance of the dots from the center. Picture taken from [25].

the analysis that the experiment requires.

Spatial precision of EEG is low. This is due to the fact that a signal generated by the same biophysical sources can be detected by different electrodes on a subject's scalp. This surely induces autocorrelation phenomena between the recordings, that can be attenuated or removed by using spatial filtering techniques such as surface Laplacian or adaptive source-space-imaging. Alternatively, to increase spatial precision, a high number of electrodes can be used ( $> 200$ ) and placed on a subject's scalp by following individual anatomical precise models of brain and skull. However, with this latter method, a higher precision is harder to obtain with respect to the first one.

As for the case of precision, spatial accuracy of EEG is low. This is due to the fact that recorded activity from one electrode, does not necessarily reflect the activity of the neurons located directly below that electrode, but rather, from a mixture of activities from different regions that can be close or distant from the considered channel. Moreover, the extent to which a brain region contributes to the recorded signal depends on cortical anatomy and to what extent that brain region is active at a certain point of time. Despite the fact that spatial precision for noninvasive EEG is very low, for the purpose of studying speech envelope tracking, it was not necessary to locate precisely where auditory cortex was placed in the brain of the babies. Instead, some anatomic models were used and electrodes have been placed in proximity of auditory cortex's predicted locations.

As was said before, noninvasive EEG can not measure all kinds of events that occur in brain dynamics and only activity coming from regions of certain spatial dimensions and brain areas can contribute. In particular, focusing on cortical areas, that are the most important when it comes to noninvasive EEG,

Temporal resolution refers to the number of acquired data samples per unit of time. For EEG this sampling rate typically spans from  $250\text{ Hz}$  up to  $1000\text{ Hz}$ . Temporal resolution is what allows the experimenter to extract frequency band specific information.

Temporal precision refers to the certainty of the measurement at each time point and mainly depends on the applied analysis to the EEG signals as well as on the considered frequency band. Unfiltered signals have the highest temporal precision because the value of the measured potential at each millisecond was sampled from brain activity only at that millisecond. On the other hand, temporal precision of a filtered signal is lower because the process of filtering act as a weighted average of temporally surrounding activity. Moreover, generally, higher frequency bands have greater temporal precision.

Temporal accuracy refers to the relationship between the timing of EEG signal and the timing of the biophysical events that are directly linked to that signal. The accuracy for EEG is generally very high because brain electrical activity travels almost instantaneously and EEG is a high temporal resolution technique.

Spatial resolution is determined by the number of electrodes that are used in the measurement and, like we have already said above, it is adjusted depending on

we can differentiate three spatial scales of neural organization.

The first is microscopic scale, that refers to cortical volumes of less than a few  $mm^3$ , hence including neurons and synapses. For the reasons specified before these areas are invisible to EEG. The second spatial scale is the mesoscopic scale, that refers to patches of cortex of several  $mm^3$  to few  $cm^3$ . Dynamics occurring at this scale can be detected by EEG although it may require high spatial resolution and spatial filtering techniques. Finally there is the macroscopic scale, composed by volumes of cortex of several  $cm^3$ . This spatial scale is easily measurable with EEG using even only a few channels.

## 1.4 EEG strengths and limitations compared to other techniques

At this point, it is useful to write a brief summary of the advantages and the limitations of using EEG technique to study neural processes, in particular neurocognitive functions.

The first positive aspect of using EEG technique is that the measured signal is multidimensional. Despite it seems to contain only the two dimensions of time and space (where space refers to all the different locations of the electrodes on the subject's scalp), it comprises at least four dimensions. These are: time, space, frequency, power (the intensity, intended as the square of the amplitude of the signal's oscillations) and phase (the timing of the activity). The brain can be thus conceptualized as a complex biological system that is based on a multidimensional space and its complexity suggests that studying some of its processes in terms of critical phenomena may be useful.

A second reason for which noninvasive EEG is an optimal technique to investigate cognitive dynamics is that it can capture rapid processes. Most cognitive processes are in fact very fast and take place within tens to hundreds of  $ms$ . Furthermore cognitive events typically occur in a temporal sequence that can span hundreds of  $ms$  to a few  $s$ . High temporal resolution techniques, like EEG, are well suited to capture these fast and temporally sequenced dynamics because of their high sampling rate.

A further reason for using EEG is that it represents a direct measurement of neural activity. Electrodes record field potentials that are directly associated with underlying biophysical phenomena at the level of populations of neurons. Moreover, oscillations that are observable from EEG signals are direct reflections of neural oscillations in the cortex.

For these last two reasons, EEG is hence a technique better suited to study cognitive processes with respect to other measurements like fMRI's based BOLD. Briefly explained, fMRI (acronym for functional magnetic resonance imaging) is a noninvasive technique that measures the hemodynamic response correlated with brain neuronal activity in men and animals in order to determine which brain areas activate during the execution of a specific task, like for example, talking, reading, listening or thinking. For research purposes, fMRI allows for instance to see whether, in proximity of a brain injury, a significant brain functional activity persists. BOLD (acronym for blood oxygenation level dependent) signal is the fMRI contrast of blood deoxyhemoglobin. To carry out its fundamental functions, the neurons do not store internal reserves of glucose and oxygen. Typically, in response to a demand for information processing, an increase in neuronal activity follows and this requires more glucose and oxygen to be rapidly delivered via the blood stream. Via this hemodynamic response, blood releases glucose and oxygen to active neurons at a faster rate relative to inactive neurons. This results in a surplus of oxyhemoglobin localized to the active areas, giving rise to measureable changes in the local ratio of oxy- to deoxyhemoglobin, providing a local marker of activity for fMRI. However, the hemodynamic response is a process linked to metabolism and, for this reason, it is slow and requires a time interval of 20  $s$  to 30  $s$  to be appreciated. So, fMRI can only provide a snapshot of a process that took place relatively far in the past, but not an instant image of the present. Furthermore, it is an indirect measure of neural activity and is hence sensitive to other physiological non neural influences.

In this frame we understand that EEG reveals also some weaknesses and in general it is not suited for addressing all research questions. In particular, as already said in paragraph 1.3, it shows a poor precision in locating the sources of brain processes and is not suited for testing hypotheses about deep brain structures. Moreover, EEG is suboptimal when it comes to study cognitive processes that are slow and that have an uncertain and variable time course, like for example, social and emotional processes. EEG and fMRI seem thus to achieve complementary functions. While EEG is fast in detecting neural processes but does not provide a precise localization of active brain areas, fMRI is

convenient for locating the processes, but at the same time, requires a lot more time to measure the activities. Again, for the purpose of studying speech envelope tracking, a fast and less spatially precise technique is required, and EEG fits better in this setting with respect to fMRI.

Another aspect that represents an inconvenience, rather than a limitation, of EEG measured signal is that it is extremely affected by background noise. This noise is hardly traceable in terms of task related signal coming from neurocognitive processes and, instead, can be associated with mechanical noise, muscle activity from the jaw or neck, eye movement and other kinds of noises coming from different physiological or physical processes. Background activity is hence difficult to interpret with respect to task activity, however it can be easily removed or strongly attenuated by using baseline normalizations or filtering the signals.

All the above discussion that has been done for EEG can be applied, with some modifications, to another type of technique called magnetoencephalography, to which we will refer from now on as MEG. MEG uses the same principles of EEG, but instead of measuring field potentials, it measures the changes in induced magnetic fields from neuronal ensemble's electric currents. Brain's magnetic fields from cortical activities are typically on the order of the  $fT$ , so they are extremely weak, especially considering the fact that the ambient magnetic noise in an urban environment is on the order of  $0,1 \mu T$ . MEG hence allows to detect neural oscillations in an alternative way respect to EEG and often, but not always, they produce the same results. Despite the similarities, there are some differences in what these two techniques measure. Taking as a reference figure 1.3, it is possible to say that EEG can detect both radial and tangential sources (this refers to the orientation of the dipoles with respect to the skull) and is maximally sensitive to radial dipoles on gyral crowns. On the other hand MEG is maximally sensitive to tangential sources and shows low sensitivity to radial ones. This MEG characteristic would represent a problem in an experiment if brain activity is expected to occur in small localized areas, but in cognitive processes like the one we are interested in, large patches of cortex are activated and both EEG and MEG are valid.

In general, MEG is better than EEG in detecting high frequency activity (above  $60 Hz$ ). This is due to the fact that magnetic fields pass undisturbed through the skull and scalp, while on the contrary, electrical fields are conducted through these tissues and this causes a decrease in signal to noise ratio at higher frequencies.

Another reason to prefer MEG over EEG is that, while in theory source reconstruction accuracy should be similar for both of them, in practice it is often higher for MEG. This is because EEG setups have fewer electrodes compared to MEG sensors and because the precise positions of EEG electrodes are rarely measured basing on subject specific models, instead standard templates are used.

Nevertheless, EEG has several practical advantages over MEG that somehow unbalance the choice between the two techniques. MEG requires special rooms and intensive and expensive maintenance, thus for obvious reasons most universities and hospitals do not have a MEG apparatus. On the contrary, EEG is portable and can be easily transported everywhere. Most universities and hospitals have several EEG setups, also because it is less expensive compared to MEG. A practical disadvantage of EEG is that every electrode requires the application of an electroconductive gel on it.

## 1.5 Types of feasible analysis from EEG recorded signal

In general, for various reasons, microvolt measured values can be difficult to compare in absolute sense between different subjects and trials and this could lead to difficulties in interpreting EEG signals. First of all, some reasons depend on the experimenter and, in particular, on data processing and analysis decisions, including the choice of reference electrode and the time period used for baseline subtraction. Furthermore, spatial and temporal filtering will change the microvolt values. Another class of reasons depend on the subject that is studied and, in particular, on its individual skull shape and thickness, its scalp preparation during the recording (even whether the subject washed his or her head the morning of the acquisition) or the anatomy of its cortical folding and other factors. A final set of reasons is related to the utilized instrumentation, like, for example, the quality of electrodes, amplifiers and filters.

All these factors influence all trials equally for a same subject, but generally, raw voltages are difficult

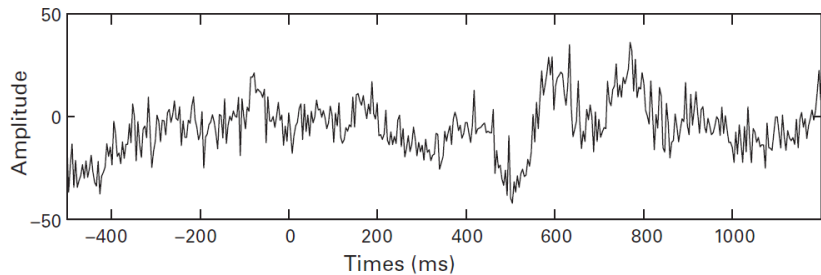
to compare between different subjects and studies. For this reason, it is good practice to interpret the general pattern of the data, rather than the absolute differences in microvolt values. To accomplish that, transformations of the original data scales are often applied in the analysis.

As we have seen in section 1.4, one of the advantages related to EEG is that it is a multidimensional measurement. As a consequence, the experimenter can chose different methods of analyzing the data, basing on convenience and on the information that is interested to extract. Below we introduce two main techniques, listing some reasons for using them, but also some of their weak points.

### 1.5.1 Event related potentials (ERPs)

ERP is a representation of the electrophysiological response elicited by an external stimulus and displays the average over repeated measurements of the variation in time of the field potential (typically in  $\mu V$ ) measured by an electrode with respect to a channel taken as reference. This process of averaging over different trials, is done in order to increase the signal to noise ratio, that, for a single EEG signal, is very high. In figure 1.5 is represented a typical example of unfiltered ERP signal.

There are several valid reasons to analyze EEG data using these types of diagrams. First of all, they are really simple and fast to obtain and require few analysis assumptions or parameters. Due to the fact that ERPs are fast to compute and easy to look at, they allow to quickly check the quality of recorded data to detect the presence of eventual artifacts. Secondly, ERPs' temporal precision and accuracy are high. They only need little



**Figure 1.5:** Example of unfiltered ERP signal coming from EEG measurement. Picture taken from [25].

data processing and and temporal filtering. Because of these characteristics, ERPs can provide a more accurate estimation of poststimulus latency than time-frequency analysis. The last point in favour of ERPs is the presence of extensive literature and findings.

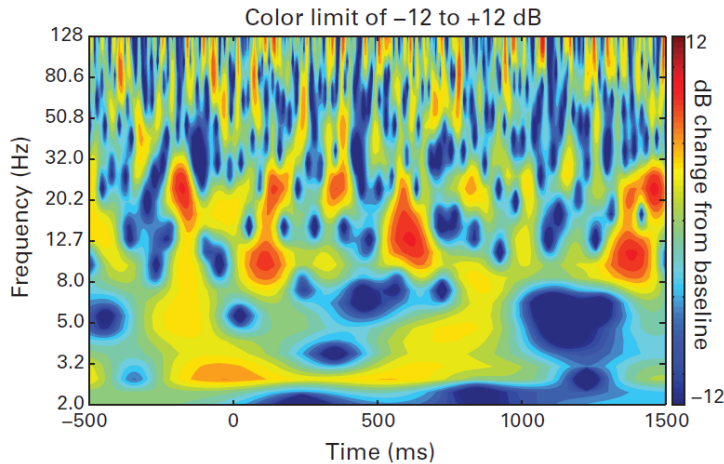
On the other hand, there are two main limitations of ERPs. The first one is represented by the particular difficulty in interpreting less significant data compared to time-frequency analysis. This happens because ERPs only reveal a small part of the collected information coming from EEG measurements. Furthermore, despite the fact that ERPs represent an optimal way to visualize a small part of EEG data, they do not give information on the dynamics underlying neurophysiological phenomena elicited by responses to external stimuli.

### 1.5.2 Time-frequency analysis

Time-frequency is a type of analysis that is particularly suited to describe auditory sensations. Instead of presenting the power of the signal as a function of the only time or frequency spectrum (as it is done in Fourier analysis), it is represented as a function of both time and frequency in particular graphs called information diagrams. In figure 1.6 is represented a typical example of time-frequency diagram. Since the measured signal does not extend for an infinite time, it is possible to define an uncertainty on its frequency spectrum:  $\Delta f$ , obtainable through Fourier analysis, which is inversely proportional to its temporal duration:  $\Delta t$ . Without entering into the details of this procedure, by using a formal mathematical apparatus analogous to that of quantum mechanics and by working with a complex expansion of the measured signal, it can be proved that an uncertainty relation between  $\Delta f$  and  $\Delta t$  holds. In particular, this relation is given by the formula:

$$\Delta f \Delta t \geq \frac{1}{2} \quad (1.1)$$





**Figure 1.6:** Example of time-frequency diagram of signal's power from EEG measurement. Picture taken from [25].

have been developed before 1945 by Dennis Gabor in order to study the analysis of hearing and develop a theory of communication [31].

Time-frequency analysis does not always provide beautiful results and it is not necessarily the best way to analyze EEG data in every situation, but certainly, in some cases, provides some advantages. The first one is that many results from time-frequency based analyses can be interpreted in terms of neural oscillations, which seem to be a fundamental mechanism that support numerous brain functions across multiple spatial and temporal scales [32]. This fact becomes important when the experimenter is interested in studying the connections between cognitive and underlying neural processes. The second advantage of using time-frequency analysis is that it allows to obtain important results from the multidimensional space described by EEG data and, in particular, it reveals a wider part of this multidimensional space with respect to ERPs diagrams.

However, there are also three main limitations of time-frequency approaches. First of all, for the general properties of rhythmic signal analysis, time-frequency decomposition causes a decrease in temporal precision. In particular, the entity of the loss depends on the frequency band of the measured data and on the parameters used for the analyses, but it can be anyway mitigated adopting some analysis strategies. However, temporal precision of time-frequency based approaches will almost always be worse than that of ERPs, but better than that of fMRI. A second limitation is that the large number of analyses that can be applied to EEG data with these techniques and their eventual complexity can be disorienting and potentially lead to improper or incorrect analyses and interpretations. The last limitation is represented by the relatively small literature that links time-frequency analyses to cognitive processes, compared to that of ERPs.

## 1.6 Causality relation between electric fields and cognition

Unfortunately, the causal connection between electrical fields and cognitive processes has not already been firmly established by a numerous amount of evidences. However, there is a significant amount of studies that points in favour of this causal relationship. One line of evidence comes from in vitro studies on the connection between local field potential oscillations and synaptic processes that are thought to be at the basis of the functioning of learning and memory [33].

Some theories suggest that oscillatory synchronization between different areas of the brain is a mechanism at the basis of information transmission across neural networks: a process of crucial importance for cognition [34, 35, 36]. The underlying idea of this mechanism is that spatially separate neural networks most efficiently cooperate and transfer information when they are phase synchronized. Due to this fact, phase-based synchronization methods are the most used techniques to study connectivity,

In the light of this relation, it is hence possible to divide the whole area of an information diagram, that is covered by the signal, into cells of fundamental dimensions equal to  $\Delta t$  on the time axis and  $\frac{1}{2\Delta t}$  on the frequency axis. In this way the cells cover an area equal to  $\frac{1}{2}$ , that is the minimum acceptable area admitted by the uncertainty principle, such that satisfy the relation with the equality sign. These fundamental cells are called logons and can be interpreted as quanta of information, containing an elementary signal. In particular, in every logon, the measured signal can be expanded in terms of these elementary signals, constituted by the combination of a periodic oscillating function with a Gaussian pulse. These observations

and, in particular, are also used to calculate phase synchronization in [10].

Other evidences that aim to determine whether oscillations are causally involved in brain computation come from methodological approaches. One of them is represented by optogenetics in mice and rats. This technique involves shining a light at a particular wavelength into the brain of a mouse and, by doing this, oscillatory activity at a desired frequency can be induced in a specific brain area. For example, it has been demonstrated that gamma band oscillations can be exogenously enhanced via optogenetics and the enhancement facilitates signal transmission and noise suppression [37]. Another methodological evidence is given by transcranial alternating current stimulation (TACS) in humans. This technique consists in passing an electrical current of frequency around  $0, 1\text{ Hz}$  and  $100\text{ Hz}$  between two electrodes placed on the scalp. It has been proved, for example, that the stimulation of human's motor cortex using TACS with a  $20\text{ Hz}$  current, increases motor evoked potential magnitude, whereas stimulation at other frequencies does not [38].

Theories on the role of oscillations in cognition may be wrong and electric fields produced by neural populations may not be causally involved in cognitive processes but, rather, be curious consequences of the mechanisms that truly underlie cognition. If electrical fields were not causally involved in cognition, this certainly would not be the end of cognitive electrophysiology. Electrical fields produced by populations of neurons are undeniably indices of brain function and the use of field potential oscillations, measured through EEG to study brain organization, would remain valid.

To make a parallelism, it is not widely believed that fMRI's BOLD signal is a causal mechanism of neural information processing. However BOLD is undoubtedly an indirect index of brain function and lack of causality does not stop scientists from using BOLD signal to understand the brain.

## 1.7 Brief insight on critical phenomena in the brain

The reasons why apparently different systems, that may display similar behaviours in certain conditions, are classifiable as critical, are represented by the three pillars of the theory of critical phenomena: scaling, universality and renormalization. Scaling principle is subdivided in two different categories of predictions. One of them is represented by scaling laws: functional relations between the critical point exponents associated to a certain phenomenon. Universality concept reflects the empirical fact that different systems can be subdivided into universality classes. In particular, two systems belong to the same universality class if they have the same critical point exponents and scaling functions. Finally, the principle of renormalization states that the critical point of a system can be mapped into a fixed point of a transformation of the system's Hamiltonian. This is the less useful principle for our scopes, because we start from measured EEG data and not from a model of brain's Hamiltonian.

The fact that different critical systems share the same critical exponents and scaling functions creates thus an interesting parallelism between well known examples of critical phenomena in physics and other emergent critical processes in other fields (in this case, neural processes and cognition), and may highlights the possibility that the same fundamental physical principles underlie the studied phenomena.

The concept of criticality in neurophysiological systems was for the first time introduced by Walter Freeman in 1987, from empirical analyses and computational models of the olfactory bulb in rabbits [39]. Freeman's studies were the first proof that science of critical phenomena could be used to investigate the brain and cognition processes and, since then, a large number of studies investigated the brain through perspective of criticality. More recent works have been concentrating on fluctuations of EEG and MEG rhythmic signals and found that they display a frequency scaling close to  $1/f$  [40]. Moreover, another result that may be useful in speech envelope tracking, is that a scale free statistics has been observed in the fluctuations' levels of synchronization between different pairs of electrodes [41].

Another work was conducted by Beggs and Plenz in 2003 and aimed to investigate evidences of criticality in the erratic spontaneous activity of neurons in neuronal cultures in vitro, finding temporal power law behaviour in the duration of burst activity and spatial power law behaviour in the number of electrodes involved in every burst [42]. In addition to the obtained results, the importance of this study comes from the fact that critical activity, observed in the cultures, maintained itself stable for several hours, without any need to tune sistem's parameters, like pH, temperature and so on. This

fact can be justified by assuming that self organized criticality occurs in the brain, in addition with the hypothesis that the interactions between system's components, the neurons, are characterized by a certain plasticity. It is believed that the system slowly accumulates energy that is then quickly dissipated through large scale avalanche processes. Thus, a certain amount of time passes between the dynamics of fast release and slow build up of energy and, this slow plasticity, is thought to play a crucial role in the diffusion of critical dynamics across broader brain areas. Several models have been elaborated in order to explain these processes of self organized criticality and slow build up and subsequent dissipation of energy, like, for example, balanced processes of excitation and inhibition [43] and replenishment and depletion of intracellular metabolic resources including ATP [44].

The above studies only represent a brief list of the most important evidences in favour of a relationship between brain processes and a plausible underlying critical phenomenology, but other evidences are gradually accumulating. In synthesis, a significant number of neuronal systems and neural models placed at critical point display several adaptive benefits from criticality, including maximum dynamic range, optimal information capacity, storage and transmission and selective enhancement of weak inputs [45]. Furthermore, it has been shown that resting state EEG and fMRI data display evidence of critical dynamics and, for this reason, it has been supposed that a critical state may be crucial for the brain to keep an "explorative" resting state and that the onset of a specific cognitive function may reflect the stabilization of a particular subcritical state under the influence of sustained attention [45]. Finally, other phase transitions to pathological states have been observed in several neurological disorders such as epilepsy and neonatal encephalopathy.

However, in general, the existence of criticality in neural activity has not already been firmly established, and the presence of  $1/f$  power laws does not necessarily imply self organized criticality.

Progress in the field of criticality in neuroscience needs thus to proceed with due caution, using appropriate analysis methods, like maximum likelihood estimation based approaches [46]; considering alternative complex processes as possible justifications of  $1/f$  scaling, like possible filtering properties of extracellular media [28]; and using computational models in partnership with data analysis.



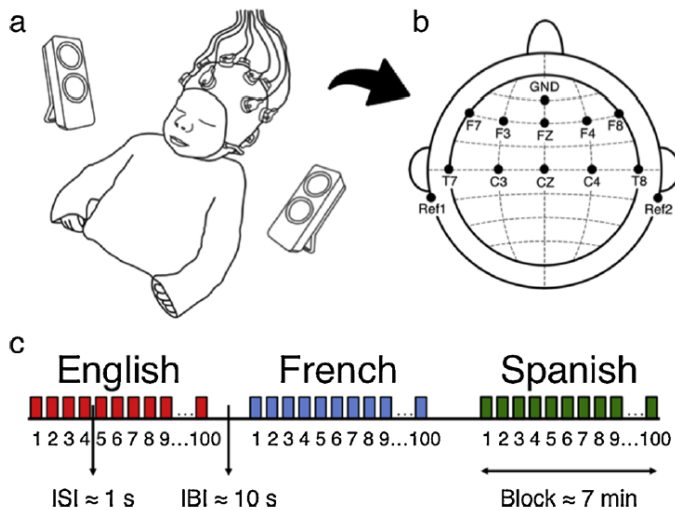
## Section 2

# Experiment

### 2.1 Subjects and data acquisitions

#### 2.1.1 Newborns

In the present and following subsections will be briefly summarized some characteristics of the subjects and the materials involved in the experiment [10]. A total of 55 newborns participants have been tested during their hospital stay in Robert-Debré Hospital in Paris. The requisites for the participation to the test were: being full term and healthy; having a birth weight  $> 2800\text{ g}$ ; being maximum 5 days old and being born to French native speaker mothers who spoke this language at least 80% of the time during the last trimester of pregnancy. For reasons of technical problems, bad data quality and difficulties in finishing the experiment due to fussiness and crying, 8 participants had to be excluded from the experiment. Hence, analysis included EEG data from 47 newborns: 20 girls and 27 boys of average age of  $2,45 \pm 1,18$  days.



**Figure 2.1:** Illustration of the experiment. (a) Newborn with EEG cap. (b) Location of EEG electrodes. (c) Grouping of sentences and languages. ISI: interstimulus interval, IBI: interblock interval. Picture taken from [10].

Participants have been divided into three groups and every group listened to one of three possible sets of repeated sentences, presented with naturally spoken characteristics. Auditory stimuli were delivered bilaterally through two loudspeakers positioned on each side of the bassinet in which babies were in a state of rest or asleep. The volume was set to a typical conversational level ( $\sim 65 - 70\text{ dB}$ ). As was said in the introduction, each group of sentences was subdivided in three blocks of 100 repeated sentences. One for each language: French (native language), Spanish (unfamiliar rhythmically similar language) and English (unfamiliar rhythmically different language). The order of the language blocks was pseudo-randomized and counterbalanced across participants and a pause of 10s between the blocks have been introduced. Furthermore, sentence repetitions were separated by a time interval of 1 – 1,5 s.

The whole listening process lasted around 21 minutes (7 minutes for each block) and, simultaneously with the delivery of the auditory stimuli, newborns' neural activity was recorded using EEG.

### 2.1.2 Six-month-olds

6 months old infants have been recruited in two different ways. 20 of them came from the group of participants who took part to the newborn experiment, while 14 of them have been contacted for the very first time. The requisites for the participation were: being full term and healthy; being 6 months old  $\pm 15$  days; being exposed to French at least 80% of the time since birth. Due to some difficulties encountered in contacting the participants, 4 of them have been tested outside the intended age range: one 5 months old, one 10 months old and two 11 months old. 9 participants have been excluded from data analysis because of bad data quality, technical problems or not finishing the experiment due to crying. Hence the analysis comprehended 25 infants: 10 girls and 15 boys of average age of  $204 \pm 48$  days.

As was done for newborns, participants have been divided in three groups: one for each possible set of sentences. The listening modalities remained similar to those of newborns, but with some important differences. 6-months-olds have, in fact, less attention span than newborns, thus, for this reason, testing time have been reduced through a choice of a more appropriate number of repetitions: 50 instead of 100. By doing this, the whole listening process took half of the time with respect to that of newborns' experiment. Furthermore, to better maintain 6-month-olds' attention, babies were seated on a parent's lap while a silent video of a children's animated cartoon was presented to them. For the participants who were tested also at birth, it has been ensured that they would not hear the same auditory stimuli, to avoid any memory effect. Another important difference is that, at this age, infants can discriminate rhythmically similar languages [15, 16], whereas they could not do it at birth [13, 14]. For these reasons the rhythmically similar languages: French and Spanish, were always presented in the first two blocks in a pseudo-randomized and counterbalanced order.

### 2.1.3 Stimuli and data acquisition

Sentences were drawn from a children's fairy tale and recorded in infant directed speech by a female native speaker of every language at a sampling rate of  $44,1 \text{ kHz}$ . After making sure that there were no differences between the sentences in the three languages in terms of pitch, it has been computed the amplitude and frequency spectra of the sentences and found that utterances were similar in every spectral decomposition.

To record EEG data, 10 active electrodes have been used and placed in locations that can be seen in figure 2.1b. Two additional electrodes were used as online reference and placed on each mastoid, while another electrode was used as ground and placed on the forehead. The electrodes were placed in proximity of areas that are typically associated with auditory and speech perception observed in infants. Data were recorded with a sampling rate of  $500 \text{ Hz}$  and electrodes' impedances were kept below  $140 \text{ k}\Omega$  for newborns and  $40 \text{ k}\Omega$  for 6-month-olds.

## 2.2 Hilbert transform

The Hilbert transform is a linear map that transforms a function of real variable  $s(t)$  into a complex function of real variable  $\mathcal{H}\{s\}(t)$ . It is an important mathematical operation in signal processing because it allows to obtain the complex version of a measured real valued signal  $s(t)$  in function of time, called analytic signal. Before applying the Hilbert transform to any signal, it is important to remember that it always has to be previously filtered, in order to isolate the frequency associated to the processes that we are interested in studying. This is particularly true for EEG signal, that is very noisy due to the presence in the measurements of other non task related noises. The idea behind the Hilbert transform applied to neural signal analysis is essentially that it consists in an alternative method for extracting the magnitude and phase information from measured EEG data. The projection onto the real axis of the Hilbert transformation of an EEG signal is the band-pass filtered part of the signal, while the square of the magnitude of the Hilbert transformation, point by point, represent the power time series of the signal and finally, from the angle of every point in the complex plane with respect to the real axis, it can be obtained the phase angle time series. Hilbert transformation is thus used because the complex version of the signal allows to extract information about the power and phase

time series, that can not be directly calculated from the real valued signal. In section 2.3 will be better explained the practical utility of the Hilbert transform in this experiment's methods of analysis. The following is the explicit analytic formula to calculate the Hilbert transform of a function.

$$\mathcal{H}\{s\}(t) = s(t) * \frac{1}{\pi t} = \frac{1}{\pi} p.v. \int_{-\infty}^{+\infty} \frac{s(\tau)}{t - \tau} d\tau = -\frac{1}{\pi} \lim_{\epsilon \rightarrow 0} \int_{\epsilon}^{+\infty} \frac{s(t + \tau) - s(t - \tau)}{\tau} d\tau \quad (2.1)$$

Where the symbol  $*$  indicates the operation of convolution between the two functions and  $p.v.$  indicates the principal value of the following integral.

In order to have an intuitive interpretation of the action of the Hilbert transform, it is useful to write its connection with the Fourier transform.

$$\mathcal{F}\{\mathcal{H}\{s\}\}(\omega) = -i \operatorname{sgn}(\omega) \mathcal{F}\{s\}(\omega) \quad (2.2)$$

Where  $\omega$  represent the frequency variable.

Thus, in general, the Hilbert transform act by shifting the original signal by  $\pi/2$  in the frequency domain, leaving unchanged the power spectrum.

## 2.3 Methods for amplitude tracking analysis

In this section we briefly introduce the analysis processes that have been applied to the data in [10] and the main reasons they were applied in order to assess amplitude tracking.

Let us start by describing EEG signal processing. First of all, EEG raw signal, directly measured by the electrodes, have been band-pass filtered between 1 *Hz* and 40 *Hz*. This range of filtering is typical for the cognitive processes that are involved in this experiment. Furthermore, it strongly damps the numerous non task-related noises and eventual disturbs coming from the power grid that affect EEG. The whole recorded EEG signals have been then segmented into epochs in such a way that they corresponded to every repetition of a sentence, and epochs have been properly temporally shifted to take into account the estimated pre-stimulus period of 200 *ms*. The resulting epochs have been then submitted to a rejection process that aimed to exclude epochs with unwanted features. In particular, epochs with amplitude exceeding  $\pm 75 \mu V$  have been rejected; together with those whose standard deviation was higher than 3 times the mean standard deviation of all non rejected epochs or lower than one third of the mean standard deviation and with those that did not satisfy a visual inspection with the purpose of removing other residual artifacts. At this point, by averaging all non rejected epochs per participant, per language and per channel, signals called evoked responses have been calculated.

Moving now to auditory stimuli processing, for the moment we will say that amplitude envelopes of the speech signals were obtained by calculating the magnitude of the Hilbert transform of the waveforms associated to the signal of auditory stimuli. Below will be clarified the meaning of Hilbert transformation. At this point, in order to match the characteristics of the evoked responses, the envelopes were low-pass filtered at 40 *Hz* and down-sampled to 500 *Hz*. In figure 2.2 are reported some results of the speech signals' analysis, starting directly from audio file of the recording of a sentence in English. There are two channels for every single signal: left and right. This because speech signals were recorded in stereo mode. The followed procedure to obtain the plotted results was the same described above. Data processing, starting from audio files, has been made by using Python programming.

Only now that signals associated to EEG and to auditory stimuli have been processed, amplitude tracking can be assessed. Amplitude tracking was estimated basically by performing cross-correlation analysis between the speech envelopes and the evoked responses, valued at the average delay time between stimulus and response for every subject, where the average was calculated across the three different languages.

Let us point an important fact. Apparently, amplitude tracking could have been estimated simply by performing cross-correlation analysis between EEG raw signals and auditory stimuli's associated signals, without any need to process the them. However this would have led to absence of synchronization. One of the reasons for which this would happen is due to the noisiness of EEG signals, that have

to be filtered to isolate electrical activity directly elicited from external auditory stimuli. Furthermore, and perhaps most importantly, what is indeed crucial for speech comprehension is the assessment of the synchronization between signals' modulation. In particular, modulation represent the process of varying one or more properties of a periodic waveform, called carrier signal, with a separate signal, called modulation signal, that typically contains the information to be transmitted. For example, like in this case, the modulation signal can be an audio signal representing the sound coming from a microphone or two loudspeakers. A typical characteristics of the carrier wave is that it is higher in frequency than the modulation signal. The purpose of modulation is to impress the information on the carrier wave, which is used to convey information to other locations through propagation of different physical entities, like electromagnetic waves, for radio or television communications or acoustic waves, like happens for example for a speech. Another purpose is to transmit multiple channels of information through a single communication medium, using a technique called frequency division multiplexing. We can find an example of this principle in cable television, which uses it to propagate many carrier signals carrying multiple television channels through a single cable. Since each carrier is characterized by a different frequency, channels do not interfere with each other. Once arrived at its destination (a radio receiver or a television), the signal is demodulated in order to extract the information bearing modulation signal. In the case of the present study, the role of the Hilbert transform is to be an analytic method of demodulation to extract information from speech signals [47].

The last part of this analysis regarded the assessment of eventual habituation effect. In fact, stimulus repetition may introduce habituation that could have had unwanted effects on the calculated correlation results. To investigate this, correlations between speech envelopes and EEG's single trials have been calculated and then Fisher transformed before statistical analysis. To evaluate the effects of habituation, the transformed correlation values from each language and channel have been submitted to linear regression analyses.

## 2.4 Methods for phase tracking analysis

In this other section we explain the processes of analysis that have been applied to the data in [10] and the main reasons they were applied in order to assess phase tracking.

The first step to assess phase tracking was to determine the syllabic rate of each sentence, simply by counting the number of peaks per second in their speech envelopes, rounded to the closest integer. From this analysis resulted that the syllabic rate for the recorded sentences spanned in a range of  $3Hz - 6Hz$ . Once determined the syllabic rate, EEG raw signals have been band-pass filtered between  $3Hz$  and  $6Hz$ , differently to what was done in the previous section. After filtering, EEG data have been submitted to the same processes of epoching, baseline correction, rejection and averaging explained in section 2.3 to obtain the evoked responses. The speech envelopes were filtered between  $3Hz$  and  $6Hz$  too, to match with the characteristics of the filtered EEG signals.

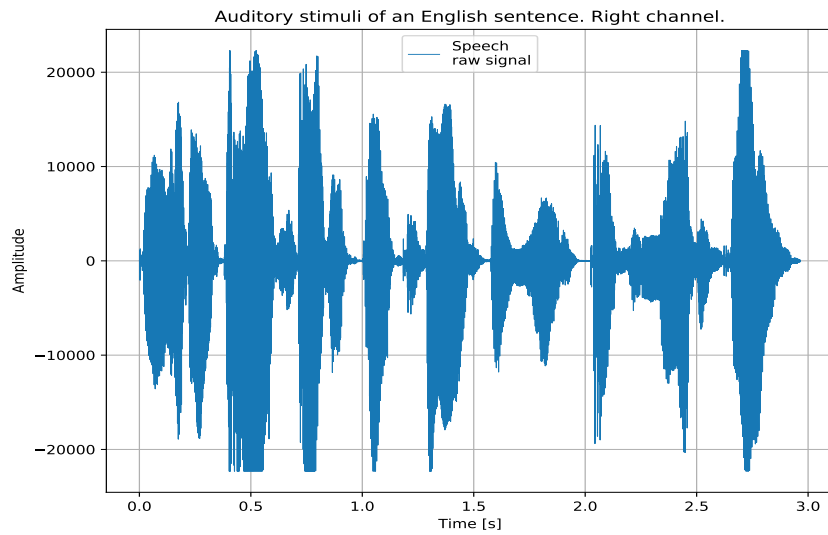
Phase tracking has been assessed by computing the phase coherence between the evoked responses and the filtered speech envelopes. First of all, the Hilbert transforms of both signals have been obtained. secondly, the unitary signals of each analytic signal have been calculated and then, finally, phase synchronization have been calculated between unitary signals. This analysis was performed per participant, per language and per channel.

In order to assess whether the distributions of results coming from correlation and phase coherence were significantly different zero, they have been submitted to t-tests.

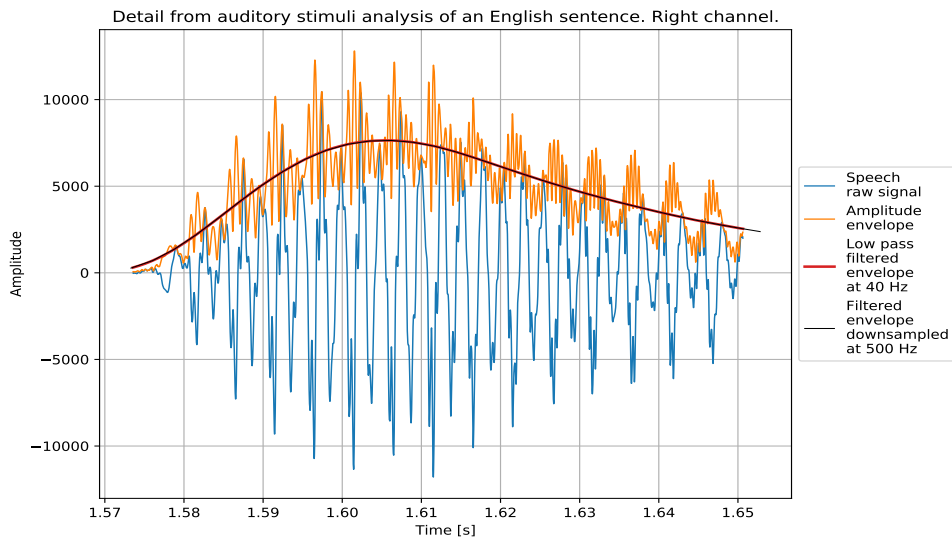
Furthermore, in order to assess whether the correlation and phase-coherence distributions were significantly different from chance, they have been submitted to a permutation test. To do so, EEG data have been shuffled 1000 times along the time dimension per subject, per language and per channel. At each iteration, amplitude and phase tracking analyses have been performed and the so obtained correlation and phase coherence distributions served as a baseline comparison for the amplitude and phase tracking results of the non permuted data.

Finally, to assess whether amplitude and phase tracking were different across languages and/or electrodes, correlation and phase-coherence values from each age group have been submitted to analysis of variances (ANOVA).





(a) Raw signal from an English sentence.



(b) Brief time window detail of the analysis of the signal from an English sentence.



(c) Final result from the analysis of the signal from an English sentence.

**Figure 2.2:** Results of speech signal’s analysis from recorded stereo audio file of an English sentence. The analyzed signal came from the right channel.

## 2.5 Mean phase coherence

A first notion of synchronization was introduced in physics by Huygens for two coupled frictionless harmonic oscillators [48]. For this simple case, phase synchronization is defined as:

$$\varphi_{n,m} = n\phi_a(t) - m\phi_b(t) = \text{constant} \quad (2.3)$$

Where  $\varphi_{n,m}$  is the relative phase,  $n$  and  $m$  are integers and  $\phi_a(t)$  and  $\phi_b(t)$  are the phases of the oscillators.

This definition was then modified in a weaker condition by Rosenblum and colleagues in order to investigate synchronization in chaotic systems [49]. This condition is called phase entrainment.

$$|\varphi_{n,m}| = |n\phi_a(t) - m\phi_b(t)| < \text{constant} \quad (2.4)$$

There exists also an even weaker condition called frequency locking:

$$\langle \omega_{n,m} \rangle = n\langle \omega_a \rangle - m\langle \omega_b \rangle = n \left\langle \frac{d\phi_a(t)}{dt} \right\rangle - m \left\langle \frac{d\phi_b(t)}{dt} \right\rangle = 0 \quad (2.5)$$

Where  $\langle \rangle$  represents averaging over time and  $\omega_{n,m}$  the relative frequency of the systems.

In general, it is difficult to know the phase variables  $\phi_a(t)$  and  $\phi_b(t)$ , in particular for nonlinear and noisy systems. However, the Hilbert transform allows to calculate analytically the instantaneous phase of an arbitrary signal  $s(t)$ , simply by calculating:

$$\phi(t) = \arctan \left( \frac{\mathcal{H}\{s\}(t)}{s(t)} \right) \quad (2.6)$$

As we can see, the so calculated phase is restricted to the interval  $[0, 2\pi]$ , thus it has to be unwrapped before taking its derivative.

However, the above definitions are not optimal to assess phase synchronization in case of noisy or chaotic signals like the one of EEG, because the relative phase can randomly exhibit rapid phase jumps, introducing noise in the relative phase too.

Keeping in mind that we are interested in measuring synchronization between time series obtained from the same physiological system, it is most likely to encounter phase synchronization at a phase locking ratio of  $n : m = 1 : 1$ . Restricting to the case of  $n = m = 1$ , as a measurement for synchronization, we can define a quantity called mean phase coherence as follows [29]:

$$R = \left| \frac{1}{N} \sum_{j=0}^{N-1} e^{i\varphi_{1,1}(j\Delta t)} \right| \quad (2.7)$$

Where  $1/\Delta t$  is the sampling rate of the discrete time series.

It is evident that  $R$  belongs to the interval  $[0, 1]$  and reaches the value 1 only in case of strict phase locking, while for unsynchronized time series it is equal to 0.

Due to its stability in case of noisy time series, the mean phase coherence may represent an accurate measure of phase synchronization for EEG signals.

## Section 3

# Results

### 3.1 Speech envelope tracking at birth

#### 3.1.1 Amplitude tracking

Amplitude synchronization was assessed by performing cross correlation analysis between the speech envelopes and the evoked responses per subject. The cross-correlation function displayed a peak in correspondence of the time lag between stimulus and response at which the envelope tracking took place. Time lags resulted similar between the three languages and, for this reason, it has been computed the average between these delays for each participant. Mean time lag for English language was  $293\text{ ms}$ , for French  $282\text{ ms}$  and for Spanish  $308\text{ ms}$ . In figure 3.1 are shown envelope tracking results for a single subject for a French sentence at each channel. The computed mean correlation values across the whole group of infants are resumed in table 3.1 below. Similar results were observed in adults tested with electrocorticography (ECoG) [50].

Language	Mean correlation values	Range across channels
English	$0,099 \pm 0,0240$	$0,048 - 0,128$
French	$0,122 \pm 0,0191$	$0,090 - 0,142$
Spanish	$0,118 \pm 0,0273$	$0,059 - 0,152$

**Table 3.1:** Mean correlation values for the whole group of newborn participants.

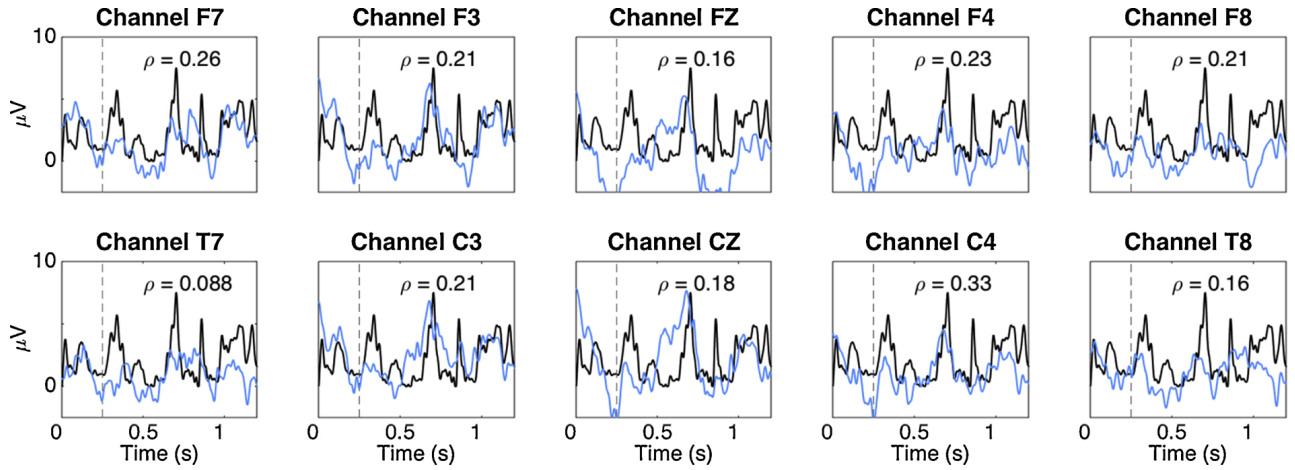
The distributions of cross - correlation values resulted significantly higher than zero and different from chance for every channel and language, with the exception of electrode T7 for French and English and T8 for all the three languages. These results thus indicate a successful amplitude envelope tracking for all channels and languages in newborn subjects.

Regarding the possible tracking differences across different languages or channels, analysis of variances yielded to no significant effects of language or channel. This result imply that newborns track the amplitude of the speech envelope equally well in all the three tested languages and nine electrodes.

Since habituation to repetition of sentences may decrease brain response and hide differences across channels or languages, habituation effects have finally been investigated in order to assess whether repetition had undesired effects on envelope tracking results. Analyses showed that EEG signal's trial number did not predict amplitude tracking, proving that habituation effects did not impact on amplitude tracking results.

#### 3.1.2 Phase tracking

After finding the syllabic rate of the recorded sentences and after processing the EEG data and the speech envelopes, band-pass filtering them between  $3\text{ Hz}$  and  $6\text{ Hz}$ , as explained in 2.4, mean phase coherence has been calculated between the phases of both filtered signals. Mean phase coherence values, calculated across the whole group of newborns, are listed in table 3.2.



**Figure 3.1:** Amplitude envelope tracking of a French sentence in a newborn subject. Black curves represent the speech envelope of the sentence; the blue ones represent the measured cortical activity per channel. EEG data were here shifted backwards using the computed time lag between stimulus and response.  $\rho$  indicates the Spearman correlation between the two signals. The vertical line at 250 ms indicates the beginning of the envelope-following period. Picture taken from [10].

These results are similar to those obtained in adults for the study [7], whose mean phase coherence values were approximately equal to 0,27. Permutation tests and t-tests showed that the distributions of mean phase coherence values were significantly different from zero and chance, highlighting the successful occurring of phase tracking for all languages and channels.

Analysis of variance’s results showed no differences in phase synchronization across languages and channels. Thus, this imply that newborns track the phase of the speech envelope equally well for all languages and channels.

Habituation effects due to repetition of sentences have not been evaluated in phase tracking because phase is not affected by the decreasing amplitude of the stimuli responses.

## 3.2 Speech envelope tracking at 6 months

### 3.2.1 Amplitude tracking

By repeating the same procedure that has been followed in order to assess amplitude synchronization in newborns, the results in table 3.3 can be obtained. These results are similar to those observed in [50], as in the case of newborns.

Important difference come from the evaluation of whether correlation distributions are significantly different from zero and chance. In fact, amplitude tracking has been

Language	Mean phase coherence values	Range across channels
English	$0,319 \pm 0,0200$	0,292 – 0,350
French	$0,331 \pm 0,0147$	0,302 – 0,349
Spanish	$0,343 \pm 0,0198$	0,312 – 0,377

**Table 3.2:** Mean phase synchronization values for whole the group of newborn participants.

Language	Mean correlation values	Range across channels
English	$0,163 \pm 0,0324$	0,094 – 0,217
French	$0,114 \pm 0,0195$	0,082 – 0,149
Spanish	$0,089 \pm 0,0195$	0,068 – 0,130

**Table 3.3:** Mean correlation values for the whole group of 6 months old participants.

found for all channels only for English. For Spanish, only F3 and Fz channels showed significant amplitude tracking, while for French, none of the channels exhibited it.

Analysis of variances showed important tracking differences between English and French and between English and Spanish, but not between French and Spanish. Furthermore, no significant difference between channels, nor habituation effects have been found.

These results mean that 6-month-olds track the amplitude of the speech envelope for the rhythmically different unfamiliar language (English), but not for the native one (French) and only weakly for the rhythmically similar unfamiliar one (Spanish).

### 3.2.2 Phase tracking

Language	Mean phase coherence values	Range across channels
English	$0,312 \pm 0,0181$	$0,276 - 0,336$
French	$0,288 \pm 0,0288$	$0,253 - 0,351$
Spanish	$0,300 \pm 0,0295$	$0,243 - 0,342$

**Table 3.4:** Mean phase synchronization values for whole the group of 6 months old participants.

To assess phase synchronization the same procedure as for newborns has been followed. Mean phase coherence values, calculated across the whole group of 6-month-olds are summarized in table 3.4.

In this case too, results are similar with those obtained in [7].

T-tests and permutation test showed that phase-coherence

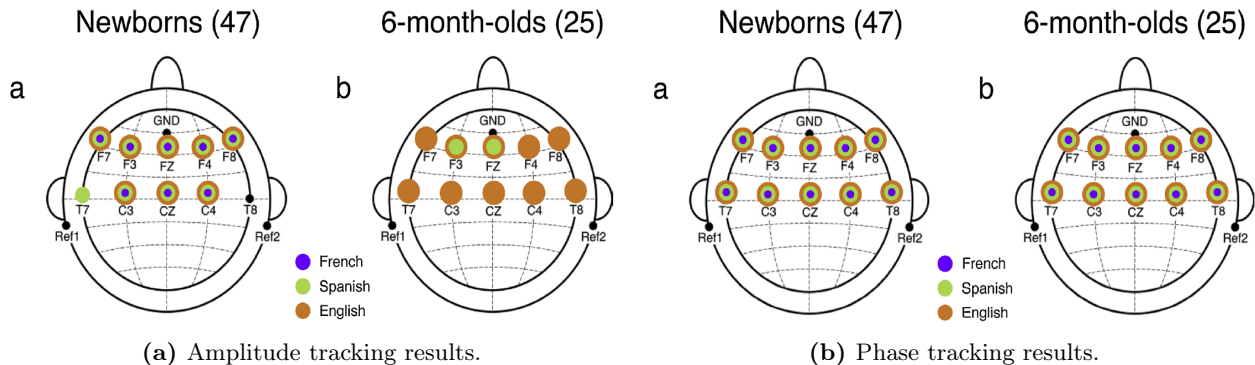
distributions are significantly different from zero and chance for every channel and language.

Analysis of variances showed no significant tracking differences between languages and electrodes.

Habituation effects have not been evaluated for the same reasons explained in 3.1.2.

The obtained results prove that phase envelope tracking for 6-month-olds occurred equally well for all the three languages and for all channels.

In figure 3.2 are graphically summarized all the amplitude and phase tracking results discussed above.

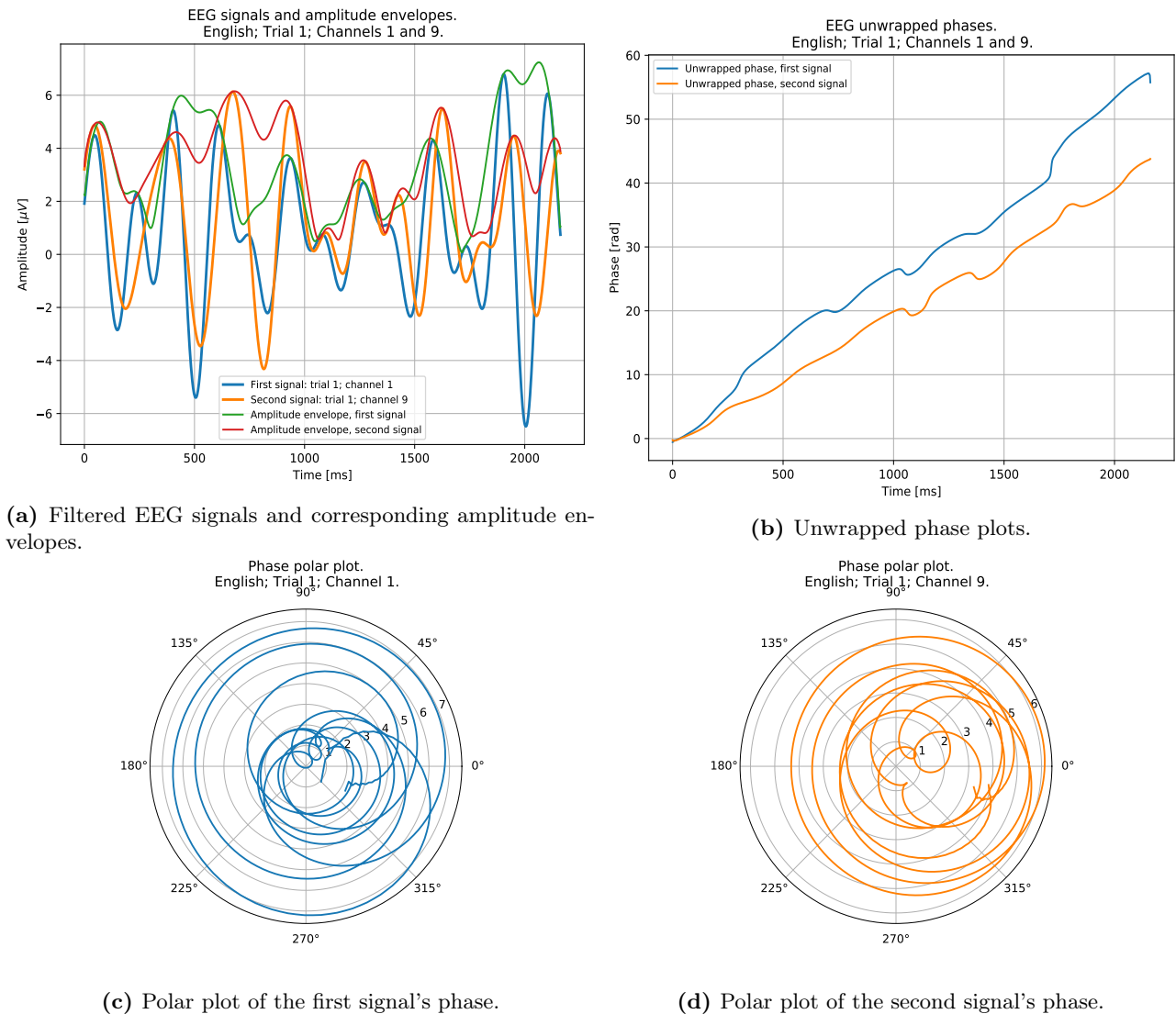


**Figure 3.2:** Topographical distribution of channels with significantly different from zero and from chance amplitude and phase tracking results for each language and group of subjects. Pictures taken from [10].

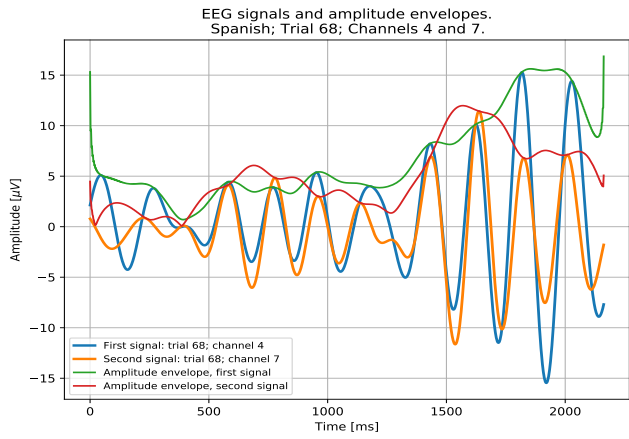
### 3.3 Phase synchronization between different channels

After reviewing the main results of [10] in the previous sections 3.1 and 3.2, we did a small extension to it, the first steps in a fully connectivity analysis. The idea was to calculate phase synchronization between channels, for a few channels on the same data set used in [10]. The data was kindly provided by the authors. It contained band pass filtered EEG signals between  $3\text{ Hz}$  and  $6\text{ Hz}$  coming from all the ten channels and sampled at a frequency of  $500\text{ Hz}$  from a newborn's brain activity for all the three languages. Specifically, signals were composed by 1082 samples for each trial and, recorded data from each language, contained 53 trials for English, 70 for Spanish and 41 for French.

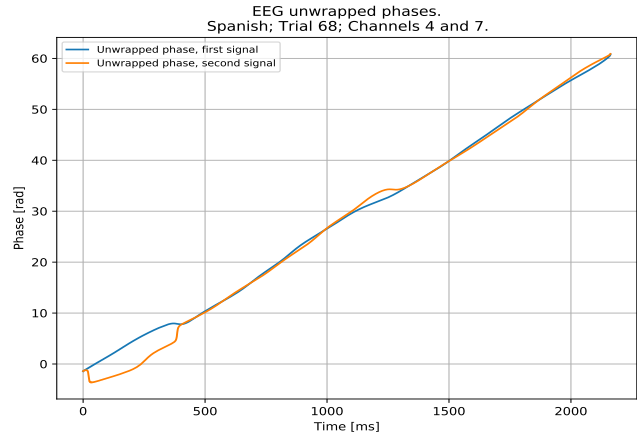
The aim, using these data, was to assess connectivity between different parts of the brain by plotting pairs of randomly chosen signals coming from the same recorded trial, for the same language at different channels, then, plotting their amplitude envelopes and their unwrapped phase spectra from signals' Hilbert transforms and finally calculating the mean phase coherence  $R$  between the two EEG signals as explained in equation 2.7. In order to obtain these results, Python programming language has been used. As an example, in figures 3.3, 3.4 and 3.5 are shown three groups of results: one for each language with their relative computed mean phase coherence values.



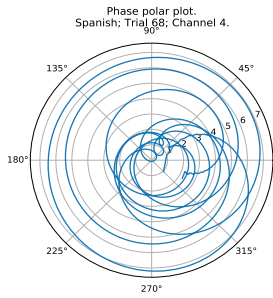
**Figure 3.3:** Computed results for English language. Analyzed signals came from trial number 1 and channels 1 and 9. Computed mean phase coherence value for this pair of signals is:  $R = 0,665$ .



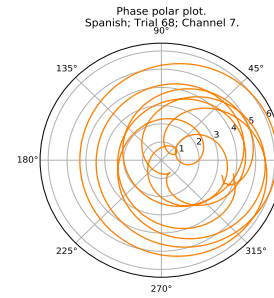
(a) Filtered EEG signals and corresponding amplitude envelopes.



(b) Unwrapped phase plots.

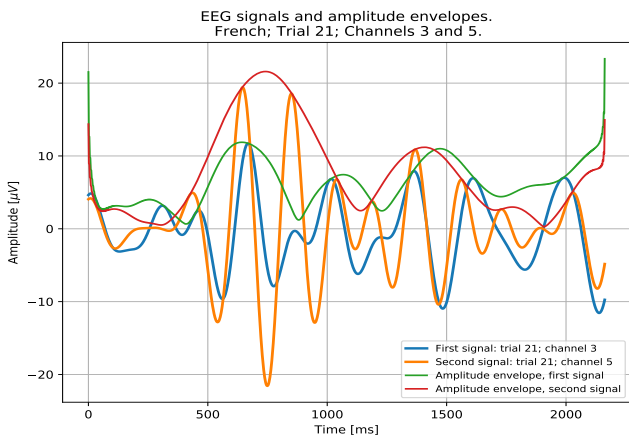


(c) Polar plot of the first signal's phase.

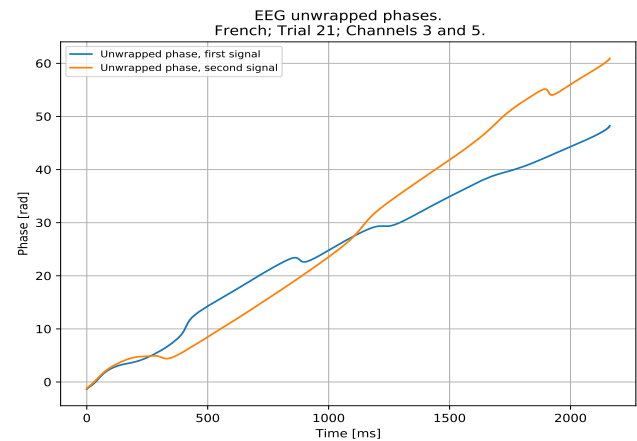


(d) Polar plot of the second signal's phase.

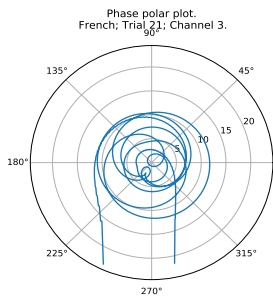
**Figure 3.4:** Computed results for Spanish language. Analyzed signals came from trial number 68 and channels 4 and 7. Computed mean phase coherence value for this pair of signals is:  $R = 0,699$ .



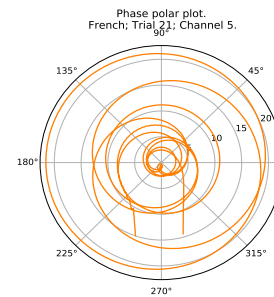
(a) Filtered EEG signals and corresponding amplitude envelopes.



(b) Unwrapped phase plots.



(c) Polar plot of the first signal's phase.



(d) Polar plot of the second signal's phase.

**Figure 3.5:** Computed results for French language. Analyzed signals came from trial number 21 and channels 3 and 5. Computed mean phase coherence value for this pair of signals is:  $R = 0,458$ .





## Section 4

# Discussion and conclusions

In the light of the obtained speech envelope tracking results exposed in the article [10] and resumed in sections 3.1 and 3.2, it is possible to state the following observations.

The fact that, in the newborn participants, both amplitude and phase speech envelope tracking occur for all the tested languages, suggests that speech envelope tracking likely represents a basic auditory ability. From this, it can be thus deduced that speech envelope tracking may only be one of the mechanisms at the basis of speech comprehension, but it is not certainly sufficient alone in this purpose. On the other hand, in six months old subjects, it was found that phase tracking continues to occur for all languages, while amplitude tracking is only observed for the rhythmically different unfamiliar language (English), but not for the native and the rhythmically similar unfamiliar ones (respectively French and Spanish). These results, together with previous findings [51], support the hypothesis that observed changes from birth to 6 months likely reflect language development and may highlight different roles for amplitude and phase tracking. In particular, the nature of envelope tracking may change from acoustic to linguistic, between birth and 6 months of age for the native language and, only slightly, for the rhythmically similar unfamiliar language, while it remains acoustic for the rhythmically different unfamiliar language. Moreover, this work on speech envelope tracking, together with other previous studies suggests that amplitude tracking of the speech envelope for the native language follows a U-shaped curve during development: as shown in above results, it is present at birth, absent at 6 months of age, but then reappears in older children [3] and adults [50]. This represent a common trend in language acquisition [52] and, possibly, reflects a perceptive and neural reorganization. In particular, a shift of the focus from syllabic units of a speech to phonetic ones, that are thought to be relevant for grammar learning. An evidence in support to this idea is that  $\gamma$  frequency band neural oscillations, believed to be responsible for phonemic processing, are enhanced between birth and 6 months of age especially for the native language [51]. Since amplitude tracking continues to be found for unfamiliar languages, it is thus evident that, for these languages, the shift in the focus from syllabic to phonetic units does not occur.

On the contrary, based on the obtained results, phase tracking of the speech envelope takes place in the same way for both newborn and six months old participants, but also for adult subjects [7]. Since this mechanism is observed at birth, like amplitude envelope tracking, it is likely to occur even in absence of comprehension. This hypothesis is supported by previous evidences of phase tracking in adult subjects even for unintelligible speech (accelerated and time reversed speech) [9, 8, 7]. The fact that phase tracking in 6 months old babies takes place for both native and unfamiliar languages in the same way, represents a possible evidence that it may be less influenced by experience and comprehension with respect to amplitude tracking. From the obtained results, in addition with observations of phase and amplitude tracking, respectively at lower ( $\delta$  and  $\theta$  band) [2, 9] and higher ( $\gamma$  band) [3, 4, 50] frequencies, it can be hypothesized that these two mechanisms play complementary roles in speech comprehension. In particular, it is proposed that phase tracking represents a low level mechanism of speech processing that occur for many auditory stimuli, even unintelligible ones, and that it may precede and modulate amplitude tracking, that instead modulates the auditory stimuli at syllabic level.

In any case, the observed differences in speech envelope tracking during development are not likely to be due to different states of attention between the two studied groups of participants (during the

experiment newborns were awake at rest or asleep, while 6-month-olds were awake at rest). In fact it has been proven that newborn's auditory cortex is active during sleep [53] and, for this reason, they are even able to learn from auditory stimuli while asleep [54]. Furthermore, effects of this type should impact all languages uniformly, whereas differences have been observed only for specific languages.

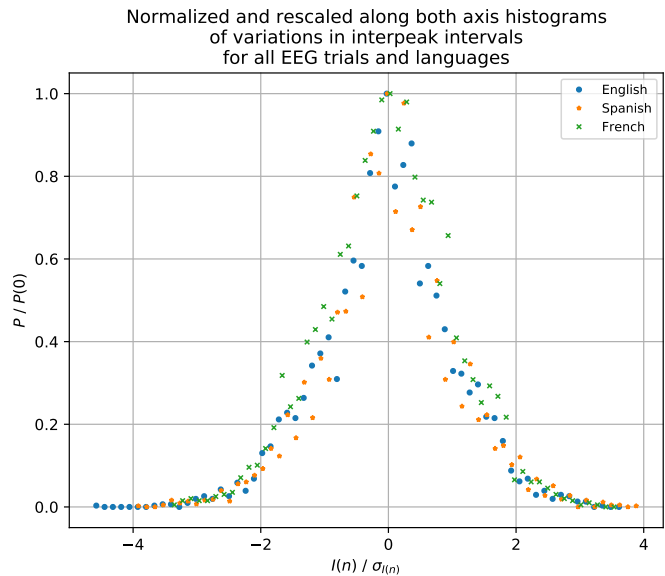
Future studies may investigate in which moment of childhood, amplitude tracking reappears, maybe assessing this mechanism in 12, 18 or 24 months old infants. Furthermore, it may be interesting to adopt time-frequency methods in order to study EEG data from a different and less exploited perspective. In relation to section 3.3, for future work, we propose to extend this preliminary analysis for all channels and conditions in order to perform a full connectivity analysis and assess the level of synchronization between different brain areas in infants during the listening of a speech.

Since, due to the shortness of the signals, we did not fully investigate the presence of critical dynamics in measured EEG data from infants' auditory cortex, in this section we also briefly resume the few applied methods and obtained results. In order to do a proper statistical inference on criticality, based on the observation of power law scaling, it is in fact important that the data range always span at least more than two orders of magnitude. As future perspective we propose thus to perform a full analysis on criticality by taking longer measurements of EEG cortical activity, with respect to those provided by experiment [10].

For our analysis we followed some of the steps that were performed in a study that found evidence of criticality in the human heartbeat [24]. In this experiment, the authors studied the properties of the time series  $B(n)$  obtained by plotting the sequential heartbeat intervals between beat  $n$  and beat  $n + 1$ , for healthy and heart diseased subjects. These time series typically reveal a complex variability and, in order to show possible differences between healthy and diseased subject it is useful to study the probability distributions of interbeat increments, defined as:  $I(n) = B(n + 1) - B(n)$ . With a different meaning, in our case of study, we can hence define  $B(n)$  as the time interval between two sequential peaks in EEG filtered signal (to better visualize this concept we can take as reference the filtered EEG signals plotted in figures 3.3a, 3.4a and 3.5a) and  $I(n)$  as the variation between successive interpeak intervals. With these conceptual modifications, it is thus possible to investigate the variability in EEG signals between the three examined languages by plotting an histogram of interpeak intervals' variations for each language and by comparing them to each other.

Proceeding with analysis results, in order to facilitate comparisons between languages, we first plotted the histograms of frequencies. Then, by renormalizing them to 1, we obtained the respective histograms of probabilities and finally, by dividing variables  $I(n)$  by their respective standard deviations and by rescaling probabilities  $P$  for  $P(0)$ , for each language, we obtained the histograms of interpeak intervals' variations for each language, represented in figure 4.1.

By calculating Kurtosis indexes, it has been then quantified the deviation of the obtained distributions from a normal distribution (represented by a Kurtosis index equal to 0). For English and Spanish languages, calculated Kurtosis indexes are respectively equal to 0,587 and 0,688, indicating that these distributions are sharper than a Gaussian. While, for French, we obtained a Kurtosis index equal to



**Figure 4.1:** Histogram of the variability between the three studied languages.

$-0,711$ , which reflects the fact that its distribution is flatter

than a Gaussian. However, all the distributions are not so different from a normal distribution and it would be interesting to assess whether they are represented by Lévy stable distributions, as found in [24].

Language	Mean	Standard deviation
English	0,221	0,273
Spanish	0,224	0,274
French	0,314	0,320

**Table 4.1:** Relevant statistical quantities in variability between languages.

Other interesting statistical properties of the distribution in figure 4.1 are listed in table 4.1.

We can thus conclude that, since the distributions of differences in interpeak intervals are similar with each other, the variability between the languages is very low. As suggested above, for future studies on critical dynamics, it would be interesting to investigate an eventual  $1/f$  power law behaviour in  $I(n)$  power spectrum as has been done in [24].



# Bibliography

- [1] Ehud Ahissar, Srikantan Nagarajan, Merav Ahissar, Athanassios Protopapas, Henry Mahncke, and Michael M Merzenich. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, 98(23):13367–13372, 2001.
- [2] Huan Luo and David Poeppel. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6):1001–1010, 2007.
- [3] Daniel A Abrams, Trent Nicol, Steven Zecker, and Nina Kraus. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *Journal of Neuroscience*, 28(15):3958–3965, 2008.
- [4] Kirill V Nourski, Richard A Reale, Hiroyuki Oya, Hiroto Kawasaki, Christopher K Kovach, Haiming Chen, Matthew A Howard, and John F Brugge. Temporal envelope of time-compressed speech represented in the human auditory cortex. *Journal of Neuroscience*, 29(49):15564–15574, 2009.
- [5] Nai Ding, Aniruddh D Patel, Lin Chen, Henry Butler, Cheng Luo, and David Poeppel. Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, 81:181–187, 2017.
- [6] Léo Varnet, Maria Clemencia Ortiz-Barajas, Ramón Guevara Erra, Judit Gervain, and Christian Lorenzi. A cross-linguistic study of speech modulation spectra. *The Journal of the Acoustical Society of America*, 142(4):1976–1989, 2017.
- [7] Maria Pefkou, Luc H Arnal, Lorenzo Fontolan, and Anne-Lise Giraud.  $\theta$ -band and  $\beta$ -band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. *Journal of Neuroscience*, 37(33):7930–7938, 2017.
- [8] Benedikt Zoefel and Rufin VanRullen. Eeg oscillations entrain their phase to high-level features of speech sound. *NeuroImage*, 124:16–23, 2016.
- [9] Mary F Howard and David Poeppel. Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of neurophysiology*, 104(5):2500–2511, 2010.
- [10] Maria Clemencia Ortiz Barajas, Ramón Guevara, and Judit Gervain. The origins and development of speech envelope tracking during the first months of life. *Developmental cognitive neuroscience*, 48:100915, 2021.
- [11] Jacques Mehler, Peter Jusczyk, Ghislaine Lambertz, Nilofar Halsted, Josiane Bertoncini, and Claudine Amiel-Tison. A precursor of language acquisition in young infants. *Cognition*, 29(2):143–178, 1988.
- [12] Christine Moon, Robin Panneton Cooper, and William P Fifer. Two-day-olds prefer their native language. *Infant behavior and development*, 16(4):495–500, 1993.
- [13] Thierry Nazzi, Josiane Bertoncini, and Jacques Mehler. Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and performance*, 24(3):756, 1998.

- [14] Franck Ramus, Marc D Hauser, Cory Miller, Dylan Morris, and Jacques Mehler. Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288(5464):349–351, 2000.
- [15] Laura Bosch and Núria Sebastián-Gallés. Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, 65(1):33–69, 1997.
- [16] Monika Molnar, Judit Gervain, and Manuel Carreiras. Within-rhythm class native language discrimination abilities of basque-spanish monolingual and bilingual infants at 3.5 months of age. *Infancy*, 19(3):326–337, 2014.
- [17] Janet F Werker and Richard C Tees. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, 7(1):49–63, 1984.
- [18] Patricia K Kuhl, Karen A Williams, Francisco Lacerda, Kenneth N Stevens, and Björn Lindblom. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044):606–608, 1992.
- [19] Andrea Cavagna, Alessio Cimarrelli, Irene Giardina, Giorgio Parisi, Raffaele Santagati, Fabio Stefanini, and Massimiliano Viale. Scale-free correlations in starling flocks. *Proceedings of the National Academy of Sciences*, 107(26):11865–11870, 2010.
- [20] Robert Burridge and Leon Knopoff. Model and theoretical seismicity. *Bulletin of the seismological society of america*, 57(3):341–371, 1967.
- [21] Kai Nagel and Hans J Herrmann. Deterministic models for traffic jams. *Physica A: Statistical Mechanics and its Applications*, 199(2):254–269, 1993.
- [22] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality. *Physical review A*, 38(1):364, 1988.
- [23] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality: An explanation of the 1/f noise. *Physical review letters*, 59(4):381, 1987.
- [24] C-K Peng, J Mietus, JM Hausdorff, Shlomo Havlin, H Eugene Stanley, and Ary L Goldberger. Long-range anticorrelations and non-gaussian behavior of the heartbeat. *Physical review letters*, 70(9):1343, 1993.
- [25] Mike X Cohen. *Analyzing neural time series data: theory and practice*. MIT press, 2014.
- [26] Jianhong Wu. *Introduction to neural dynamics and signal transmission delay*, volume 6. Walter de Gruyter, 2011.
- [27] vivadifferences.com. 8 major difference between action potential and resting potential. <https://vivadifferences.com/difference-between-action-potential-and-resting-potential/>, accessed: 22.06.2021.
- [28] Claude Bedard, Helmut Kroeger, and Alain Destexhe. Does the 1/f frequency scaling of brain signals reflect self-organized critical states? *Physical review letters*, 97(11):118102, 2006.
- [29] Florian Mormann, Klaus Lehnertz, Peter David, and Christian E Elger. Mean phase coherence as a measure for phase synchronization and its application to the eeg of epilepsy patients. *Physica D: Nonlinear Phenomena*, 144(3-4):358–369, 2000.
- [30] Paul L Nunez, Ramesh Srinivasan, et al. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2006.
- [31] Dennis Gabor. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.

- [32] Francisco Varela, Jean-Philippe Lachaux, Eugenio Rodriguez, and Jacques Martinerie. The brain-web: phase synchronization and large-scale integration. *Nature reviews neuroscience*, 2(4):229–239, 2001.
- [33] Nikolai Axmacher, Florian Mormann, Guillen Fernández, Christian E Elger, and Juergen Fell. Memory formation by neuronal synchronization. *Brain research reviews*, 52(1):170–182, 2006.
- [34] Thomas E Akam and Dimitri M Kullmann. Efficient “communication through coherence” requires oscillations structured to minimize interference between signals. *PLoS Comput Biol*, 8(11):e1002760, 2012.
- [35] Pascal Fries. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences*, 9(10):474–480, 2005.
- [36] Wolf Singer. Synchronization of cortical activity and its putative role in information processing and learning. *Annual review of physiology*, 55(1):349–374, 1993.
- [37] Vikaas S Sohal, Feng Zhang, Ofer Yizhar, and Karl Deisseroth. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature*, 459(7247):698–702, 2009.
- [38] Matteo Feurra, Giovanni Bianco, Emiliano Santarnecchi, Massimiliano Del Testa, Alessandro Rossi, and Simone Rossi. Frequency-dependent tuning of the human motor system induced by transcranial oscillatory potentials. *Journal of Neuroscience*, 31(34):12165–12170, 2011.
- [39] Walter J Freeman. Simulation of chaotic eeg patterns with a dynamic model of the olfactory system. *Biological cybernetics*, 56(2):139–150, 1987.
- [40] Walter S Pritchard. The brain in fractal time: 1/f-like power spectrum scaling of the human electroencephalogram. *International Journal of Neuroscience*, 66(1-2):119–129, 1992.
- [41] Cornelis Jan Stam and Eveline Astrid De Bruin. Scale-free dynamics of global functional connectivity in the human brain. *Human brain mapping*, 22(2):97–109, 2004.
- [42] John M Beggs and Dietmar Plenz. Neuronal avalanches in neocortical circuits. *Journal of neuroscience*, 23(35):11167–11177, 2003.
- [43] Mikail Rubinov, Olaf Sporns, Jean-Philippe Thivierge, and Michael Breakspear. Neurobiologically realistic determinants of self-organized criticality in networks of spiking neurons. *PLoS Comput Biol*, 7(6):e1002038, 2011.
- [44] Yogesh S Virkar, Woodrow L Shew, Juan G Restrepo, and Edward Ott. Feedback control stabilization of critical dynamics via resource transport on multilayer networks: How glia enable learning dynamics in the brain. *Physical Review E*, 94(4):042310, 2016.
- [45] Luca Cocchi, Leonardo L Gollo, Andrew Zalesky, and Michael Breakspear. Criticality in the brain: A synthesis of neurobiology, models and cognition. *Progress in neurobiology*, 158:132–152, 2017.
- [46] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
- [47] Alexandros Potamianos and Petros Maragos. A comparison of the energy operator and the hilbert transform approach to signal and speech demodulation. *Signal processing*, 37(1):95–120, 1994.
- [48] Christiaan Huygens. *Horologium oscillatorium*. 1980.
- [49] Michael G Rosenblum, Arkady S Pikovsky, and Jürgen Kurths. Phase synchronization of chaotic oscillators. *Physical review letters*, 76(11):1804, 1996.
- [50] Jan Kubanek, Peter Brunner, Aysegul Gunduz, David Poeppel, and Gerwin Schalk. The tracking of speech envelope in the human cortex. *PloS one*, 8(1):e53398, 2013.

- [51] Silvia Ortiz-Mantilla, Jarmo A Hämäläinen, Gabriella Musacchia, and April A Benasich. Enhancement of gamma oscillations indicates preferential processing of native over foreign phonemic contrasts in infants. *Journal of Neuroscience*, 33(48):18746–18754, 2013.
- [52] Janet F Werker and Takao K Hensch. Critical periods in speech perception: new directions. *Annual review of psychology*, 66:173–196, 2015.
- [53] Anke Sambeth, Katja Ruohio, Paavo Alku, Vineta Fellman, and Minna Huotilainen. Sleeping newborns extract prosody from continuous speech. *Clinical Neurophysiology*, 119(2):332–341, 2008.
- [54] William P Fifer, Dana L Byrd, Michelle Kaku, Inge-Marie Eigsti, Joseph R Isler, Jillian Grose-Fifer, Amanda R Tarullo, and Peter D Balsam. Newborn infants learn during sleep. *Proceedings of the National Academy of Sciences*, 107(22):10320–10323, 2010.